PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Analysis and Prediction of Structural Motifs in the Glycolytic Enzymes [and Discussion]

M. J. E. Sternberg, F. E. Cohen, W. R. Taylor, R. J. Feldmann and J. D. G. Smit

Phil. Trans. R. Soc. Lond. B 1981 **293**, 177-189
doi: 10.1098/rstb.1981.0071

Phil. Trans. R. Soc. Lond. B **293**, 177–189 (1981)
Printed in Great Britain

Analysis and prediction of structural motifs in the glycolytic enzymes

By M. J. E. Sternberg†, F. E. Cohen†, W. R. Taylor† and R. J. Feldmann‡

† *Laboratory of Molecular Biophysics, Department of Zoology, South Parks Road, Oxford OX1 3PS, U.K.*

‡ *National Institutes of Health, Bethesda, Maryland* 20014, *U.S.A.*

[Plates 1 and 2]

Protein crystallography has determined the three-dimensional structures of 10 of the 13 enzymes of the glycolytic pathway. Diagrams and details of these enzyme structures are given in the paper. Most of the enzyme domains are variations and extensions of a many (4–9)-stranded, predominantly or totally parallel, β-sheet that is shielded from solvent by α-helices (i.e. α/β structures). There are strong structural similarities between the domains of some, but not all, of the enzymes. In particular the dinucleotide binding fold of lactate dehydrogenase and the β-barrel of triose phosphate isomerase are found in other domains. General rules governing the topology and packing of α-helices against a β-sheet provide a basis for the combinatorial prediction of the tertiary fold of glycolytic domains from their amino acid sequence and observed secondary structure. The prediction algorithm demonstrates that there are severe restrictions on the number of possible structures. However, these restrictions do not fully explain some of the remarkable structural similarities between different enzymes that probably result from evolution from a common ancestor.

1. Introduction

Over the last 10 years there has been a concerted effort by protein crystallographers to determine the three-dimensional structures of the enzymes of the glycolytic pathway (see, for example, Blake 1975). Today this goal is close to being realized with at least chain tracings and often high-resolution structures for 10 of the 13 enzymes (see table 1). One question that motivated this study is whether the enzymes, with their related functions, have similar structures. In this review we show that all the enzyme structures are variations and extensions of a basic theme of a many (4–9)-stranded, predominantly parallel, β-sheet that is surrounded by α-helices. Beyond this generalization, only some of the molecules have domains that have a strong structural similarity. The current understanding of the general rules governing the topology and the packing of α-helices against a β-sheet will be demonstrated by the results of an algorithm to predict the tertiary fold of glycolytic domains from their amino acid sequences. Relations between enzyme structure and function will be outlined. Finally, we will consider whether the glycolytic enzymes have evolved by divergence from a common ancestor or by convergence to a stereochemically and/or functionally favourable folding.

2. The structures of the enzymes

A superficial view, analogous to that of a low-resolution electron density map, suggests considerable variations between the different three-dimensional structures (see table 2). The enzymes can be monomeric, dimeric or tetrameric with subunits ranging between 230 residues

(PGAM) and 841 residues (PHO), (see table 1 for references to and abbreviations of the enzymes). Polypeptide chains longer than 250 residues will segregate into distinct folding units or domains, each involving between *ca.* 130 and *ca.* 490 residues.

The high-resolution structures are shown in figure 5 (plates 1 and 2) in terms of the packing of α-helices and β-sheets. However, to compare these structures it is helpful to use an even more schematic representation (figure 6), which shows the positions of the component β-strands (triangles) and α-helices (circles).

Levitt & Chothia (1976) have shown that most protein domains belong to one of four structural classes: all-α, predominantly involving the packing of α-helices; all-β, with primarily β-structure; α/β, where the secondary structure alternates between α-helices and β-strands;

TABLE 1. THE STRUCTURE DETERMINATION OF THE GLYCOLYTIC ENZYMES

			structure determination			
enzyme	code	reaction	source	resolution/Å	group	reference
phosphorylase	PHO	glycogen ⇌ G1P	rabbit	3.0	Johnson (Oxford)	Johnson *et al.* (1980)
			rabbit	2.5	Fletterick (San Francisco)	Spang & Fletterick (1979)
phospho-glucomutase	PGM	G1P ⇌ G6P	rabbit	—	Rossmann (Purdue)	M. G. Rossmann, personal communication
hexokinase	HK	G6P ⇌ glucose	yeast	2.1	Steitz (Yale)	Steitz *et al.* (1976); Anderson *et al* (1978)
phosphoglucose isomerase	PGI	G6P ⇌ F6P	pig	3.5	Muirhead (Bristol)	Shaw & Muirhead (1977)
phosphofructo-kinase	PFK	F6P ⇌ F1,6P_2	*B. stearo-thermophilus*	2.4	Evans (Cambridge)	Evans & Hudson (1979)
aldolase	ALD	F1, 6P_2 ⇌ G3P	human	5	Watson (Bristol)	H. Watson, personal communication
triose phosphate isomerase	TIM	G3P ⇌ DHAP	chicken	2.6	Phillips (Oxford)	Banner *et al.* (1975)
			yeast		Petsko (M.I.T.)	G. A. Petsko, personal communication
glyceraldehyde-3-phosphate dehydrogenase	GPD	G3P ⇌ 1,3P_2G	lobster	3.0	Rossmann (Purdue)	Moras *et al.* (1975)
			B. stearo-thermophilus	2.7	Wonacott (London)	Biesecker *et al.* (1977)
phosphoglycerate kinase	PGK	1,3P_2G ⇌ 3PG	horse	2.5	Blake (Oxford)	Banks *et al.* (1979)
			yeast	2.5	Watson (Bristol)	Bryant *et al.* (1974)
phosphoglycerate mutase	PGAM	3PG ⇌ 2PG	yeast	2.8	Watson (Bristol)	Campbell *et al.* (1974)
enolase	ENO	2PG ⇌ PEP	chicken	—	Watson (Bristol)	H. Watson, personal communication
pyruvate kinase	PK	PEP ⇌ pyruvate	cat	2.6	Muirhead (Bristol)	Stuart *et al.* (1979)
lactate dehydrogenase	LDH	pyruvate ⇌ lactate	dogfish	2.0	Rossmann (Purdue)	Holbrook *et al.* (1975)

Adapted from Blake (1975). The abbreviations for the substrates are: G1P, α-D-glucose-1-phosphate; 6GP, D-glucose-6-phosphate; F6P, α-D-fructose-6-phosphate; F1,6P_2, α-D-fructose-1,6-bisphosphate; G3P, D-glyceraldehyde-3-phosphate; DHAP, dihydroxyacetone phosphate; 1,3P_2G, 1,3-bisphospho-D-glycerate; 3PG, 3-phospho-D-glycerate; 2PG, 2-phospho-D-glycerate; PEP, phosphoenolpyruvate.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

and (α + β), in which there are both α-helices and β-strands but these secondary structures tend to segregate separately (figure 1). Most of the glycolytic domains are of the α/β class. The alternating α/β structure produces a predominantly parallel β-sheet, with 4–9β-strands; the α-helices, generally lying antiparallel to the strand direction, shield most of the β-sheet from solvent. This α/β structure is typified by the N-terminal (NAD) domain of LDH with its sheet formed from strands FEDABC (figure 6). However, the other structural classes occasionally

TABLE 2. STRUCTURAL DIVERSITY OF THE GLYCOLYTIC ENZYMES

			domains		structure	
enzyme	number of subunits	number of residues per monomer	number	number of residues	class	sheet type
PHO	2	841 (rabbit)	2	489	all-α + α/β	9 β mixed
				352	all-α + α/β	6 β pure d.n.b.f.
PGM	1	*ca.* 590 (rabbit)	—	—	—	—
HK	2	*ca.* 457 (yeast)	2	290	all-α + α/β	6 β mixed
				167	α/β	5 β mixed
PGI	2	*ca.* 514 (pig)	2	360	α/β	6 β?
				154	α/β	4 β?
PFK	4	316 (*B. stearothermophilus*)	2	186	α/β	7 β mixed
				130	α/β	4 β pure
ALD	2 or 4	361 (rabbit)	—	—	—	—
TIM	2	247 (chicken)	1	247	α/β	8 β pure barrel
GPD	4	334 (lobster)	2	170	α/β	9 β mixed d.n.b.f.
				164	α/β	7 β mixed
PGK	1	416 (horse)	2	210	α/β	6 β pure
				206	α/β	6 β pure d.n.b.f.
PGAM	4	230 (yeast)	1	230	α/β	5 β mixed
ENO	2	*ca.* 400 (rabbit)	—	—	—	—
PK	4	*ca.* 550 (cat)	3	*ca.* 200	α/β	8 β pure barrel
				ca. 175	all-β	7 β mixed
				ca. 175	α/β	5 β mixed
LDH	4	332 (dogfish)	2	182	α/β	6 β pure d.n.b.f.
				150	α + β	3 + 3 β mixed

The structural class refers to the classification by Levitt & Chothia (1976) (see §2). The sheet type gives the number of β-strands in the sheet, whether the sheet is pure parallel (i.e. all adjacent strands run parallel) or mixed, and indicates if there is a dinucleotide binding fold (d.n.b.f.) or a β-barrel.

occur: the (α + β) fold of the C-terminal domain of LDH, the all-β structure of the second domain of PK, and a combination of all-α with α/β in the large domains of PHO.

The polypeptide chain segregates into the different domains in a variety of ways. One frequent feature is that a chain requires more than n-1 cuts to dissect it into its n domains. For example, in GPD the chain forms most of the N-terminal domain, then all of the second domain, and finally the C-terminal α-helix associates with the rest of the N-terminal domain. This generalization does not apply to the largest of the enzymes, PHO, and the topology of domain segregation in proteins requires further analysis.

A more detailed examination of the individual α/β domains starts with the topology of the fold: the number, order and direction of the β-strands in the β-sheet, including whether all adjacent pairs of β-strands lie parallel (pure-parallel) or not (mixed β-sheet); and the

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

conformation and position of the connections between these β-strands. Figures 5 and 6 show that there is considerable variation in the topologies of the enzymes. However, certain specific sheet–helix topologies are seen in several domains resulting from the presence of a substructure formed from two parallel β-strands that generally, but not always, lie adjacent in the sheet together with the interstrand connection: a β × β unit (figure 2). Generally the connection

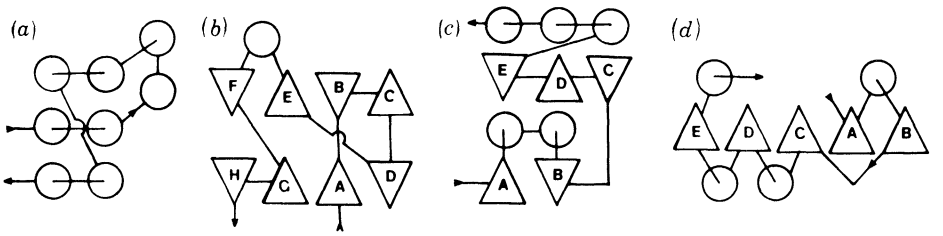


FIGURE 1. The four structural classes of globular proteins: (*a*) myoglobin (all-α); (*b*) transthyretin ('prealbumin') (all-β); (*c*) lysozyme (α + β); (*d*) flavodoxin (α/β). In the schematic diagrams each α-helix is represented by a circle and each β-strand by a triangle (see figure 6 for further details of this representation). For references to structures see Sternberg & Thornton (1976).

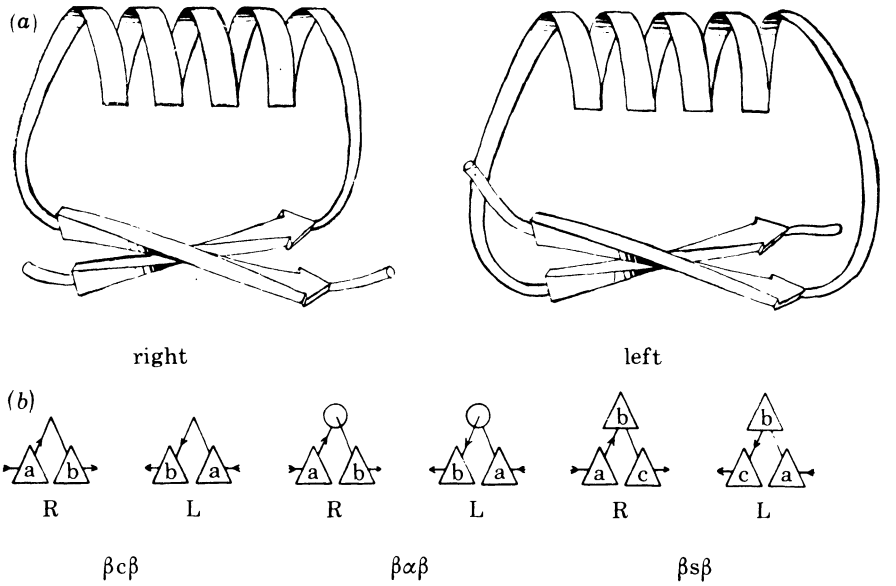


FIGURE 2. (*a*) Right-handed and left-hanced βαβ units in which the β-strands are represented by twisted ribbons and the α-helix as a spiral. (*b*) Schematic diagrams of right (R) and left (L) βcβ, βαβ and βsβ units.

involves one or more α-helices, one of which lies roughly antiparallel to the strand direction and packs against the β-sheet, and thereby forms a βαβ unit. However, the entire connection can adopt an irregular coil conformation (βcβ unit) or the connection can form part of another β-sheet (βsβ unit). An important feature of β × β units is that they have a chirality that depends on the handedness of the rotation in going along the first β-strand, back along the connection that lies on one side of the sheet, and finally along the second β-strand (Sternberg & Thornton 1976, 1977*a*; Richardson 1976). It is the repeated formation of β × β units, particularly βαβ units, by sequential sections of the polypeptide chain that produces structures that are common to different glycolytic domains.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

One such common structure is the six-stranded, pure parallel β-sheet in the NAD-binding domain of LDH (figure 3). This motif, often called the dinucleotide binding fold (d.n.b.f.), consists of five overlapping β × β units: two pairs of adjacent βαβ units (AB, BC and DE, EF) and one βcβ unit (CD), all of which are right-handed (Rao & Rossmann 1973). The d.n.b.f. occurs with minor differences in three other domains (figure 3). Both the ATP-binding domain of PGK and the second (pyridoxal-5′-phosphate binding) domain of PHO have six β-strands

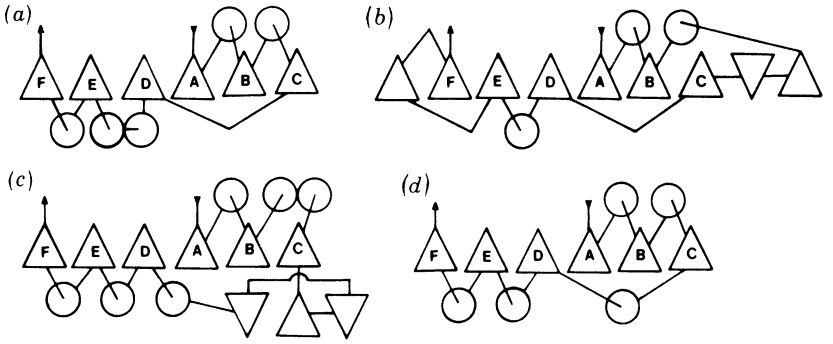


FIGURE 3. Four protein domains with the dinucleotide binding fold (d.n.b.f.): schematic diagrams (see figure 6 for details) of the NAD-binding domains of (*a*) LDH and (*b*) GPD; (*c*) the ATP-binding domain of PGK; (*d*) the pyridoxal phosphate (PLP)-binding domain of PHO. The six β-strands forming the d.n.b.f. are lettered ABCDEF sequentially.

with the same topology and with βαβ units corresponding to the four on LDH. The central six of the nine β-strands in the NAD-binding domain of GPD resemble the d.n.b.f. However, the d.n.b.f. is not restricted to the glycolytic enzymes as it forms the NAD-binding domain of alcohol dehydrogenase (Eklund *et al.* 1976) and the CTP-binding domain of aspartate carbamoyltransferase (Honzatko 1979). Furthermore, the C_α positions of most of the residues that form the α-helices and β-strands in one d.n.b.f. can be readily superimposed on the C_α positions of the other d.n.b.f. These superpositions yield an r.m.s. deviation between equivalenced *ca.* 54 C_α atoms of *ca.* 3 Å (Ohlsson *et al.* 1974). These seemingly remarkable structural similarities raise the question of whether the d.n.b.f. domains have evolved from a common ancestor and hence share a common architecture (see §5).

A similar question is raised by the structural similarity between TIM and the N-terminal domain of PK. The TIM structure consists of an eightfold repeat of a β-strand followed by one or two α-helices forming seven overlapping right-handed βαβ units. Sequential β-strands are laid down in one β-sheet adjacent and parallel to each other, with the last β-strand adjacent to the first, thereby forming a β-barrel. The connecting α-helices run antiparallel to the strand direction and pack outside and against the barrel. The N-terminal domain of PK has an eight-stranded barrel flanked by α-helices that closely resembles that of TIM. Furthermore, in TIM there is a long loop (15 residues) after strand C that interacts with other subunits, while in PK after strand C the polypeptide chain leaves the β-barrel to form the second domain; 160 C_α atoms of TIM and PK can be superimposed with an r.m.s. deviation of 3.0 Å (Levine *et al.* 1978).

The d.n.b.f. of LDH can be dissected into two substructures, mononucleotide binding folds (m.n.b.f.), involving two overlapping right-handed βαβ units (strands ABC and strands FED). Similarly the β-barrels of TIM and PK can be decomposed into m.n.b.f. units. This motif,

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

sometimes with βcβ units rather than βαβ units, is found in several other glycolytic domains – strands PQR of PK strands FGH and GHI in PGI, and strands JKL in the first domain of PHO – but also in other enzymes, e.g. strands ABC in flavodoxin (Burnett *et al.* 1974) (see figures 5 and 6).

Apart from these common motifs, the structural similarities between the different domains result from their α/β motif formed from right-handed β × β, particularly βαβ, units. Indeed, there are different topologies for the β-strands in all the other domains.

3. Topological and packing rules

The above structural observations will now be placed in the context of topological and packing rules for α/β structures that were obtained by analysis of *all* known protein structures. One seemingly remarkable feature of the glycolytic enzymes is the widespread occurrence of right-handed β × β, particularly βαβ, units. However, in all known structures there are more than 100 right-handed β × β units, many of which are βαβ-type, and only three left-handed units: a βαβ unit in subtilisin, the βsβ unit between strands E and J in PFK, and, in the tentative chain tracing of PGI, the connection between strands D and E (Sternberg & Thornton 1976, 1977*a*; Richardson 1976; Nagano 1977). This surprising tendency for right-handedness probably results from the twisted nature of the β-sheet (Chothia 1973) that produces a left-handed rotation between β-strands (figure 2). Consequently the right-handed connection provides a more direct path for the connection than is available in a left-handed unit. Consistent with this explanation is the greater length of connections in the left-handed structures than in many of the right-handed units.

There are several other topological tendencies for β-sheets (Levitt & Chothia 1976; Sternberg & Thornton 1977*b*; Richardson 1977). For example, sequential β-strands tend to be adjacent (i.e. hydrogen-bonded) in the β-sheet, probably because of the pathway along which the protein folds. However, more restrictive than these tendencies are three rules that apply to all known pure-parallel β-sheets: (1) pure-parallel sheets in α/β structures will occur rather than mixed sheets if there are sufficient residues in the connections to span the longer distance required for parallel connections rather than antiparallel ones (F. E. Cohen, M. J. E. Sternberg & W. R. Taylor, unpublished); (2) pure-parallel β-sheets will have one-chain reversals if they do not form a β-barrel ('planar' sheets) (e.g. in LDH the reversal is between strands BC and DC) and no reversals if they do (Richardson 1977); (3) 'Planar' sheets will have a difference of no more than one between the number of α-helical connections that lie above and the number below the sheet. In β-barrels all the α-helices will lie on the outside of the barrel (Cohen *et al.* 1980). These rules might result from the requirements of the folding pathway (rules 1 and 2) and the need for pure-parallel β-sheets to be shielded from the solvent by α-helices (rule 3).

At a more detailed level than these topological rules are the recent studies of the packing of α-helices against the β-sheet (Chothia *et al.* 1977; Janin & Chothia 1980; Cohen *et al.* 1981). We studied α-helix/β-sheet packing by using calculations of the area of the van der Waals surface (contact area) of protein segments that can be in contact with a hypothetical water probe that represents the solvent (Richmond & Richards 1978). The change of contact area of the non-polar atoms (ϕ-area) on docking the α-helix onto the β-sheet can be roughly related to the hydrophobic contribution to the free energy of association of these structures (1 Å ≡ 334

kJ/mol (Richmond & Richards 1978)). In seven different proteins, only four of which were glycolytic enzymes (TIM, PGK, LDH, GPD), there were 46 α-helix–β-sheet interactions with a ϕ-area change greater than 33 Å ($\equiv 10.9$ kJ/mol). For these interactions there was a well defined geometry with the mean ($\pm$ standard deviation) of the distance and the dihedral angle between the α-helix axis and an average strand axis of 9.8 ± 1.4 Å and $170 \pm 30°$. The residues that mediate the interaction have changes in contact area and the residues involved generally trace a distinct pattern on the surfaces of the α-helix and the β-sheet (e.g. figure 4). On the

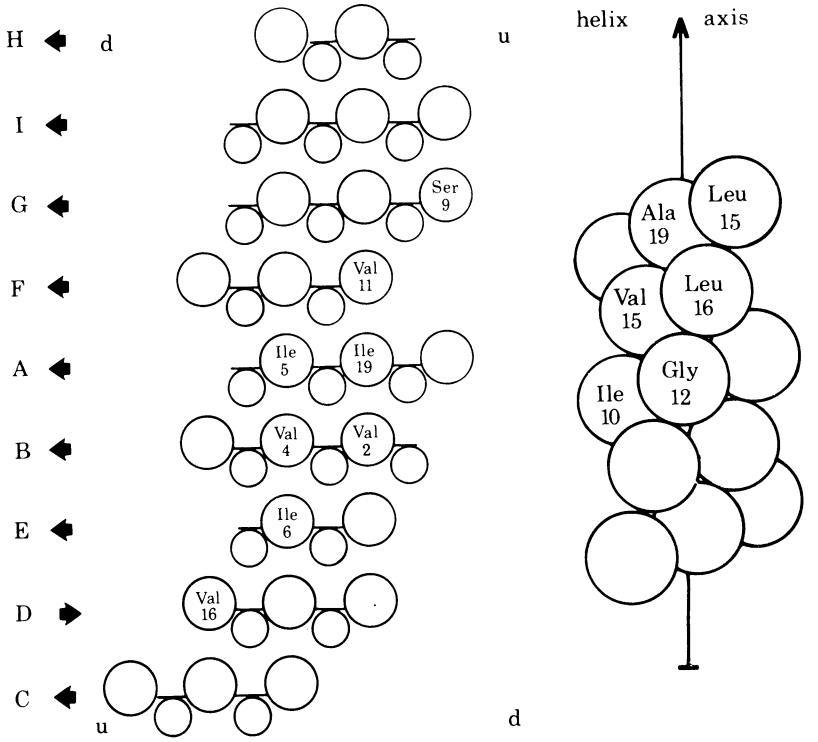


FIGURE 4. The non-polar contact area (square ångströms) changes on docking α-helix (residues 9–22) (right) onto the β-sheet in lobster GPD. The α-helix lies on top of the β-sheet. The letters u and d indicate the upward and downward pointing corners of the twisted β-sheet. The upward and downward pointing residues in the sheet are indicated by large and small circles respectively.

α-helix the pattern involves two adjacent ridges of predominantly non-polar residues at sequence positions i, $i+4$, $i+8$ and $i+1$, $i+5$, $i+9$. The β-sheet pattern is more variable but tends to involve non-polar residues, particularly branched aliphatics (Val, Ile, Leu) that form a parallelogram-shaped patch whose longer diagonal goes from the top left to the bottom right corner of the β-sheet when viewed as in figure 4. This pattern results from the twist of the β-sheet (Chothia 1973) that raises the top left and lower right corners of the β-sheet towards the interacting α-helix (see figure 4).

4. PREDICTION OF DOMAIN STRUCTURE

The extent to which these topological and packing rules restrict the number of possible structures for a glycolytic domain can be assessed from our combinatorial approach to predict the tertiary structure of proteins from their amino acid sequence and the correct (i.e. crystallographic)

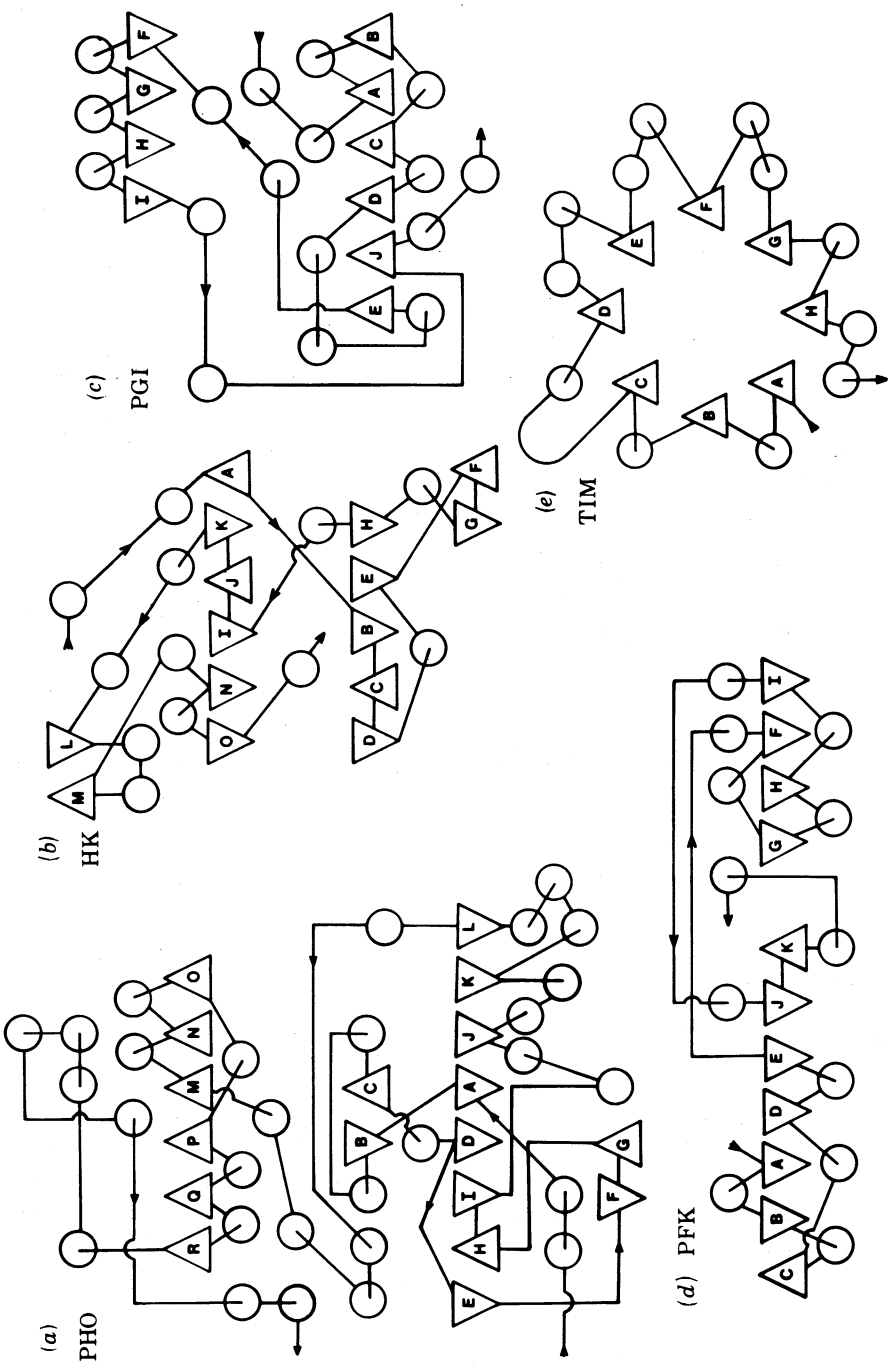


FIGURE 6. Schematic diagrams of the tertiary fold of the glycolytic enzymes. Each β-sheet is viewed along the strand direction. A triangle whose apex points up or down indicates whether the component β-strand is viewed from its N-terminus or its C-terminus. A circle represents an α-helix and if the α-helix runs roughly parallel or antiparallel to the strand direction, the connection is shown entering or leaving the circle. The views roughly correspond to those in figure 5. See table 1 for details of abbreviations used.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

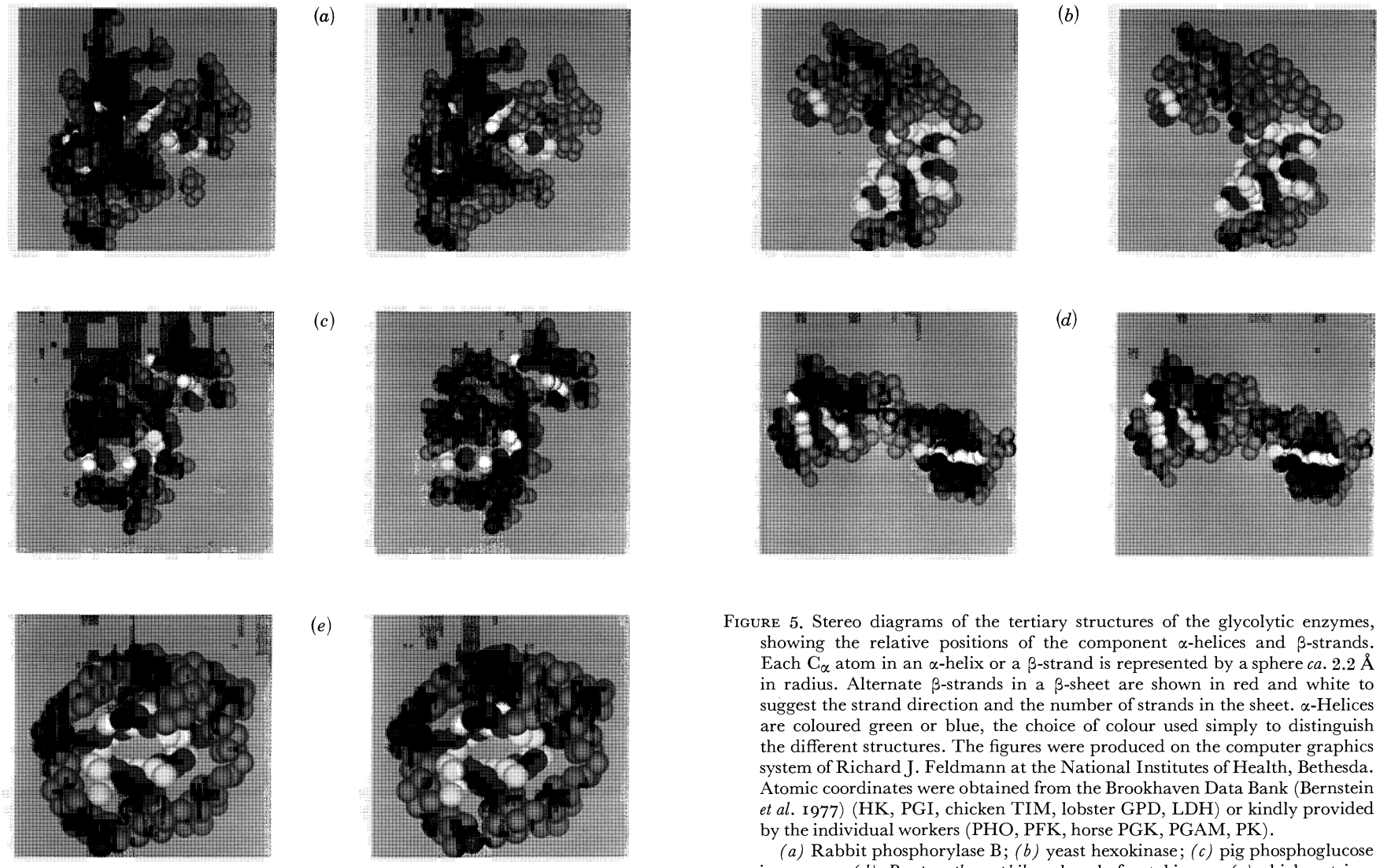


FIGURE 5. Stereo diagrams of the tertiary structures of the glycolytic enzymes, showing the relative positions of the component α-helices and β-strands. Each C_α atom in an α-helix or a β-strand is represented by a sphere *ca.* 2.2 Å in radius. Alternate β-strands in a β-sheet are shown in red and white to suggest the strand direction and the number of strands in the sheet. α-Helices are coloured green or blue, the choice of colour used simply to distinguish the different structures. The figures were produced on the computer graphics system of Richard J. Feldmann at the National Institutes of Health, Bethesda. Atomic coordinates were obtained from the Brookhaven Data Bank (Bernstein *et al.* 1977) (HK, PGI, chicken TIM, lobster GPD, LDH) or kindly provided by the individual workers (PHO, PFK, horse PGK, PGAM, PK).

(a) Rabbit phosphorylase B; *(b)* yeast hexokinase; *(c)* pig phosphoglucose isomerase; *(d) B. stearothermophilus* phosphofructokinase; *(e)* chicken triose phosphate isomerase.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

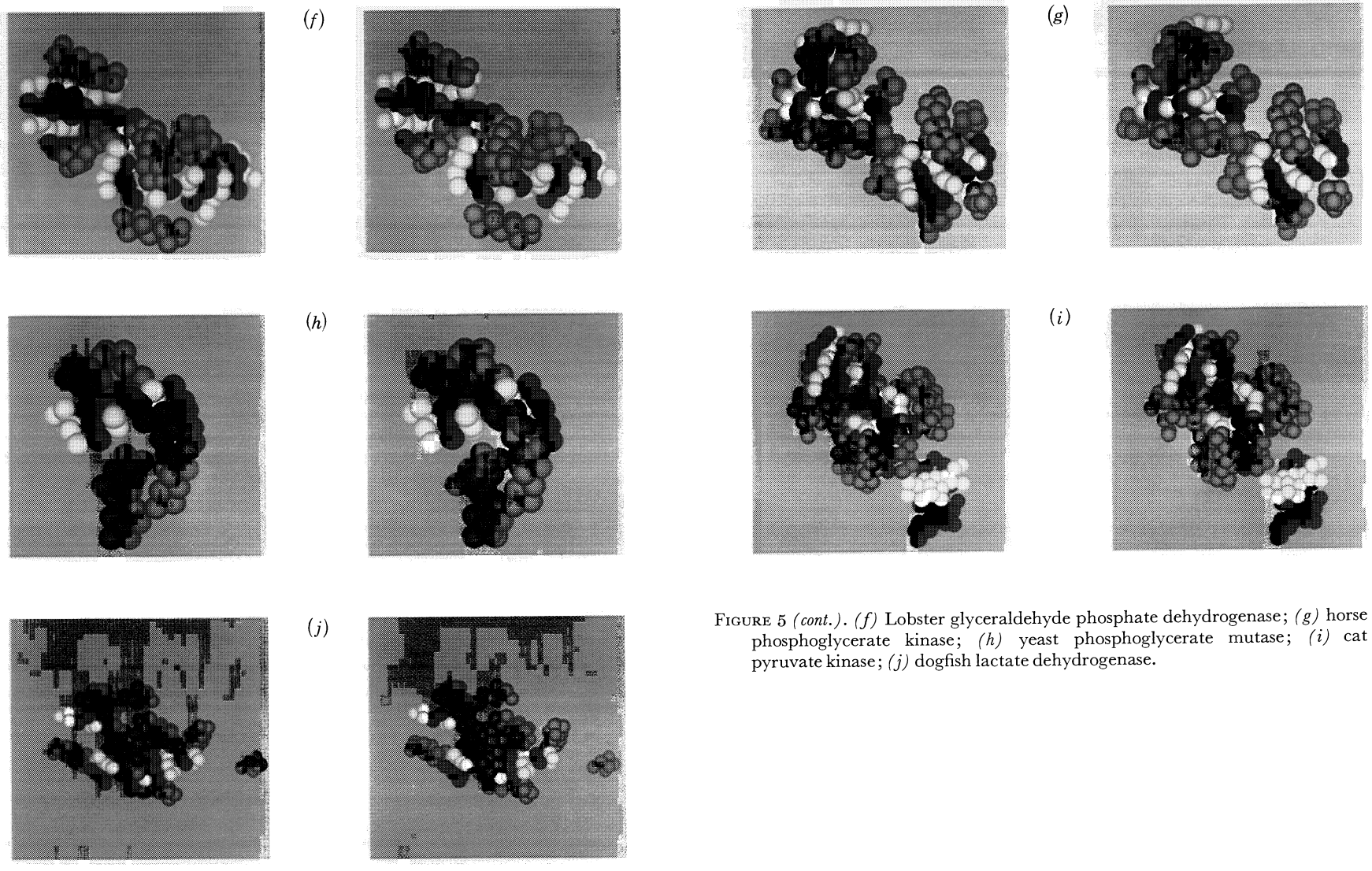


FIGURE 5 *(cont.)*. *(f)* Lobster glyceraldehyde phosphate dehydrogenase; *(g)* horse phosphoglycerate kinase; *(h)* yeast phosphoglycerate mutase; *(i)* cat pyruvate kinase; *(j)* dogfish lactate dehydrogenase.

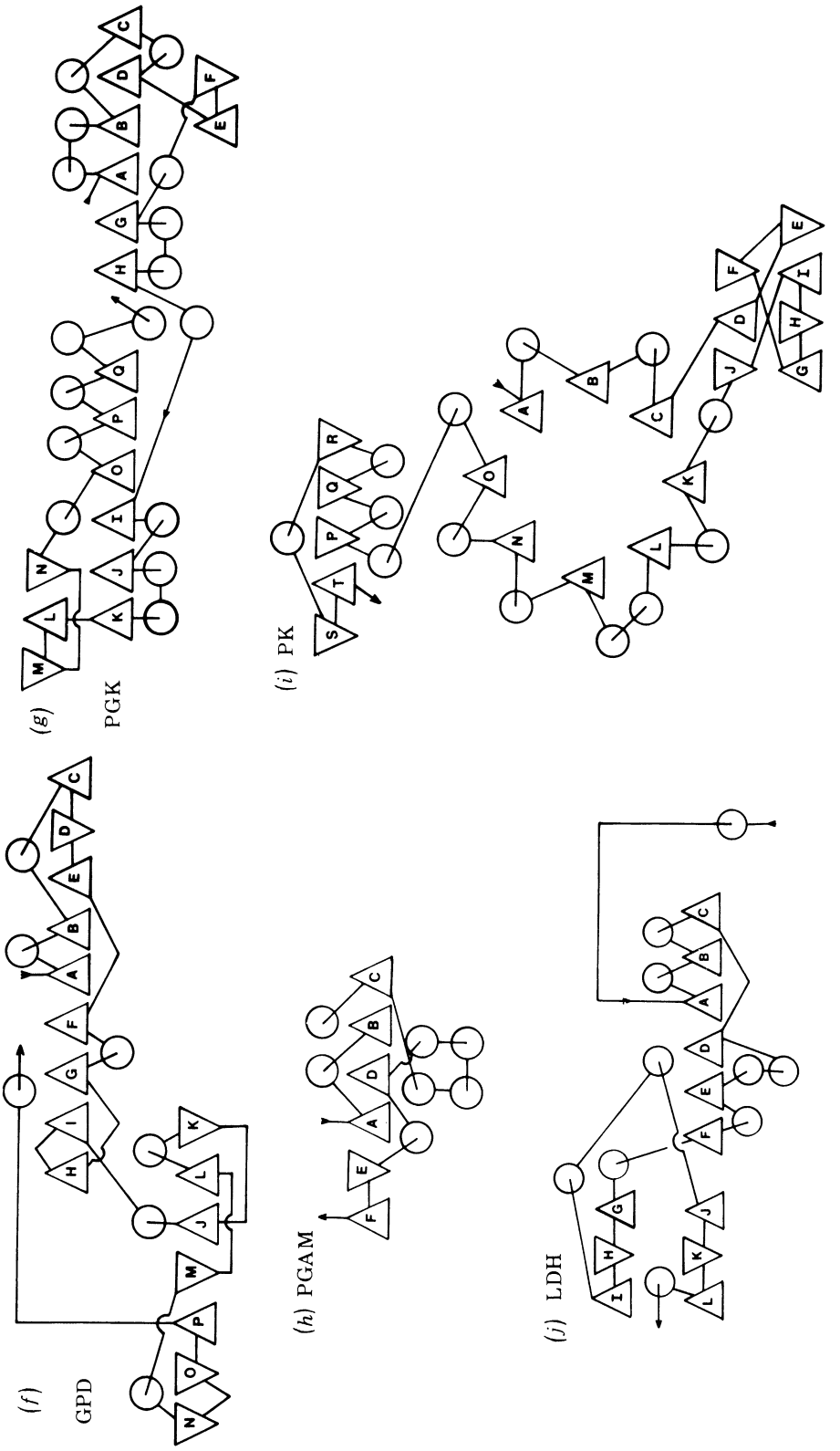


FIGURE 6 (*cont.*). For description see p. 184.

assignment of secondary structure (Cohen *et al.* 1980). In this approach one first generates all combinations of packed α-helices and β-strands, and then structures are eliminated if they violate any stereochemical rule.

Consider the pure-parallel β-sheets. The n β-strands can have $\frac{1}{2}n!$ strand orders and each of the $(n-1)$ connections can be right-handed or left-handed. Thus there are $\frac{1}{2}n!2^{n-1}$ possible topologies (table 3). For sheets with more than six strands this number is doubled to allow for the formation of β-barrels. However, the restriction on handedness and the use of rules 2 and 3

TABLE 3. THE PREDICTION OF THE TERTIARY FOLDING OF GLYCOLYTIC DOMAINS

		topologies		structures		r.m.s. deviation, predicted/crystal structures	
domain	strands	possible	allowed	possible	rank-ordered position of native	Å	residues
LDH (NAD-binding)	6	10^4	13	5×10^7	96	5.0	105
PGK (catalytic)	6	10^4	10	5×10^7	47	5.8	88
PGK (ATP-binding)	6	10^4	10	5×10^7	2	7.3	112
TIM	8	5×10^6	21	3×10^{11}	1	5.2	158

The number of possible topologies (β-strand order and whether the connections lie above or below the β-sheet) is restricted to the number allowed. The number of possible structures (topology, hydrogen-bonding and α-helix position) is restricted, and the position of the native-like structure in the list rank-ordered on strand overlap is given. The r.m.s. deviation (ångstroms) between the predicted and the crystal structures is reported.

for pure-parallel sheet topology markedly limits the number of possible topologies (Cohen *et al.* 1980).

Each topology represents several different structures depending on the stagger of the component β-strands (i.e. the hydrogen-bonding pattern), the position of the α-helix on the sheet, and the surface of the α-helix that packs against the β-sheet. A conservative estimate is that there are 4^n structures per topology for an n-stranded β-sheet. The requirements for specific patterns of hydrophobic residues on the surfaces of the β-sheet and the α-helices reduce the number of possible structures. This list of structures is then rank-ordered on the number of hydrogen bonds. Table 3 shows that less than 100 structures would have to be surveyed to locate the 'native' (cf. the notion of strand overlap in Cohen *et al.* 1980). A C_α representation of the structure can be generated by use of standard geometry for the α-helix–β-sheet packing. The r.m.s. deviations between the residues located in the predicted structures (88–158 residues) and the crystal structures range from 5.0 to 7.3 Å. Similar results were obtained for predictions of other pure-parallel β-sheets and some mixed sheets.

Before the combinatorial approach to structure prediction can be applied to unknown structures, two problems must be solved. First, secondary structure prediction must be improved beyond its present accuracy of *ca.* 80% (Schulz & Schirmer 1979). Secondly, a method is required to predict the segregation of the protein chain into domains. Nevertheless, the above results, which reduce the number of possible structures by at least a factor of 10^5, suggest that structure prediction might soon be feasible.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

5. Structure–function relations

We now consider whether the related functions of the glycolytic enzymes lead to common structural features. In most of the enzymes there is a binding site of the substrate and/or a cofactor that lies near the carboxyl ends of the predominant direction of the β-strands and consequently near the amino ends of the connecting α-helices (see, for example, Schulz & Schirmer 1974; Rossmann *et al.* 1975; Hol *et al.* 1978). This binding motif occurs in all the four domains with the d.n.b.f. involving NAD, ATP and pyridoxal phosphate (figure 3) and in the binding of dihydroxyacetone phosphate for TIM, phosphoenolpyruvate for PK, and in some other enzymes. This phenomenon might result from the alignment of polarized peptide units that produces an effective positive charge of about one half unit at the N-terminus that would assist the binding of negatively charged molecules, particularly the phosphate moiety (Hol *et al.* 1978). The four kinases (PGK, HK, PK, PFK) have quite different tertiary structures, but one common feature is the presence of two (or three for PK) distinct domains that might move relative to one another during catalysis in a similar fashion to the proposals for HK and PGK (Anderson *et al.* 1979; Banks *et al.* 1979). The structural resemblance of TIM to PK rather than to the other isomerase, PGI, might result from a similar functional mechanism on the triose substrates of TIM and PK (Levine *et al.* 1978).

Evolutionary significance

There are two evolutionary processes that could produce the structural and functional similarities that occur between different glycolytic enzymes (see, for example, Ohlsson *et al.* 1974; Rossmann *et al.* 1974, 1975; Rossmann & Argos 1977; Blake 1975; Phillips *et al.* 1978; Schulz & Schirmer 1979). The similarities could be the visible legacy of divergence of different enzymes or enzyme domains from a common ancestor that occurred more recently than the earliest differentiation of enzymes (see, for example, Eigen 1971). Alternatively stereochemical and functional requirements may have led to convergence to similar structures. There is little evidence for one common ancestor for all the glycolytic enzymes, as they have no greater structural and functional similarities between one another than they have with other α/β proteins that use the helix dipole to promote the binding of negatively charged groups. In contrast, the specific structural resemblance, particularly those of the domains with the d.n.b.f. and of the TIM and PK β-barrels at first sight seem remarkable and strongly suggestive of divergence (see, for example, Rossmann *et al.* 1975; Levine *et al.* 1978). However, the combinatorial approach of predicting α/β structures demonstrates (table 3) not only that there are severe restrictions on the number of allowed β-sheet topologies (see Schulz & Shirmer 1977) and possible structures but also that the detailed structural similarities expressed by r.m.s. deviations of *ca.* 3 Å can, to some extent, result simply from the common packing geometry of α/β proteins that leads to predicted structures with r.m.s. deviations of *ca.* 5 Å. Of course, the folding of the chain into the tertiary structure is dictated by and predicted from the amino acid sequence that is encoded in the genes. Thus the restrictions suggested by table 3 tend to underestimate the significance of structural similarities. On balance we consider that divergence of the d.n.b.f. and of the TIM and PK β-barrels from common ancestors is the more likely explanation for these structural similarities, but convergence remains a distinct possibility.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

We thank Sir David Phillips, Sec. R.S., for his interest, encouragement and advice. M.J.E.S. holds a Stothert Fellowship from the Royal Society, F.E.C. was a Rhodes Scholar, and W.R.T. held a Northern Ireland Postgraduate Studentship.

References (Sternberg *et al.*)

Anderson, C. M., Stenkamp, R. E. & Steitz, T. A. 1978 *J. molec. Biol.* **123**, 15–33.

Anderson, C. M., Zucker, F. H. & Steitz, T. A. 1979 *Science, N.Y.* **204**, 375–380.

Banks, R. D., Blake, C. C. F., Evans, P. R., Haser, R., Rice, D. W., Hardy, G. W., Merrett, M. & Phillips, A. W. 1979 *Nature, Lond.* **279**, 773–777.

Banner, D. W., Bloomer, A. C., Petsko, G. A., Phillips, D. C., Pogson, C. I., Wilson, I. A., Corran, P. H., Furth, A. J., Milman, J. D., Offord, R. E., Priddle, J. D. & Waley, S. G. 1975 *Nature, Lond.* **255**, 609–614.

Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, E. F., Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T. & Tasumi, M. 1977 *J. molec. Biol.* **112**, 535–542.

Biesecker, G., Harris, J. I., Thierry, J. C., Walker, J. E. & Wonacott, A. J. 1977 *Nature, Lond.* **266**, 328–333.

Blake, C. C. F. 1975 In *Essays in biochemistry*, vol. 11 (ed. P. N. Campbell & W. N. Aldridge), pp. 37–79. London: Academic Press.

Bryant, T. N., Watson, H. C. & Wendell, P. L. 1974 *Nature, Lond.* **247**, 14–17.

Burnett, R. M., Darling, G. D., Kendall, D. S., LeQuesne, M. E., Mayhew, S. G., Smith, W. W. & Ludwig, M. L. 1974 *J. biol. Chem.* **249**, 4383–4392.

Campbell, J. W., Watson, H. C. & Hodgson, G. I. 1974 *Nature, Lond.* **250**, 301–303.

Chothia, C. 1973 *J. molec. Biol.* **75**, 295–302.

Chothia, C., Levitt, M. & Richardson, D. 1977 *Proc. natn. Acad. Sci. U.S.A.* **74**, 4130–4134.

Cohen, F. E., Sternberg, M. J. E. & Taylor, W. R. 1980 *Nature, Lond.* **285**, 378–382.

Cohen, F. E., Sternberg, M. J. E. & Taylor, W. R. 1981 In preparation.

Eigen, M. 1971 *Naturwissenschaften* **58**, 465–523.

Eklund, H., Nordstrom, B., Zeppezauer, E., Soderlund, G., Ohlsson, I., Boiwe, T., Soderberg, B-O., Tapia, O. & Branden, C.-I. 1976 *J. molec. Biol.* **102**, 27–59.

Evans, P. R. & Hudson, P. J. 1979 *Nature, Lond.* **279**, 500–504.

Hol, W. G. J., van Duijnen, P. T. & Berendsen, H. J. C. 1978 *Nature, Lond.* **273**, 443–446.

Holbrook, J. J., Liljas, A., Steindel, J. & Rossmann, M. G. 1975 In *The enzymes* (ed. P. D. Boyer), 3rd edn, vol. 11, pp. 191–292. New York: Academic Press.

Honzatko, R. B., Monaco, H. L. & Lipscomb, W. N. 1979 *Proc. natn. Acad. Sci. U.S.A.* **76**, 5105–5109.

Janin, J. & Chothia, C. 1980 *J. molec. Biol.* **143**, 95–128.

Johnson, L. N., Jenkins, J. A., Wilson, K. S., Stura, E. A. & Zanotti, G. 1980 *J. molec. Biol.* **140**, 565–580.

Levine, M., Muirhead, H., Stammers, D. K. & Stuart, D. I. 1978 *Nature, Lond.* **271**, 626–630.

Levitt, M. & Chothia, C. 1976 *Nature, Lond.* **261**, 552–558.

Moras, D., Olsen, K. W., Sebesan, M. N., Buehner, M., Ford, G. C. & Rossmann, M. G. 1975 *J. biol. Chem.* **250**, 9137–9162.

Nagano, K. 1977 *J. molec. Biol.* **109**, 235–250.

Ohlsson, I., Nordstrom, B. & Branden, C.-I. 1974 *J. molec. Biol.* **89**, 339–354.

Phillips, D. C., Sternberg, M. J. E., Thornton, J. M. & Wilson, I. A. 1978 *J. molec. Biol.* **119**, 329–351.

Rao, S. T. & Rossmann, M. G. 1973 *J. molec. Biol.* **76**, 241–256.

Richardson, J. S. 1976 *Proc. natn. Acad. Sci. U.S.A.* **73**, 2619–2623.

Richardson, J. S. 1977 *Nature, Lond.* **268**, 495–500.

Richmond, T. J. & Richards, F. M. 1978 *J. molec. Biol.* **119**, 537–555.

Rossmann, M. G. & Argos, P. 1977 *J. molec. Biol.* **109**, 99–129.

Rossmann, M. G., Moras, D. & Olsen, K. W. 1974 *Nature, Lond.* **250**, 194–199.

Rossmann, M. G., Liljas, A., Branden, C.-I. & Banaszak, L. J. 1975 In *The enzymes* (ed. P. D. Boyer), 3rd edn, vol. 11, pp. 61–102. New York: Academic Press.

Schulz, G. E. & Schirmer, R. H. 1974 *Nature, Lond.* **250**, 142–144.

Schulz, G. E. & Schirmer, R. H. 1977 *Principles of protein structure*. New York: Springer-Verlag.

Shaw, P. J. & Muirhead, H. 1977 *J. molec. Biol.* **109**, 475–485.

Sprang, S. & Fletterick, P. J. 1979 *J. molec. Biol.* **131**, 523–551.

Steitz, T. A., Fletterick, R. J., Anderson, W. F. & Anderson, C. M. 1976 *J. molec. Biol.* **104**, 197–222.

Sternberg, M. J. E. & Thornton, J. M. 1976 *J. molec. Biol.* **105**, 367–382.

Sternberg, M. J. E. & Thornton, J. M. 1977*a* *J. molec. Biol.* **110**, 269–283.

Sternberg, M. J. E. & Thornton, J. M. 1977*b* *J. molec. Biol.* **110**, 285–296.

Stuart, D. I., Levine, M., Muirhead, H. & Stammers, D. K. 1979 *J. molec. Biol.* **134**, 109–142.

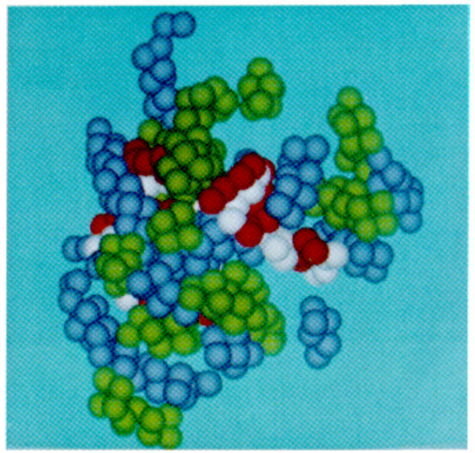
PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Discussion

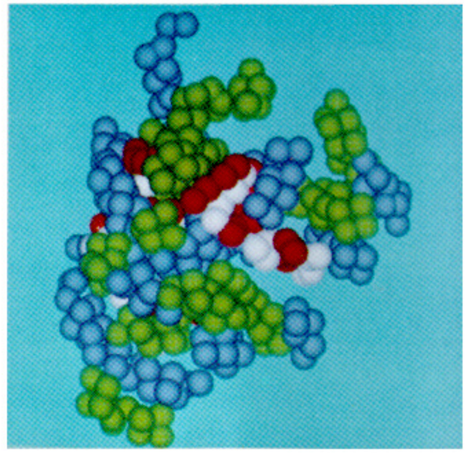
J. D. G. Smit (*Laboratorium für Biochemie, Eidgenössische Technische, Hochschule Zürich, Switzerland*). I noticed that the ranking orders for the observed structures of TIM and PK (part) are much better than those for the NAD-binding site domains in the dehydrogenases in terms of the allowable structures prediction. Is it possible that the ranking orders for TIM and PK are better than those for NAD-binding domains because TIM and PK are essentially structures with all nearest neighbouring strands (1–2–3–4...) whereas the NAD-binding domains are not? (6–5–4–1–2–3).

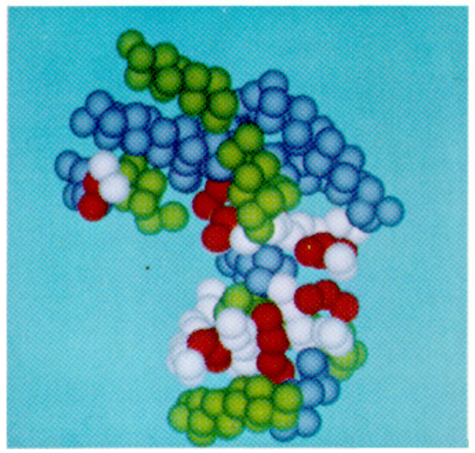
M. J. E. Sternberg. In the prediction of pure parallel β-sheets, there is no explicit consideration of adjacency that favours all strands to be adjacent as in TIM (1–2–3–4–...8) as opposed to the NAD-binding domains (6–5–4–1–2–3). The success of the TIM prediction arises from the lengths of the β-strands and the position of the non-polar residues along the β-strand. The β-barrel is a highly constrained structure and the native structure is the best structure as defined by the notions encoded in our algorithm.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

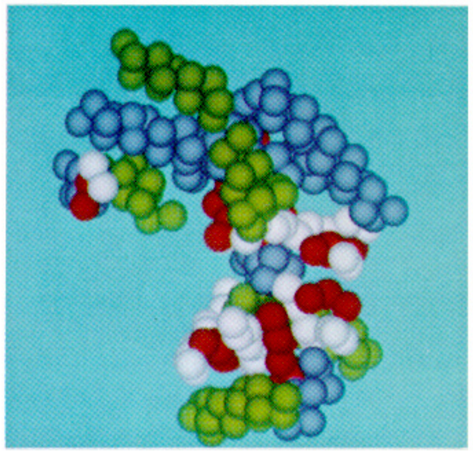


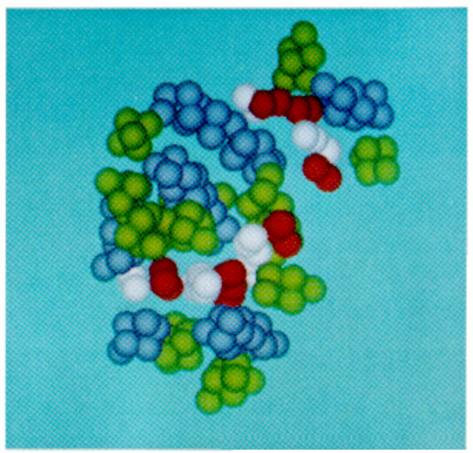
(a)



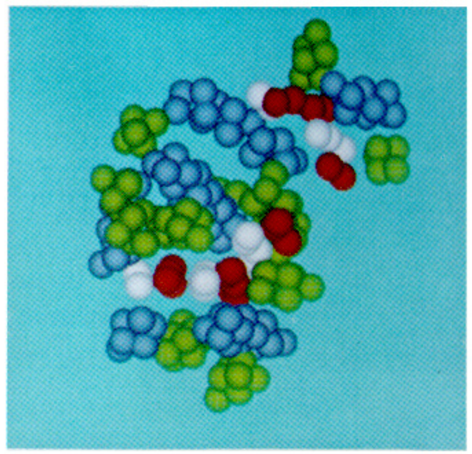


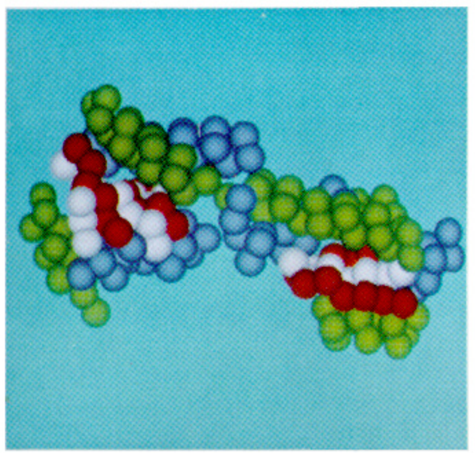
(b)



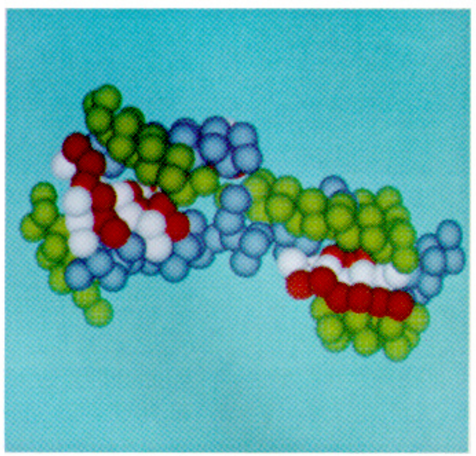


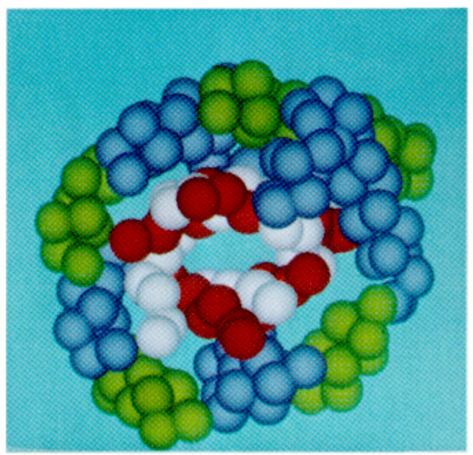
(c)





(d)





(e)

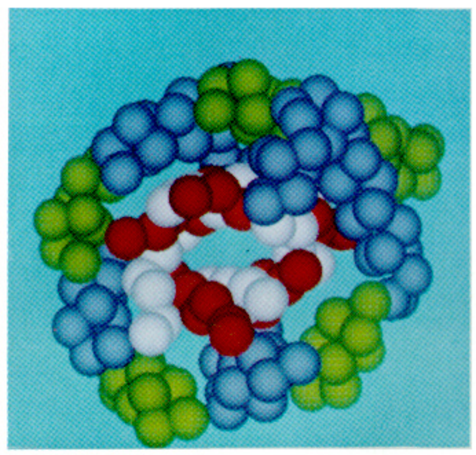
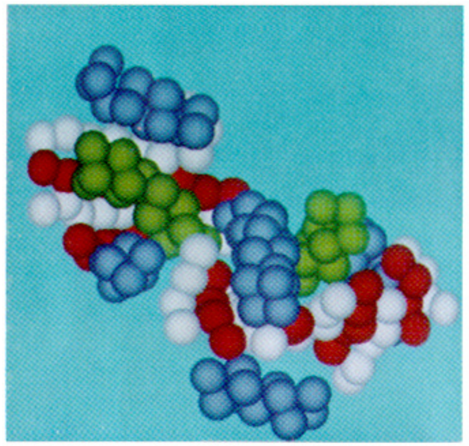


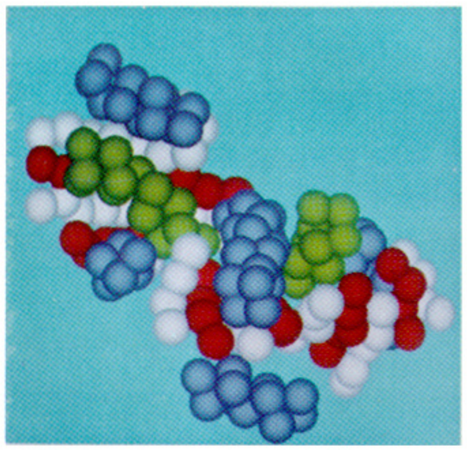
FIGURE 5. Stereo diagrams of the tertiary structures of the glycolytic enzymes, showing the relative positions of the component α-helices and β-strands. Each C_α atom in an α-helix or a β-strand is represented by a sphere *ca.* 2.2 Å in radius. Alternate β-strands in a β-sheet are shown in red and white to suggest the strand direction and the number of strands in the sheet. α-Helices are coloured green or blue, the choice of colour used simply to distinguish the different structures. The figures were produced on the computer graphics system of Richard J. Feldmann at the National Institutes of Health, Bethesda. Atomic coordinates were obtained from the Brookhaven Data Bank (Bernstein *et al.* 1977) (HK, PGI, chicken TIM, lobster GPD, LDH) or kindly provided by the individual workers (PHO, PFK, horse PGK, PGAM, PK).

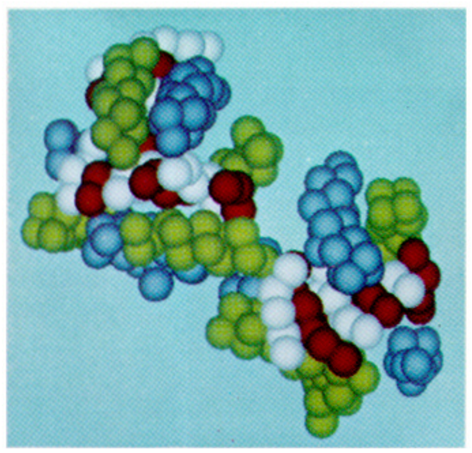
(a) Rabbit phosphorylase B; *(b)* yeast hexokinase; *(c)* pig phosphoglucose isomerase; *(d) B. stearothermophilus* phosphofructokinase; *(e)* chicken triose phosphate isomerase.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

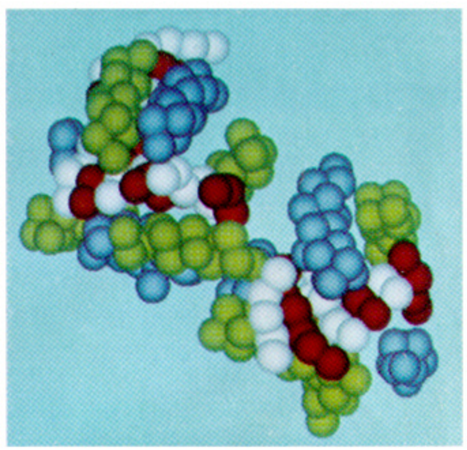


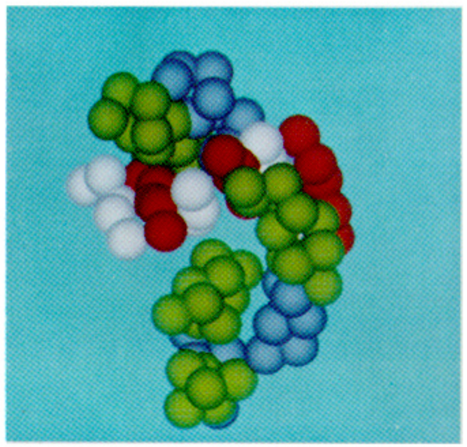
(f)



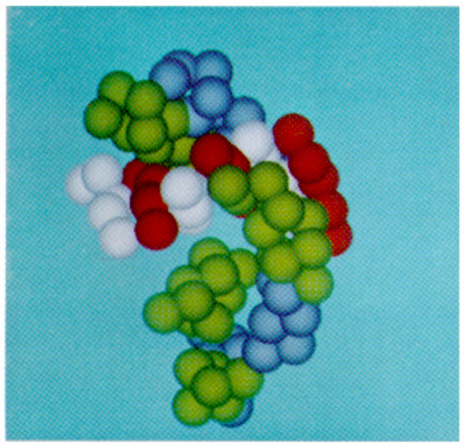


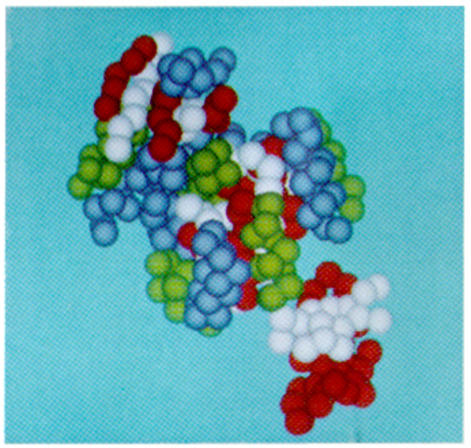
(g)



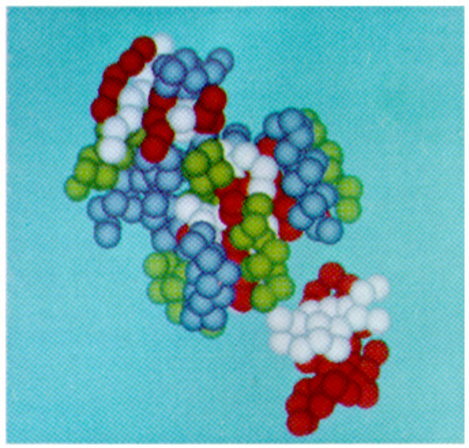


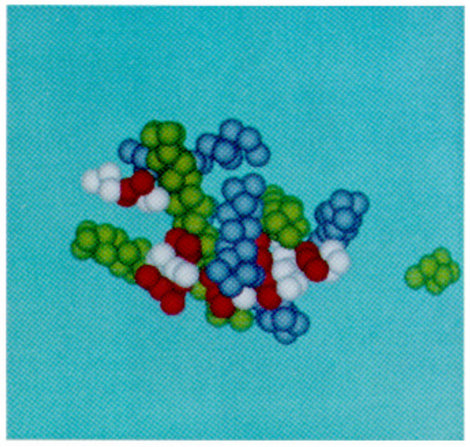
(h)





(i)





(j)

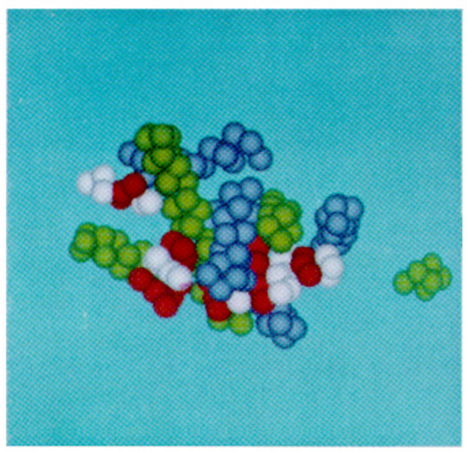


FIGURE 5 *(cont.)*. *(f)* Lobster glyceraldehyde phosphate dehydrogenase; *(g)* horse phosphoglycerate kinase; *(h)* yeast phosphoglycerate mutase; *(i)* cat pyruvate kinase; *(j)* dogfish lactate dehydrogenase.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES