The α/β hydrolase fold

Ollis, David L.; Cheah, Eong; Cygler, Miroslaw; Dijkstra, Bauke; Frolow, Felix; Franken, Sybille M.; Harel, Michal; Remington, S. James; Silman, Israel; Schrag, Joseph

Published in:
%22Protein Engineering%2C Design and Selection%22

DOI:
10.1093/protein/5.3.197

Protein Engineering vol.5 no.3 pp.197–211, 1992

The α/β hydrolase fold

David L.Ollis, Eong Cheah, Miroslaw Cygler[1], Bauke Dijkstra[2], Felix Frolow[3], Sybille M.Franken[2], Michal Harel[3], S.James Remington[4], Israel Silman, Joseph Schrag[1], Joel L.Sussman[3], Koen H.G.Verschueren[2] and Adrian Goldman[5]

Research School of Chemistry, Australian National University, Canberra, ACT 2601, Australia, [1]Biotechnology Research Institute, National Research Council of Canada, Montreal, Quebec H4P 2R2, Canada, [2]Laboratory of Chemical Physics, Department of Chemistry, University of Groningen, Nijenborgh 16, 9747 AG Groningen, The Netherlands, [3]Department of Structural Biology, Department of Neurobiology, The Weizmann Institute of Science, Rehovot 76100, Israel, and [4]Institute of Molecular Biology, and Department of Physics, University of Oregon, Eugene, OR 97403, USA

[5]Visiting Scientist, The Weizmann Institute, Permanent address: Waksman Institute, Rutgers University, New Brunswick, NJ 08855, USA

We have identified a new protein fold—the α/β hydrolase fold—that is common to several hydrolytic enzymes of widely differing phylogenetic origin and catalytic function. The core of each enzyme is similar: an α/β sheet, not barrel, of eight β-sheets connected by α-helices. These enzymes have diverged from a common ancestor so as to preserve the arrangement of the catalytic residues, not the binding site. They all have a catalytic triad, the elements of which are borne on loops which are the best-conserved structural features in the fold. Only the histidine in the nucleophile–histidine–acid catalytic triad is completely conserved, with the nucleophile and acid loops accommodating more than one type of amino acid. The unique topological and sequence arrangement of the triad residues produces a catalytic triad which is, in a sense, a mirror-image of the serine protease catalytic triad. There are now four groups of enzymes which contain catalytic triads and which are related by convergent evolution towards a stable, useful active site: the eukaryotic serine proteases, the cysteine proteases, subtilisins and the α/β hydrolase fold enzymes.

Key words: catalytic triad/evolution/hydrolases/protein structure

Introduction

Even though protein structure, function and evolution are closely related, there are very few systems which provide a clear picture of how protein structures evolve. There are two models for the evolution of similar structures. In convergent evolution, which is less common, similar structures evolve independently to perform a similar function. The best example of this in proteins is the relationship between subtilisins (including proteinase K and thermitase) and the eukaryotic serine proteases, where the enzymes have a completely different tertiary structure but a similar active site—one that is well-suited for peptide hydrolysis (Wright *et al.*, 1969; Drenth *et al.*, 1971). The more common divergent evolution is most obvious in protein families, like the globin family, where all the proteins share a common biological function and bind the same substrate. For example, the common ancestry of the globins is evident in their similar overall fold and very similar active sites, even though there is only modest sequence homology between different globins; widely varying sequences can code for the same structure (Dickerson and Geis, 1983).

Structural similarity is preserved much longer than sequence similarity. Consequently, it has been difficult to use sequence data alone to determine the evolutionary relationship between enzymes which have different functions, even though it has been possible to use such data to determine the relationship between proteins with the same function (such as the cytochrome *c*s) (Fitch and Margoliash, 1967). With enzymes in particular, the problem is more complex: they both bind a substrate and catalyze a chemical reaction, and so it is unclear *a priori* which capability would be more conserved during evolution. Some groups of enzymes might have similar binding subsites but different catalytic subsites in the active site, while others might have different binding subsites but similar catalytic subsites. In the former case, the three-dimensional fold of the enzyme presumably forms an environment especially well-suited for binding a particular type of substrate. In the latter case, the three-dimensional fold presumably forms an environment well-suited for a particular kind of catalysis. We believe that the proteins described in this paper represent the first clear example of the structural conservation of a catalytic subsite framework during the evolution of enzymes with different activities.

Two examples of binding site conservation during divergent evolution are the dehydrogenases, which have a dinucleotide fold, and α-lactalbumin and lysozyme. The dehydrogenases all have the same NAD^+ binding pocket, but bind different substrates. The substrate binding domains, which contain the catalytic machinery, have been grafted onto the conserved NAD^+ binding pocket and there is only modest sequence homology between different dehydrogenases (Rossmann *et al.*, 1975). In contrast, lysozyme and α-lactalbumin have very similar structures and sequences, even though α-lactalbumin does not show lysozyme activity. Here, the catalytic subsite has been lost, but the sugar binding subsite remains (Acharya *et al.*, 1989).

A cursory glance at enzymes which contain α/β-barrel domains leaves the impression that they are related by divergent evolution, presumably to conserve a catalytic subsite. More detailed studies, however, suggest that their evolutionary history is less clear. Farber and Petsko (1990) grouped the α/β-barrel domains into four distinct classes which, they claim, diverged from each other by circular rearrangement of the gene. Although circularly premuting the *N*-(5′-phosphoribosyl)anthranilate-isomerase gene (Luger *et al.*, 1989) yields a stable, active α/β-barrel enzyme *in vivo*, it is not clear that such a rearrangement has happened *in vivo*. Even though some members of the α/β-barrel family of enzymes clearly evolved from a common ancestor (Neidhardt *et al.*, 1990), divergent evolution may not be the only, or even the most parsimonious, explanation of the evolution of the α/β-barrel enzymes. Lesk *et al.* (1989), for instance, argue that the α/β barrel domains in enzymes can be placed in two separate classes related by convergent evolution, because the two classes show different packing arrangements. If the α/β-barrel domain was invented more than once (i.e. convergent evolution), it must

be easy to form (Marchionni and Gilbert, 1986; Rice *et al.*, 1990) and provide a convenient framework which can accommodate different catalytic activities (Brändén, 1986). Thus α/β-barrels could be related by convergent evolution towards a stable scaffold, rather than by divergent evolution from a common ancestor.

Materials and methods

The enzymes

The enzymes whose structures we compared are acetylcholinesterase (AChE) from *Torpedo californica* (Sussman *et al.*, 1991), carboxypeptidase II (CPW) from wheat—a member of a large family of serine carboxypeptidases (Breddam, 1986; Cooper and Bussey, 1989; Liao and Remington, 1990; Thomas *et al.*, 1990), dienelactone hydrolase (DLH) from *Pseudomonas* sp. B13 (Pathak *et al.*, 1988; Pathak and Ollis, 1990), haloalkane dehalogenase (HAL) from *Xanthobacter autotrophicus* (Franken *et al.*, 1991) and lipase (GLP) from *Geotrichum candidum* (Schrag *et al.*, 1991). Most of these enzymes have very different sequences, substrates (Figure 1) and physical properties (Table I). The exceptions are GLP and AChE, whose substantial sequence homology (and structural similarity)

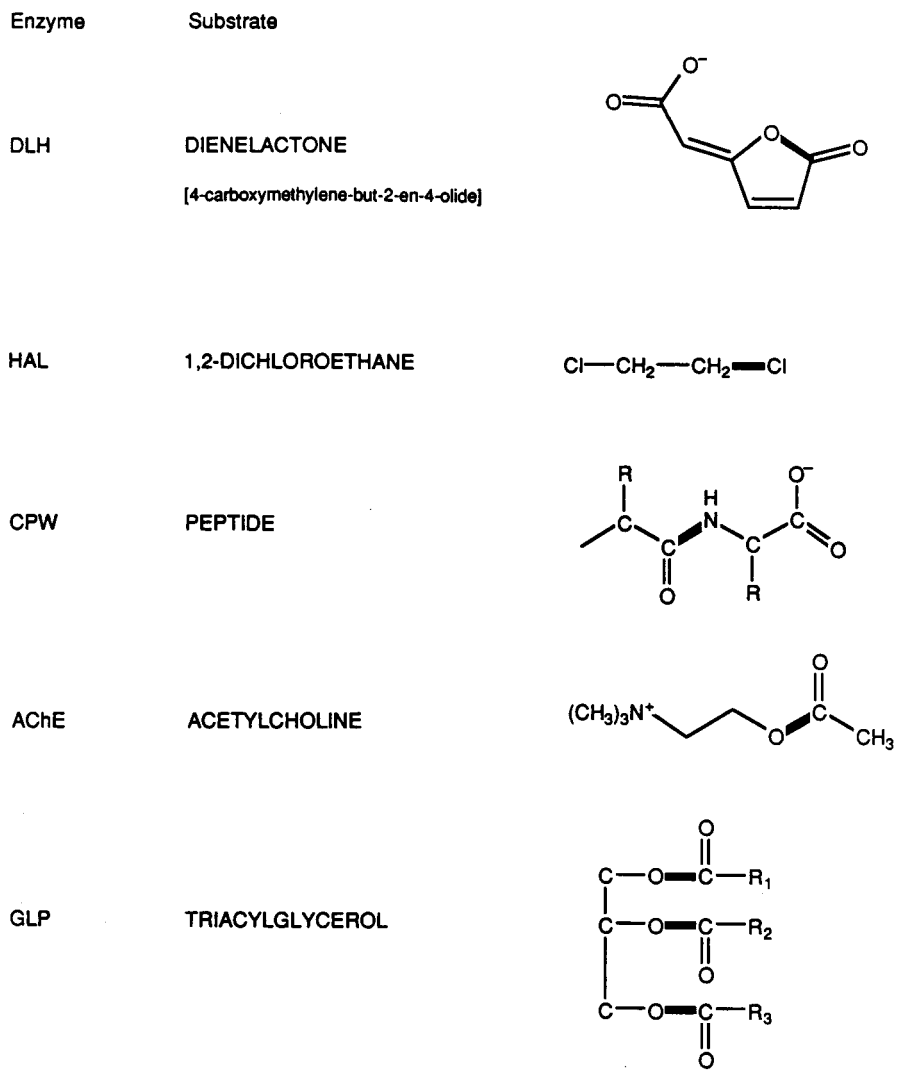

Fig. 1. The substrates of the enzymes discussed in this paper. The bonds drawn with thick lines are cleaved by the enzymes.

Table I. Selected physical and crystallographic data for the enzymes

Protein	Subunit		Oligomer structure	Modifications	Source	Refinement	
	Mol. wt (kDa)	Sequence length (amino acids)				Resolution (Å)	R-factor (%)
Acetylcholine esterase (AChE)	60	537	Dimer	Glycosylated	Fish: *Torpedo*	2.8	19
Carboxypeptidase II (CPW)	60	423	Dimer	Glycosylated	Plant: wheat	2.2	17
Dienelactone hydrolase (DLH)	25	236	Monomer	None	Bacteria: *Pseudomonas*	1.8	15
Dehalogenase (HAL)	35	310	Monomer	None	Bacteria: *Xanthobacter*	2.4	18
Lipase (GLP)	60	544	Monomer	Glycosylated	Fungus: *Geotrichum*	2.2	19

was known before their structures were determined (Shimada *et al.*, 1990; Slabas *et al.*, 1990; Schrag *et al.*, 1991). Some of these proteins are monomers while others are dimers; some are glycosylated while others are not, and their subunit sizes range from 25 to 60 kDa. None of these properties even hint that these enzymes might be related. The structures of all these enzymes have only recently been determined and they differ in their levels of refinement (Table I). All the structures are, however, sufficiently well determined to allow a detailed comparative study.

The α/β hydrolase fold

All five enzymes contain a central catalytic domain of unique topology and three dimensional structure, which we have named the 'α/β hydrolase fold' (Figure 2a). In four of the five enzymes, the topology (Richardson, 1981) of the central eight strands is identical: +1, +2, −1x, +2x, $(+1x)_3$ (Figure 3), while CPW has a β-hairpin loop inserted between the seventh and eighth strands of the sheet so that the resulting topology is: +1, +2, −1x, +2x, +1x, +1x, +3x, −1, −1 (Figure 3). All five proteins possess catalytic triads and in each case, the triad residues occur at the same topological location (Figure 3). Finally, as is typical of the active sites in α/β proteins, the nucleophile is located at the strand crossover point of the parallel β-sheet (Brändén, 1980).

The secondary structure labels of the α/β hydrolase fold (Figures 2 and 3) were chosen to emphasize the similarity in the

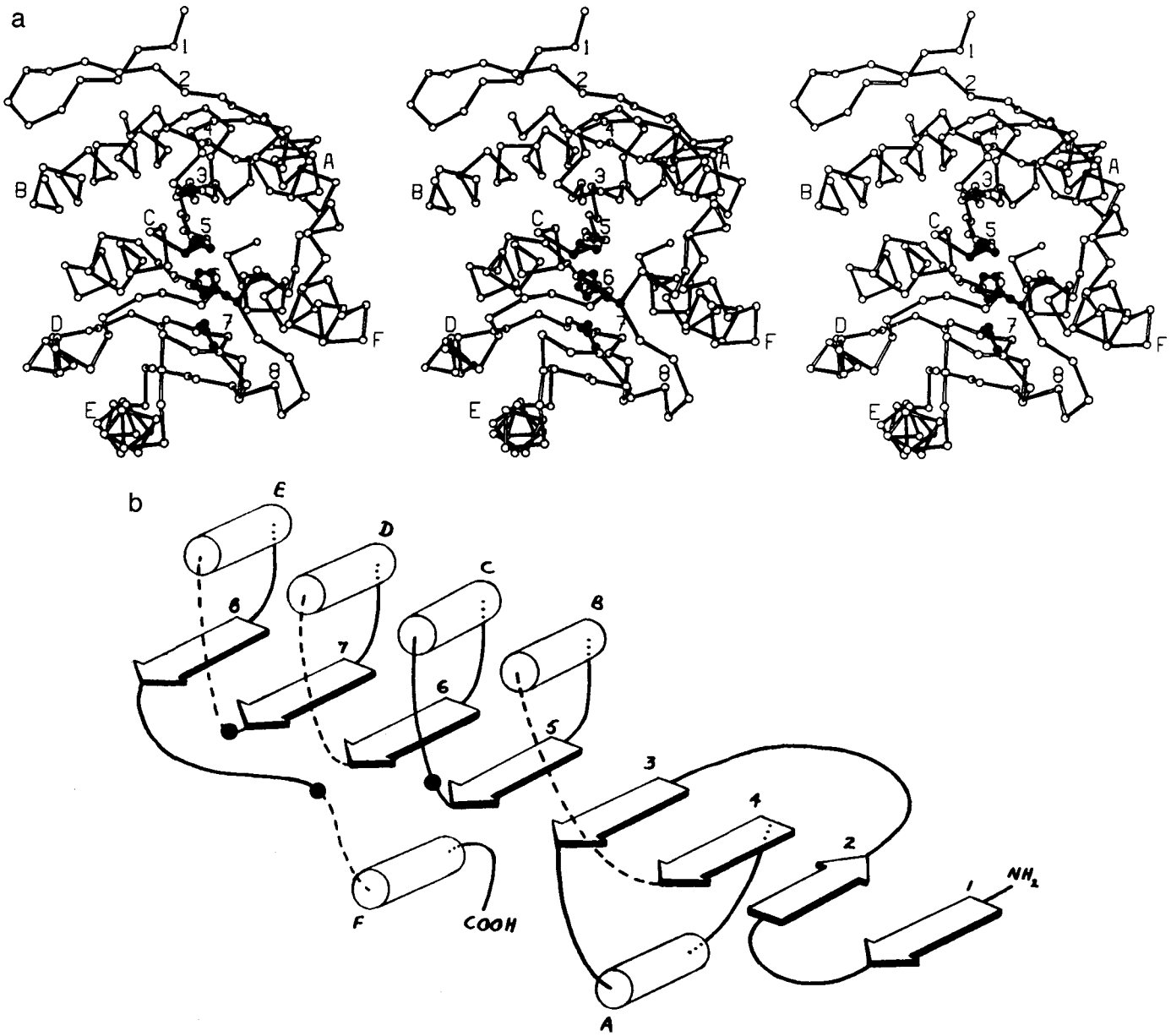

Fig. 2. (**a**) An α-carbon diagram of CPW, showing the overall fold of a hydrolase domain enzyme. The β-strands (1−8) and α-helices have been labelled A−F (B'_1 has been omitted). (**b**) A schematic diagram of the α/β hydrolase fold. The naming scheme has been chosen to emphasize the similarity between different members of the α/β hydrolase fold family. Consequently, β-strands before the start of the eight-stranded α/β hydrolase fold domain are labelled $-n. \ldots -2$, -1, and β-strands after the α/β hydrolase fold are labelled 9,10. . . The broken lines indicate places where some of the structures have excursions. The excursion between strands 3 and 4 is the A′ excursion, between strands 4 and 5, the B′ excursion, and so on. The crossover helices are part of the hydrolase domain and are called A, B, C, D, E and F. The helices in the excursions between the strands are primed (′) and numbered sequentially. Consequently, DLH has the following extra helices: B'_1 before B; HAL: D'_1, D'_2, D'_3, D'_4, D'_5 before D; CPW: D'_1, D'_2, D'_3 before D; AChE: A'_1 before A, D'_1, D'_2, D'_3 before D and E'_1, E'_2, E'_3 before E; GLP: same as AChE.

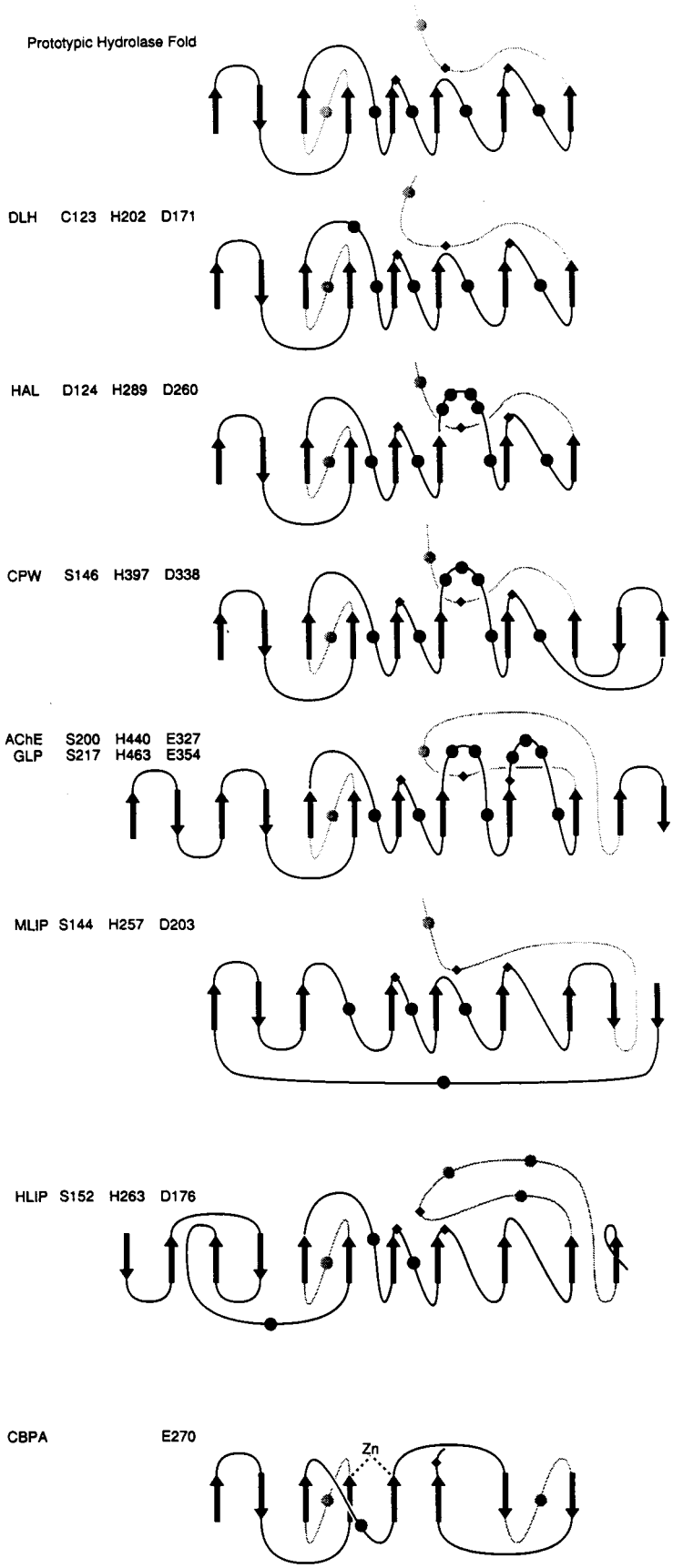

Fig. 3. The topology of the hydrolase domain and related enzymes. Arrows represent β-strands and circles represent α-helices. The helices drawn between the strands are 'crossover helices', while those drawn above the strands (after strands 2, 6, 7 and 8) are involved in substrate binding. The squares indicate the positions of triad residues, whose identities are given on the left hand side of the diagram. Proteins not already specified in the text are triacyl glycerol lipase from *R.miehei* (MLIP), triacyl glycerol lipase from human pancreas (HLIP) and carboxypeptidase A (CBPA).

secondary structures of the different members of the α/β hydrolase fold family of enzymes and are not, therefore, consistent with the earlier secondary structure labels assigned to these enzymes. The five enzymes constitute four different structural families of α/β hydrolase fold enzymes; AChE and GLP (collectively referred to as AChE/GLP) have very similar sequences (24% identity) (Shimada *et al.*, 1990; Slabas *et al.*, 1990; Schrag *et al.*, 1991), and, not surprisingly, have rather similar structures. The four groups have no sequence homology and differing amounts of structural similarity. CPW and the other α/β hydrolase fold enzymes, despite slight differences in topology, all have a similar three-dimensional arrangement of the central eight strands. Their β-strands are superhelically twisted so that the surface of the sheet covers about half a cylinder and the first and last strands cross each other at an angle of $\sim 90°$.

Table II. Automatic alignment of selected structures

Proteins compared	Number of corresponding residues		RMS distance between corresponding Cα atoms
	Total	As % of smaller protein	
DLH/CPW	160	69	2.77
DLH/AChE	157	67	2.74
DLH/HAL	146	63	3.04
DLH/GLP	154	66	2.57
AChE/GLP	399	74	1.90

The superpositions were done using the William Bennetts program SUPPOS, which uses the algorithm of Rossmann and Argos (1975), which automatically superimposes two structures.

William Bennett's program OVERLAP which uses an automated superposition procedure (Rossmann and Argos, 1975) superimposed 74% of the α-carbons of the very similar AChE and GLP, including all the major secondary structural elements and the catalytic triad, with a root mean square deviation (r.m.s.d.) per α-carbon of 1.90 Å (Table II). Automatic alignment of other pairs of α/β hydrolase fold enzymes is not as successful, but the triad residues and most of the secondary structure superimpose. The HAL/DLH superposition, with an r.m.s.d. of 3.07 Å, is the least successful. R.m.s.d.s of ~3 Å per Cα occur

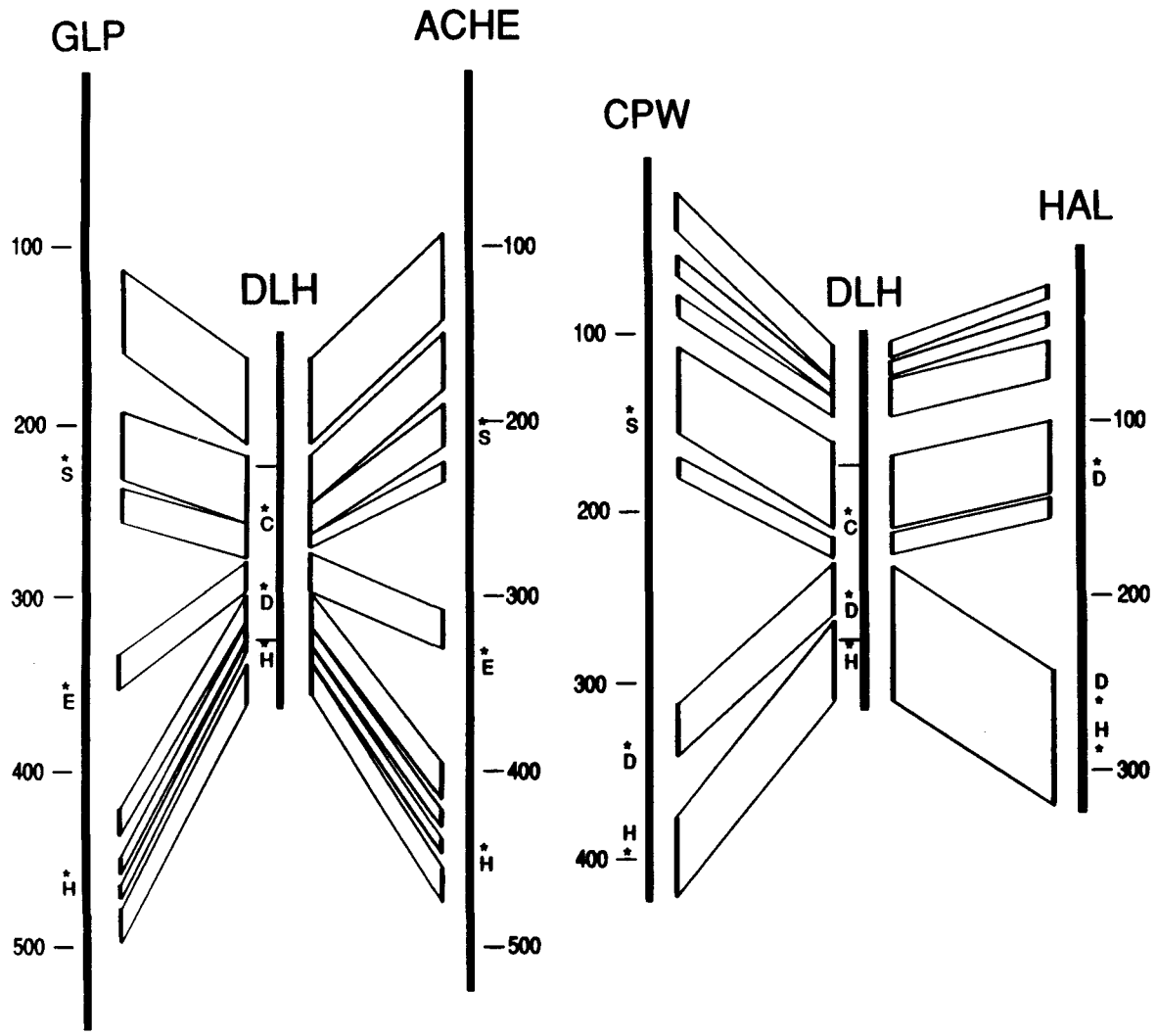

Fig. 4. Structural correspondence between different α/β hydrolase fold enzymes. The connected solid bars indicate parts of the polypeptide backbone that correspond structurally when the proteins are compared, as described in the text. It is clear that DLH is by far the smallest and simplest of the hydrolase domain enzymes. We have used Richardson's (1981) notation $(+/-)n<x>$ to describe the secondary structure of a protein in terms of its sequence. The number 'n' refers to the number of strands in the secondary structure between the current strand and the next strand in sequence, while the '+' or '−' indicates whether the strand is to the left or right of the current strand. The 'x' indicates that there is a crossover between the strands, so that they are parallel in direction (even though they may not be hydrogen bonded to each other). Thus a hairpin loop is either +1 or −1, and a $\beta\alpha\beta$ without any intervening strands is either +1x or −1x.

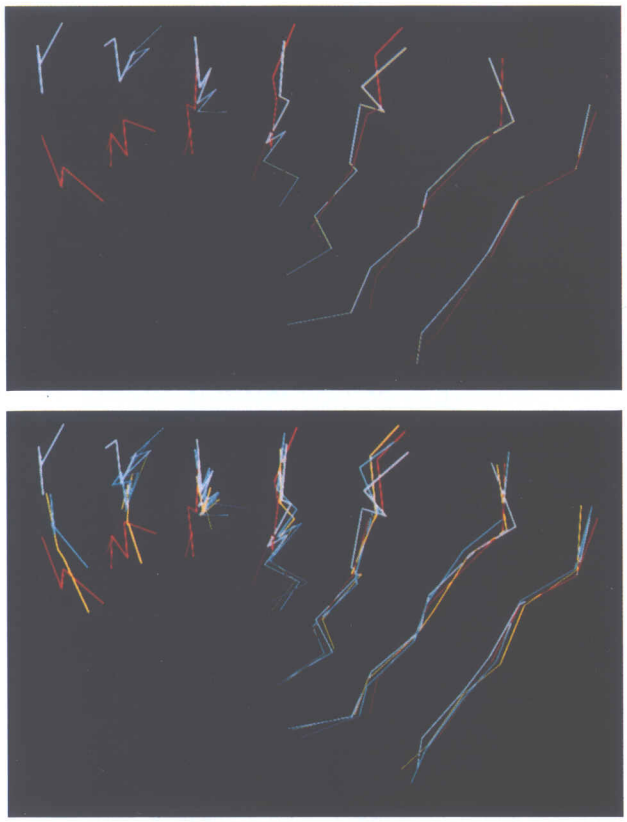

Fig. 5. (**a**) DLH (white) and HAL (red), superimposed using strands two to five. This and other alignments which required a predefined correspondence between atoms was done with program FITATOM. Visual comparison of structures was done with the program FRODO (Jones, 1978; Pflugrath *et al.*, 1984). (**b**) AChE (blue), CPW (yellow), DLH (white) and HAL (red) superimposed using strands two to five. The difference in curvature of the sheets is clearly visible.

when two structures have a similar three-dimensional architecture overall, but also have some large-scale differences. For instance, r.m.s.d.s of this order are observed when different α/β-barrel domains are compared because the barrels have different degrees of ellipticity (Goldman *et al.*, 1987). The correspondence of residues found in automated alignment are differently distributed in each of the structures (Figure 4). DLH is clearly the simplest of the enzymes; all the others have large additions in the center of the molecules (discussed later under substrate binding). AChE and GLP also have extra polypeptide at both their N- and C-termini.

The large-scale differences in the three-dimensional structures of the α/β cores of AChE/GLP, CPW, DLH and HAL are primarily due to the differences in the degree of curvature of the β-sheet (Figure 5b); the DLH sheet is the flattest and the HAL sheet, the most curved (Figure 5a). Most of the difference in curvature is due to the degree of bend between strands five and six. If strands two to five are superimposed, strands six to eight overlap poorly and vice versa (Figure 5b; Table III). In all such comparisons involving DLH or HAL, the r.m.s.d. between corresponding α-carbons is much larger when strands two to eight are used for alignment. Using the strand two to five alignment, the β-sheet of HAL can, however, be generated from that of DLH merely by rotating strands six to eight by $\sim 20°$ about strand five. The β-sheets of CPW and AChE/GLP, which all have a similar degree of β-sheet bend, lie between these two extremes, and consequently require smaller rotations about strand five to be superimposed precisely on DLH.

Table III. Alignment of secondary structural elements

Protein	Strands superimposed	R.m.s. distances between corresponding Cα in Å				
		AChE	CPW	DLH	HAL	GLP
AChE	2–5	–				
	6–8	–				
	2–8	–				
CPW	2–5	0.76	–			
	6–8	1.01	–			
	2–8	1.00	–			
DLH	2–5	0.97	0.88	–		
	6–8	0.84	1.08	–		
	2–8	1.34	1.58	–		
HAL	2–5	0.98	1.05	1.14	–	
	6–8	0.88	1.04	0.89	–	
	2–8	1.38	1.45	1.87	–	
GLP	2–5	0.38	0.74	1.00	0.97	–
	6–8	0.44	1.00	0.88	0.86	–
	2–8	0.54	1.00	1.15	1.47	–

The number of Cα atoms used in the superposition of strands 2–5 was 32, except in GLP, where there were only 31 atoms. The number of atoms used in the superposition of strands 6–8 was 19, and the number of atoms used in the comparison of strands 2–8 was 51, except in GLP, where it was 50. The assignment of corresponding residues was done by superimposing the two structures manually, and judging which residues were close to each other.

Residues used in comparisons

Protein		HAL	DLH	CPW	AChE	GLP
Strand	2	36–42	16–22	30–36	95–101	106–112
	3	48–56	29–37	45–53	110–118	123–130
	4	75–82	56–63	88–95	141–148	157–164
	5	117–124	116–123	139–146	193–200	210–217
	6	142–148	139–145	170–176	220–226	243–249
	7	250–257	161–168	328–335	317–324	344–351
	8	281–284	194–197	389–392	420–423	446–449

The three-dimensional positions of helices A, B and C are also well conserved; superimposing strands two to five superimposes the α-helices in the N-terminal parts of these enzymes. When the α-carbons in the N-terminal parts of DLH and HAL (the most different pair) are superimposed, 65 atoms ($\sim 50\%$ of the N-terminal part of DLH) align with an r.m.s.d. of 1.5 Å (Figure 6). This is surprising considering that DLH has an extra helix (B_1') before the crossover helix B while HAL has none (Figure 3). In the other enzymes, the helices in the N-terminal portions also superimpose well when β-strands two through five are aligned. In contrast, the positions of the four crossover helices (C, D, E and F) in the C-terminal half of the α/β hydrolase fold are less conserved. Although these helices are topologically equivalent in all the enzymes, they do not overlap spatially except in AChE and GLP. All the proteins except DLH have one or more large excursions on the C-terminal end of strands six, seven or eight, and so the position of the crossover helices that connect the strands varies from protein to protein. These large excursions form the binding subdomains of the α/β hydrolase fold and will be discussed later.

Results

Active site loops

The nucleophile elbow. Although the nature of the nucleophile varies (Figure 3), it is always the central residue in an extremely sharp, γ-like (Matthews, 1972) turn between strand five and helix C. The sharpness of the turn results in the nucelophile backbone

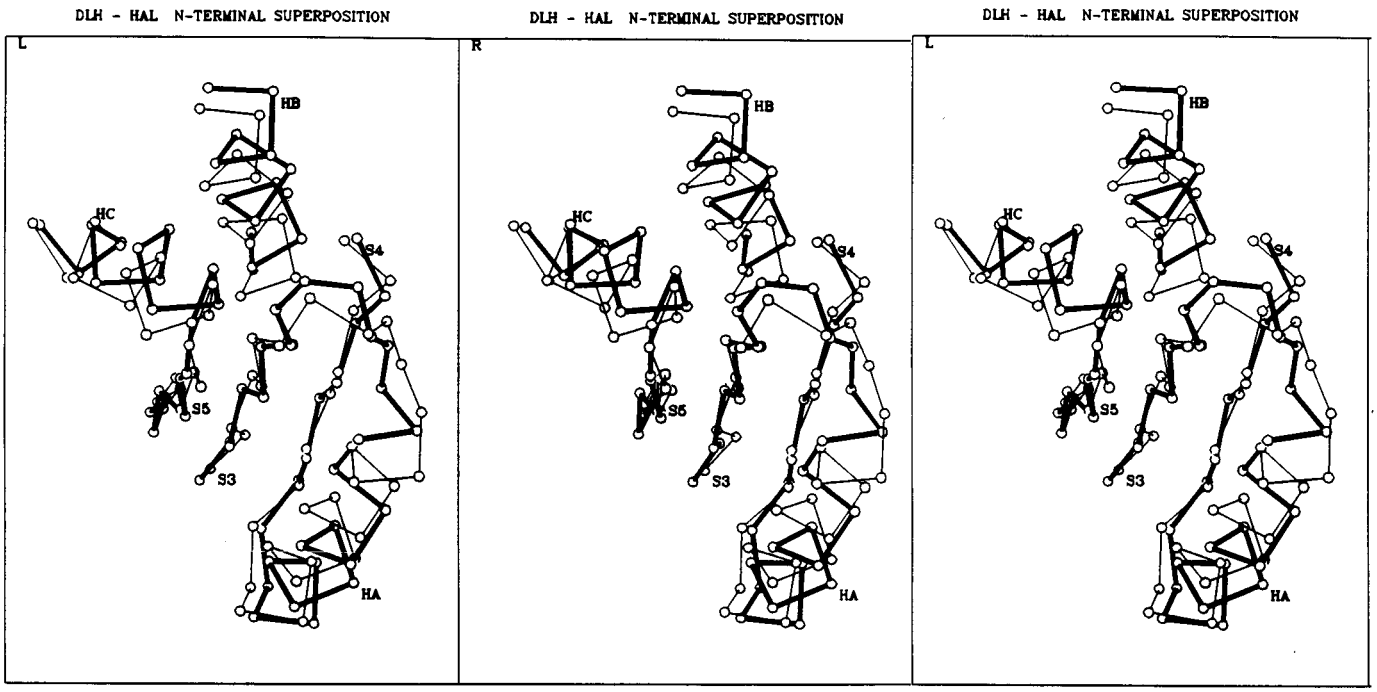

Fig. 6. A stereo Cα diagram showing the secondary structure of the N-terminal half of DLH (thick line) and HAL (thin line) after superposition. Strands two to five and helices A, B and C are shown.

phi and psi angles being in an unfavorable region of the Ramachandran plot. The strand−nucleophile−helix feature—the 'nucleophile elbow'—is the most conserved structure within the α/β hydrolase fold (Table IV; Figure 7); a search of the Brookhaven protein databank (Bernstein *et al.*, 1977) revealed that similar structures also occur on the surface of arabinose binding protein (Gilliland and Quiocho, 1981) and insulin (Blundell *et al.*, 1972). The elbow bend is quite different from the peptide surrounding the nucleophile in papain (Figure 8) although in this enzyme, as in the α/β hydrolase fold, the nucleophile is at the start of a helix.

Helix C lies unusually close to sheets four, five and six because the nucleophile is the only residue between it and strand five. The potential steric problems which might be expected to arise at the nucleophile elbow are avoided by having residues with small side-chains at four key positions (Table V). If the nucleophile is designated Nu, then the side-chains of residues Nu−2 (on strand five) and Nu+2 (on helix C) will be very close to each other (Figure 9). One, and usually both, of these residues must be glycine. The third residue which must be small is at Nu+3 (on helix C) to avoid steric overlap with sheet four. Consequently the sequence around the nucleophile must be Sm-X-Nu-X-Sm-Sm (Sm = small residue); this is different from G-X-Nu-G-G, the pattern seen in the trypsin-like proteases (Brenner, 1988). The fourth key position is the residue immediately following sheet six; its side-chain points towards helix C and, as a result, must be small.

The acid turns. The second member of the triad, which can either be Asp (DLH, HAL and CPW) (Figure 10a) or Glu (AChE/GLP), is on a loop following strand seven. While the sequence of the aspartic acid turns varies in DLH, CPW and HAL, the local structure (from the end of strand seven to two residues past the acid) is almost identical (Figures 10a and b). At the end of strand seven, two reverse turns occur in the space of seven amino acids with the acid as the joining residue: last residue of the first turn and first residue of the second turn (Figure 10b). A hydrogen bond forms between the Asp side-chain oxygen distal to the His and the backbone nitrogen of the residue two amino acids towards the C-terminus from the Asp (Figure 10b). Thus the second reverse turn stabilizes the position of the Asp side-chain, a role fulfilled by Ser214 in the serine proteases (Meyer *et al.*, 1988).

Table IV. Comparison of the nucleophile elbows in the different hydrolase fold enzymes

Protein	R.m.s. distances between corresponding Cα in Å				
	AChE	CPW	DLH	HAL	GLP
AChE	–				
CPW	1.32	–			
DLH	0.74	1.04	–		
HAL	1.42	1.47	1.23	–	
GLP	1.26	0.83	1.06	1.23	–

18 Cα atoms were used in comparison; seven before nucleophile and 11 after. The different nucleophile bends were superimposed using the program FITATOM. The assignment of corresponding residues was done by superimposing the two structures manually, and judging which residues were close to each other.

The loop bearing the Glu in AChE and GLP and the loop bearing the Asp in the other three α/β hydrolase fold enzymes are quite similar. In the AChE/GLP glutamic peptide, the Glu occupies the same position as the Asp does in DLH, CPW and HAL: it is the fourth residue in a reverse turn (Figure 10b). In both types of 'acid turn', the peptide is structurally conserved only as far as the acid and the residues which form hydrogen bonds to the acid residue side-chain. Consequently, AChE and GLP only have one reverse turn at the end of strand seven (Figure 10b). The Glu side-chain oxygen distal to His appears to form a hydrogen bond with the backbone nitrogen of the residue three amino acids before it in the sequence (Figure 10b).

The histidine loop. The structure of the peptide containing the catalytic triad histidine is highly conserved in DLH, HAL and CPW, despite the fact that the CPW topology in this region is

Fig. 7. A Cα diagram of the nucleophile elbow (see text) of the α/β hydrolase fold enzymes. Shown are AChE (green), CPW (gold), DLH (white), HAL (red), GLP (yellow) and MLIP (blue).

completely different (Figures 3 and 10a). The His peptide, which lies at the end of strand eight, comprises a turn, one amino acid and then the histidine which is the first residue in a reverse turn, but the structural conservation only extends up to the histidine. The His peptides of AChE and GLP differ from those of the other three proteins, but are similar to each other (except for a stretch of five residues at the end of strand eight) (Figure 10c). In the His peptides of AChE and GLP, extended loops bring the histidine to a position appropriate for forming the triad, and similar to that in the other proteins (Figure 10c).

Active site design

In the five α/β hydrolase fold enzymes discussed here, the catalytic triad residues always occur in the same order in the primary sequence: nucleophile, acid, histidine; this order is different from that observed in any of the other proteins that contain catalytic triads (Table VI). Furthermore, the catalytic triad residues in all the α/β hydrolase fold enzymes have astonishingly similar topological and three dimensional positions despite the lack of sequence homology (Figures 3 and 11; Table VII). A spare, simple, economical design is possible for the catalytic triad because of its topological organization (Figure 11). With the exception of the histidine in AChE/GLP, the triad residues all lie in loops close to the C-terminus of a β-strand. The parallel β-sheet has the normal left-handed twist (Richardson, 1976) found in α/β proteins, which means that very short loops beyond the ends of strands five (nucleophile), seven (acid) and eight (histidine) are sufficient to position these residues correctly to form a catalytic triad. Furthermore, there is little steric hindrance of the catalytic triad by the β-sheet.

The nucleophile and histidine of a catalytic triad must stand proud of the rest of the active site surface. This is not easy to achieve with the serine side-chain, because it is so short. Placing the nucleophile on an extremely sharp turn, as occurs in the α/β hydrolase fold, achieves this effectively and allows easy access on one side by His and on the other by substrate. The sharp turn in the peptide also optimally positions the nucleophile at the end of helix C so that the helix dipole can help stabilize both the tetrahedral intermediate in the catalytic process and the ionized form of the nucleophile.

The varying relative lengths of the strands in the β-sheet also help achieve a simple active site design. Both strands seven (the acid turns) and eight (the histidine loop) overhang strand six. If strand six were even one amino acid longer, the acid residue could not form a hydrogen bond to the histidine imidazole. The loop following strand six also must turn abruptly to avoid a close contact with the neighboring helix C (the nucleophile elbow). The shape of this bend is conserved in all five α/β hydrolase fold proteins, and the first amino acid in the bend is a small residue, usually glycine.

In the α/β hydrolase fold enzymes, the twist of the sheet imposes a 'handedness' on the catalytic triad, which is approximately a mirror-image of that seen in the serine proteases. It has previously been noted that the triad residues of chymotrypsin and papain have the opposite handedness (Garavito *et al.*, 1977). When the imidazole rings of DLH and chymotrypsin are superimposed, it can be seen that, although the position of the key atoms of the catalytic triad do not change much with respect to each other, they change substantially with respect to the rest of the protein (Figure 12). The peptides leading to the histidine and the nucleophiles approach from opposite sides of the plane defined by the imidazole rings. If the positions of key atoms in the triad are kept constant, chymotrypsin can be produced from DLH by rotating the whole molecule about the line joining serine Cβ to histidine Cγ, and the same holds true for the other members of the α/β hydrolase fold family.

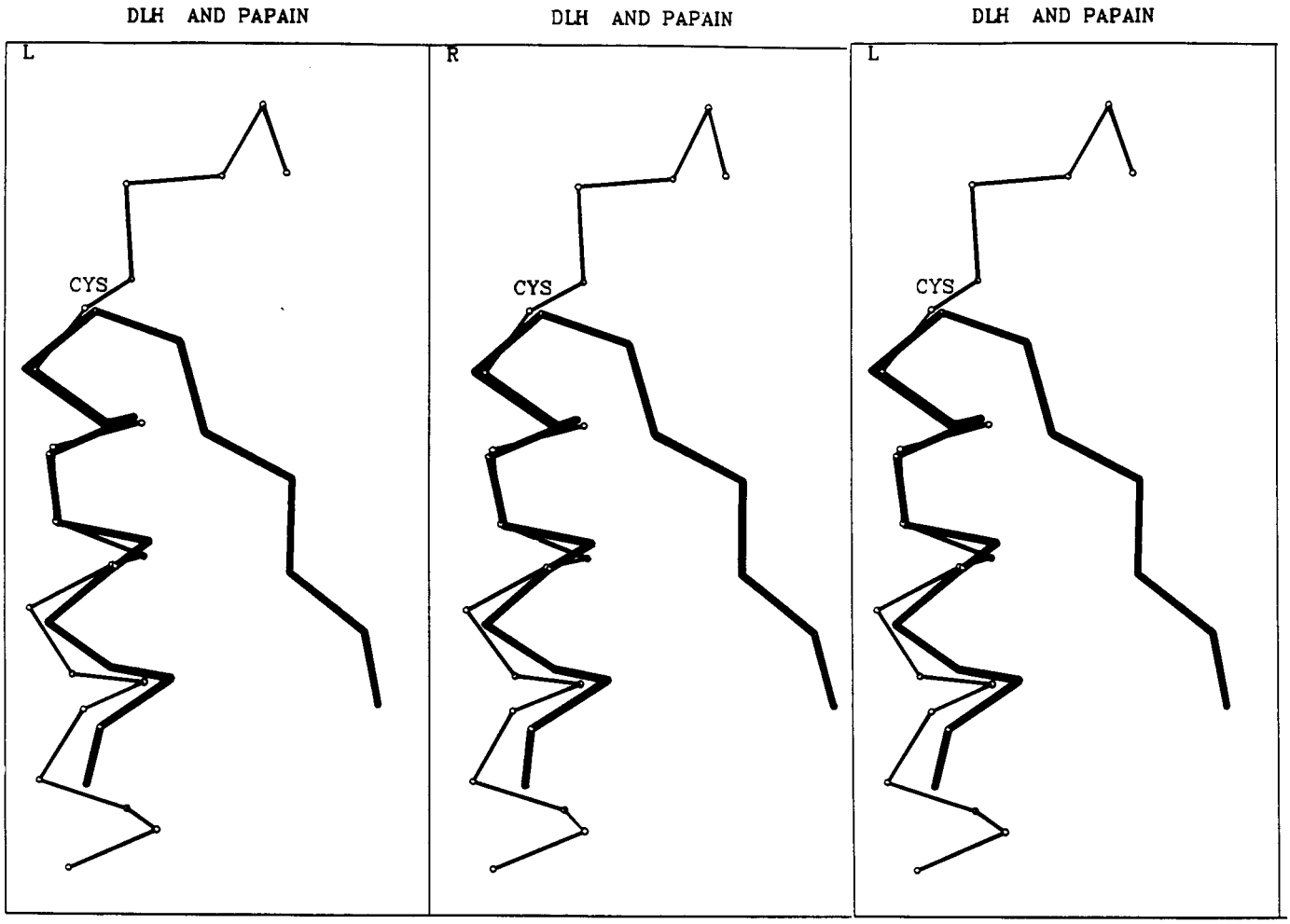

Fig. 8. A stereo Cα diagram of the nucleophile elbow of DLH (thick line) superimposed upon the peptide of papain (thin) that surrounds the catalytic triad cysteine (Cys25 in papain). Odd numbers refer to the papain residue numbering scheme; even numbers, the DLH residue numbering scheme.

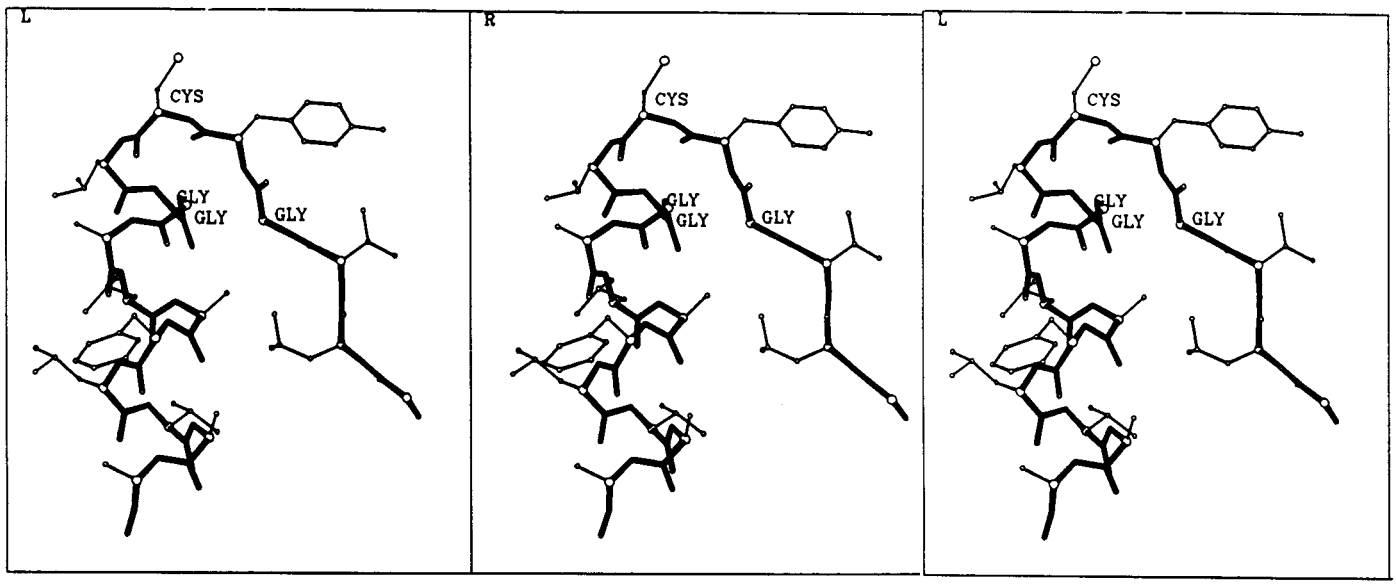

Fig. 9. A stereo diagram of the nucleophile elbow of DLH, showing the closeness of approach of 122 to 126, and 121 to 129.

Finally, there is a candidate for an oxyanion hole (Henderson, 1970) in a similar place in all five enzymes: in a turn between strand three and helix A. DLH and HAL both have a similar sharp bend at the end of strand three, while AChE, CPW and GLP share a different, somewhat longer, turn. The backbone amides of residues in this loop and of the amides of the residue

immediately following the nucleophile point into a small cavity between the loop and the nucleophile. Consequently, this cavity, like the oxyanion hole in chymotrypsin, appears to be well-designed to stabilize the tetrahdedral intermediate. This argument is supported by modelling studies and, in DLH, by inhibitor binding studies (D.L.Ollis and E.Cheah, unpublished results). The putative oxyanion hole is another example of the way in which the α/β hydrolase fold active site is a 'mirror-image' of that of the serine proteases. Thus, the role played by amide nitrogen of the active site serine in the serine proteases (Hender-

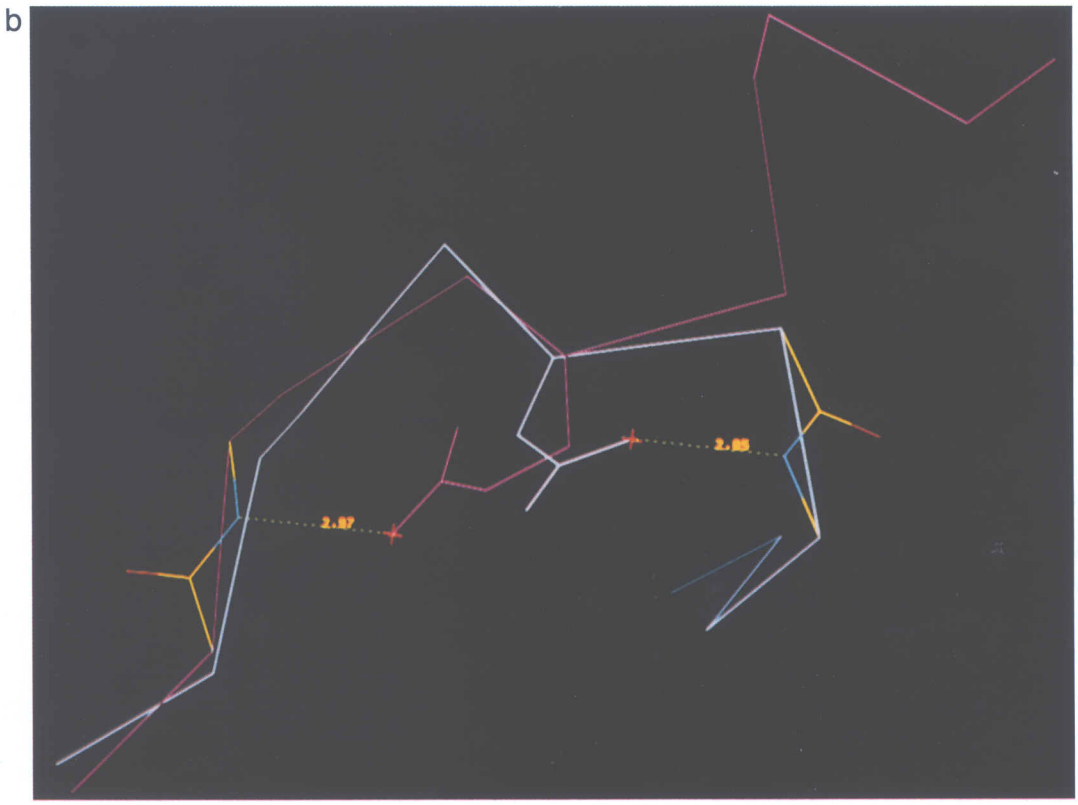

Fig. 10. (**a**) A Cα diagram showing portions of CPW (yellow), DLH (white) and HAL (red) superimposed on each other. The polypeptide backbone shown includes strand seven and the acid peptide, and strand eight and the histidine peptide. The acid and histidine residues are also shown. (**b**) A Cα diagram of the superposition of the acid loop of AChE (Glu-loop) in magenta on the acid loop of DLH (Asp-loop), in white. Dotted lines indicate the hydrogen bonds between the acid side-chains and the backbone. (**c**) A Cα diagram of the superposition of the histidine loops AChE (blue), DLH (white) and GLP (green).

son, 1970) is played by the amide nitrogen of the residue following the nucleophile (the first residue in helix C) in the α/β hydrolase fold enzymes.

The binding site excursions

The five enzymes differ in size because, besides having the α/β hydrolase fold which carries the catalytic machinery, each enzyme has different excursions of varying length which lie close to the catalytic subsite and bind substrate. The smallest excursions occur in DLH, whose substrate specificity is fully determined by helix B_1' and by an eight amino acid peptide immediately following the histidine. AChE and GLP have the largest excursions, some of which are ~100 residues long, and are the regions in which they are the least homologous in sequence. The excursions are all at the C-termini of the β-strands, as in other parallel β-sheet proteins.

Although the excursions differ in size, most occur in the second half of the β-sheet (Figure 3)—in particular between strands six and seven (excursion D′), and strands seven and eight (excursion E′). In DLH, which has the smallest α/β hydrolase fold, helix D is a bent 3_{10} helix that forms a short connection between strands six and seven. All the other proteins have D′ excursions which start from strand six, extend over the surface of the protein, and, by a variety of routes, make their way to strand seven. The crossover helices D, E and F do not overlap well, partly because there are long loops between the C-terminus of a strand and the N-terminus of the crossover helix. The long loop results in few constraints on the position of the crossover helix, and, consequently, little overlap between different α/β segments, as has been observed in other α/β structures (Rice *et al.*, 1990).

Table V. Conserved residues

Protein	Conserved residues						
	Around nucleophile						After strand 6
	−2	−1	Nu	+1	+2	+3	
AChE	**Gly**	Glu	Ser	Ala	**Gly**	**Gly**	**Gly**
CPW	**Gly**	Glu	Ser	Tyr	**Ala**	**Gly**	**Gly**
DLH	**Gly**	Tyr	Cys	Leu	**Gly**	**Gly**	**Gly**
HAL	**Val**	Gln	Asp	Trp	**Gly**	**Gly**	**Ala**
GLP	**Gly**	Glu	Ser	Ala	**Gly**	**Ala**	**Gly**

Sequence similarities among α/β hydrolase fold enzymes

Before the structures of AChE and GLP were determined, sequence comparisons suggested that the two enzymes had similar structures (Shimada *et al.*, 1990; Slabas *et al.*, 1990; Schrag *et al.*, 1991), but comparing the sequence of AChE/GLP, DLH, HAL or CPW gives no indication of similarity. Furthermore, when the structural alignments described above are used as a basis

Table VI. Order of catalytic triad residues in the primary sequence

DLH	Cys	Asp	His
HAL	Asp (Nu)	Asp (Acid)	His
CPW	Ser	Asp	His
AChE	Ser	Glu	His
GLP	Ser	Glu	His
R.miehei lipase (Brady *et al.*, 1990)	Ser	Asp	His
Human lipase (Winkler *et al.*, 1990)	Ser	Asp	His
Trypsin	His	Asp	Ser
Papain	Cys	His	Asn
Subtilisin	Asp	His	Ser

for sequence comparisons, there is no significant sequence similarity except around the nucleophile. Although it would have been impossible to predict the similarity of these enzymes based on their sequences, they are clearly members of a large class of similar proteins.

Sequence comparisons can be used to identify additional hydrolase domain proteins whose structures have not yet been determined. The sequences of AChE and GLP are similar to the C-terminal part of thyroglobulin (Schumacher *et al.*, 1986) and to other lipases. Prior studies have found a low level of sequence similarity between HAL and an epoxide hydrolase (Jansen *et al.*, 1989), and a further search of the sequence database revealed sequence similarity between HAL, 2-hydroxymuconic semialdehyde hydrolase (Nordland and Shingler, 1990) and 2-hydroxy-6-oxo-6-phenyl-hexa-2,4-dieneoate hydrolase (Kimbara *et al.*, 1989). The three enzymes are most similar over the first half of the α/β hydrolase fold (strands two to six). They all appear to have a substrate binding domain at the end of strand six and this domain is different in each enzyme. From the end of the substrate binding domain to the triad aspartate, the three proteins have similar sequences, but therafter the sequence of HAL is different. Although the two new proteins may have α/β hydrolase folds after strand seven, we cannot show that they do by sequence comparisons with the proteins discussed in this paper.

Structural relationships betwen the α/β hydrolase fold and other proteins

Two recently determined lipases have structures similar to the α/β hydrolase fold and may, indeed, be members of the α/β hydrolase fold family of enzymes. *Rhizomucor miehei* lipase (Brady *et al.*, 1990) has a somewhat different topology to that of the α/β hydrolase fold enzymes described here: ignoring strand one of its nine-stranded sheet, it is $+1$, $+1$, $(+1x)_4$, $+1$, while the α/β hydrolase fold is $+1$, $+2$, $-1x$, $+2x$, $(+1x)_3$ (Figure 3). It is not immediately clear how one topology could easily be converted to the other. However, its catalytic triad is on the same strands as the α/β hydrolase fold's: Ser after strand five, Asp after strand seven, and His after strand eight. This protein has a nucleophile elbow (Figure 7) and the catalytic triad has the same hand as that of the α/β hydrolase fold enzymes (Brady *et al.*, 1990).

The topology of the first domain of human pancreatic lipase (HPL) is even more like that of the α/β hydrolase fold enzymes: ignoring the first two strands, it is -1, $+3$, $-1x$, $+2x$, $(+1x)_3$, $+1x$, which differs from the α/β hydrolase fold only in the connection of the first two strands (Figure 3) (Winkler *et al.*, 1990). However, the catalytic triad residues of HPL are organized differently than in the canonical α/β hydrolase fold presented here. In HPL, the active site Ser and His are in the expected places: after strands five and eight respectively, but the

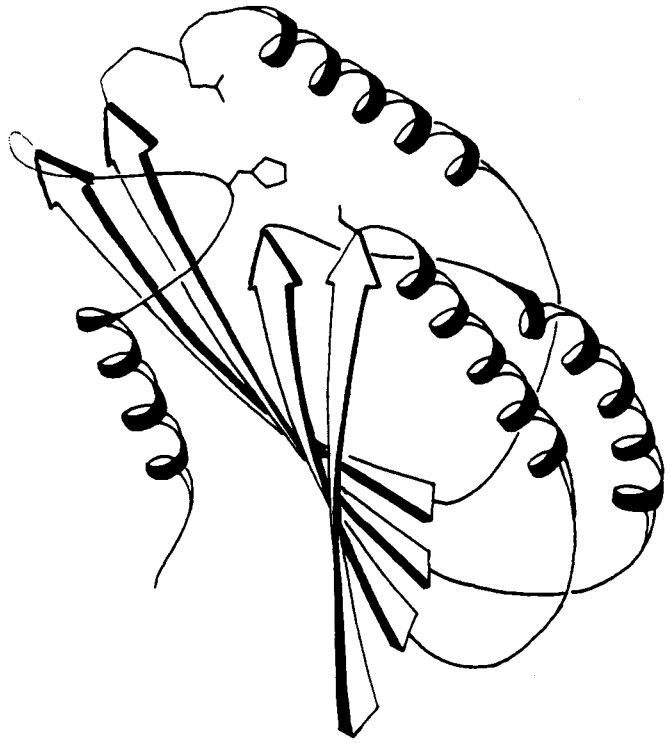

Fig. 11. A schematic of the C-terminal half of DLH. The side-chains of the triad residues are shown. Strands are represented by arrows and helices by coils. Strands five (closest) through eight are shown along with helices C–F. The other four proteins have large substrate binding loops in this region.

Table VII. Superpositions of the active site triad peptides

Protein	R.m.s. distances between corresponding Cα in Å				
	AChE	CPW	DLH	HAL	GLP
AChE	–				
CPW	0.81	–			
DLH	0.90	0.91	–		
HAL	0.81	0.80	0.94	–	
GLP	0.42	0.94	0.75	0.84	–

Nine Cα atoms were used in each comparison: three amino acid residues centered around each member of the catalytic triad (i.e. each of the catalytic triad residues, and one on either side of that residue). The different enzymes were superimposed using the program FITATOM. The assignment of corresponding residues was done by superimposing the two structures manually, and judging which residues were close to each other.

Asp follows strand six, not seven. The structures of the five α/β hydrolase fold proteins suggest how this could happen. In them, the end of strand six is close to the triad acid turn; if the turn were moved, strand six could be extended to put a new acid where the old acid (from strand seven) was. Although the triad acid in HPL does not emanate from near the end of strand seven, there is an acid residue in that position which conceivably may have formed part of the triad in ancestral proteins. Further comparative studies will be needed to understand the relationship between these lipases and the α/β hydrolase fold enzymes.

Carboxypeptidase A (CBPA) (Rees *et al.*, 1983) also shows some similarity to the enzymes of the α/β hydrolase fold. The topology of the mixed sheet in CBPA is $+1$, $+2$, $-1x$, $+2x$, $+2$, $+1x$, -2 (Figure 3), so the topology of the first five strands is the same as the α/β hydrolase fold enzymes, but the strands

Table VIII. Sequence comparisons of HAL, SEMI and ENOA

```
      1                         *                  *  .  *  *       50
ENOA  .......... .MSELNESST SKFVTINEKG LSNFRIHLND AGQGERVIML
SEMI  .......... ....MNAPQN SPEIG.REII AAGIRTNLHD SGAGFPLMMI
 HAL  MINAIRTPDQ RFSNLDQYPF SPNYLDDLPG YPGLRAHYLD EGNSDAEDVF

     51                              **  *  *  **  **          100
ENOA  HGGGPGAGGW SNYYRNIGPF VEAGYRVLLP DAPGFNKSDT VVMDEQRGLV
SEMI  HGSGPGVTAW AN.WRLVMPE LAKSRRVIAP DMLGFGYSER PADAQYNRDV
 HAL  LCLHGEPTWS YLYRKMIPVF AESGARVIAP DFFGFGKSDK PVDEEDYTFE

    101           *      *  *    **    *       * *    **  *150
ENOA  NARS.VKGMM DVLGIEKAHL VGNSMGGAGA LNFALEYPER TGKLILMGPG
SEMI  WVDH.AVGVL DALEIEQADL VGNSFGGGIA LALAIRHPER VRRLVLMGSA
 HAL  FHRNFLLALI ERLDLRNITL VVQDWGGFLG LTLPMADPSR FKRLIIMNAC
                                  ↑
    151                     *  *           **              200
ENOA  GLGNSL.... ...FTAMPME GIKLLFKLYA EPSLETLKQM LNVFL.....
SEMI  GVSFPI.... .........TE GLDAVWGY.. NPSFAEMRRL LDIFA.....
 HAL  LMTDPVTQPA FSAFVTQPAD GFTAWKYDLV TPSDLRLDQF MKRWAPTLTE

    201*                  *                 *               250
ENOA  FDQSVITDEL LQGRW.ANIQ RNPEHLKNFI LSAQKVPLSA WDVSARLG..
SEMI  FDRNLVNDEL AELRYQASIR PGFHESFAAM FPAPRQRWVD GLASAEAAIR
 HAL  AEASAYAAPF PDTSYQAGVR KFPK...... MVAQRDQACI DISTEAISFW

    251       *     *  *                                    300
ENOA  ..EIKAKTLV TWGRDDRFVP LDHGLKLIAN MQDAHVHV.F PRCA..IGRS
SEMI  ..ALPHETLV IHGREDQIIP LQTSLTLADW IARAQLHV.F GQCGHWTQIE
 HAL  QNDWNGQTFM AIGMKDKLLG PDVMYPMKAL INGCPEPLEI ADAGHFVQ.E
                     ↑                                  ↑
    301                        322
ENOA  GSTRTPSTG. .......... ..
SEMI  HAARFASLVG DFLAEADAAA IS
 HAL  FGEQVAREAL KHFAETE... ..
```

A sequence alignment of 2-hydroxy-6-oxo-6-phenyl hex-2,4-dienoate hydrolase (ENOA) from *Pseudomonas* spp. (Kimbara *et al.*, 1989), 2-hydroxy-muconosemialdehyde hydrolase (SEMI) from *Pseudomonas putida* (Nordland and Shingler, 1990) and HAL, using the program PILEUP in the GCG sequence analysis package (Devereux *et al.*, 1984). Identical residues are marked in bold; similar (Asp = Glu; Lys = Arg) residues, with an underline. Positions in the sequence where all three enzymes are either the same or similar are marked with an asterisk above the sequence. There are more such residues, and fewer insertions and deletions in the N-terminal part of these enzymes than anywhere else. Residues 1–150 (β1–6) are more highly conserved than both the substrate binding subdomain (151–240; $\alpha D_1'-D_5'$) and the last two strands (251–310). The catalytic triad residues in HAL are marked under the sequence by an arrow (↑); the nucleophile Asp in HAL is replaced with Ser in ENOA and SEMI, and the acid Asp is conserved in all three proteins. There is no clear candidate for the histidine in ENOA and SEMI; although there is a His in SEMI at the position of the histidine in HAL, the sequence at that point appears to have been deleted in ENOA.

in its sheet all tend to be much shorter. Nevertheless, certain features of the active site of CBPA are similar to the α/β hydrolase fold. The metal ion is bound by residues located on loops at the C-termini of strands three and five; equivalent loops in the α/β hydrolase fold form the oxyanion pocket and hold the nucleophile.

Discussion

Evolutionary history of the hydrolases

Divergent evolution of the α/β hydrolase fold. The facts presented above clearly suggest that HAL, DLH, CPW, AChE and GLP have diverged from a common ancestor and that they have evolved so as to preserve the positions of key catalytic components. This is in contrast to lysozyme and α-lactalbumin, which evolved to preserve a common binding site and also in contrast to the α/β barrel proteins where it is unclear if they are all related by divergent evolution.

The enzymes share a striking similarity in their central catalytic domain. They have similar overall topology, a conserved sequence order for the catalytic triad residues, and conserved loops for the catalytic triad and oxyanion hole. They also have similar three-dimensional structures. The structural and mechanistic conservation is all the more remarkable, given the lack of sequence similarity between the proteins. Secondly, as expected of enzymes where divergent evolution has conserved an active site, the structure of the peptides around the active site

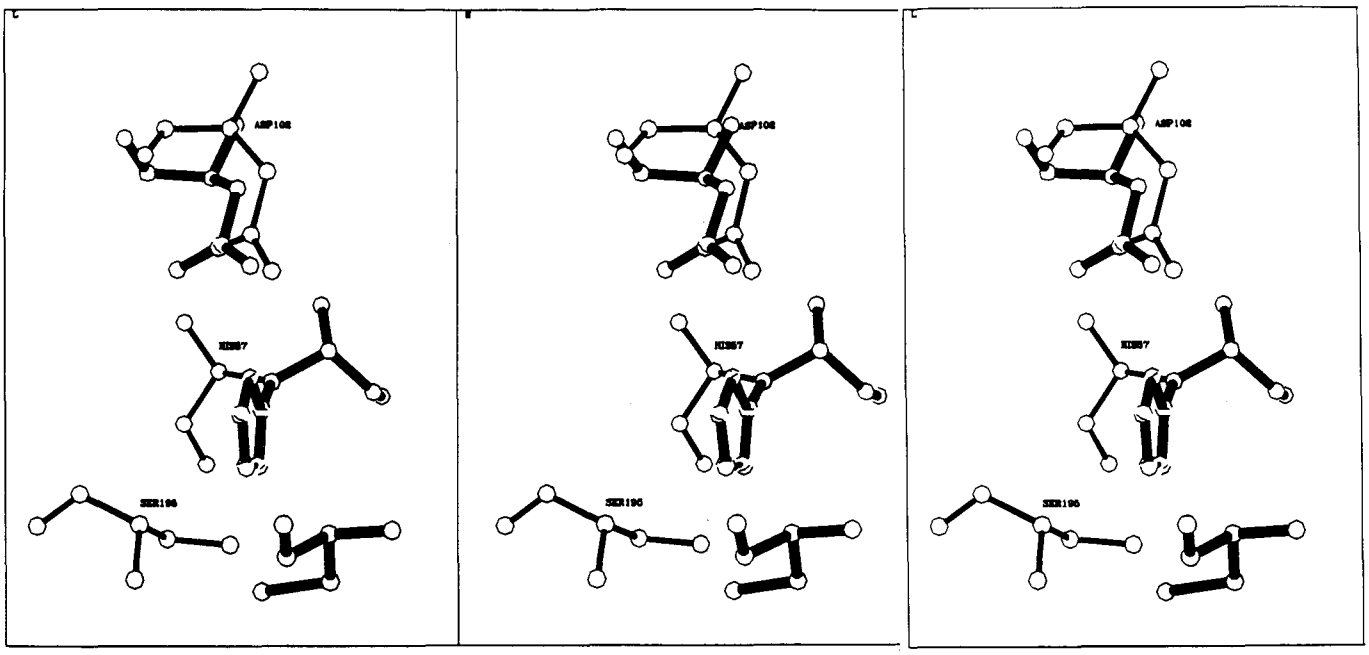

Fig. 12. A stereo diagram of the catalytic triads of DLH (thick line) and chymotrypsin (thin line). The imidazole rings of the two histidines have been superimposed. It can be seen that the position of the oxyanion hole is different in the two enzymes.

is highly conserved, and those forming the scaffolding for the catalytic triad are the most conserved. The α/β hydrolase fold also provides an effective catalytic framework from which different binding pockets can be arranged (Brändén, 1980, 1986).

The α/β hydrolase fold has been preserved because it is a simple, stable and effective way of building a variety of different catalytic triads which can cause a hydrolysis reaction. The α/β hydrolase fold enzymes catalyze a wider variety of hydrolysis reactions than any other class of catalytic triad enzymes (Figure 1). They also show more variation in the identity of the members in the catalytic triad; the only conserved residue is the histidine. The loops that contain the catalytic triad residues are extremely well-conserved structurally, and, it appears, can accommodate various different residues at the nucleophile and acid positions. The acid can either be glutamate (AChE, GLP) or aspartate (CPW, DLH, HAL); these are the first instances where glutamate has been observed in a catalytic triad. The nucleophile can either be Ser, Cys or Asp, and it is possible to replace the active site Cys of DLH by Ser and retain a significant level of wild-type catalytic activity (Pathak *et al.*, 1991). In contrast, the serine proteases require Ser, not Cys, in the catalytic triad (Higaki *et al.*, 1989) and papain Cys, not Ser (Andrew Storer, personal communication; Clark and Lowe, 1978) to retain a significant level of activity. The α/β hydrolase fold obviously provides a convenient general-purpose framework on which different catalytic triads can be arranged without losing catalytic efficiency.

Despite the varying residues in the catalytic triad, all the enzymes except HAL have substrates (amides or esters) and mechanisms similar to that of the serine proteases. In AChE, CPW, DLH and GLP, the reaction occurs in two steps, with a covalent intermediate formed between the nucleophile and the substrate. Furthermore, in the case of AChE, CPW and DLH, there is evidence that a covalent intermediate forms (Douglas *et al.*, 1976; Rees *et al.*, 1983; Quinn, 1987), presumably via a tetrahedral intermediate stabilized by the oxyanion hole of strand three and helix C (see above). The covalent intermediate presumably then breaks down, as in the serine proteases, by general base catalysis.

HAL, by way of contrast, hydrolyzes halogenated alkanes and has an Asp for the nucleophile; the first step of its hydrolysis mechanism is unusual. It has been proposed that this first step involves an S_N2 displacement of the halogen atom of the haloalkane by an oxygen of the Asp. The transition state contains a penta-coordinated carbon atom, as in other S_N2 reactions, which here appears to be stabilized by two tryptophans. This S_N2 displacement generates a covalent enzyme–ester intermediate like that found in the other hydrolase enzymes. The second half of the hydrolysis reaction then proceeds as in the other hydrolases.

Convergent evolution and other hydrolases. There is, it should be emphasized, no global similarity between the α/β hydrolase fold enzymes and the other enzymes that have catalytic triads: the serine proteases, subtilisins and the cysteine proteases. The only similarity between the α/β hydrolase fold enzymes and the other hydrolases is in the organization of the atoms of the catalytic triad (Ser Oγ, acid carboxylate group, His imidazole ring) and, consequently, the mechanism of the enzyme. The overall fold of these enzymes, the order of the catalytic triad residues in the sequence, the placement of residues contributing to the oxyanion hole, and even the direction of the Cα–Cβ bond of the nucleophile are different. Consequently, the relationship of the active site to the rest of the enzyme in the α/β hydrolase fold is, as previously described, qualitatively different from the organization of the active sites found before.

It is remarkable that there are now four different examples of the catalytic triad which are related by convergent, not divergent evolution: the serine proteases, subtilisin, papain and the α/β hydrolase fold. We believe that this reflects the primordial nature of hydrolysis. It may have been one of the very first reactions for which a protein enzyme evolved, and consequently several different stuctural solutions were found. It also reflects how central hydrolysis is to biochemical pathways, and how few solutions are possible, at the level of individual amino acid side-chain chemistry, to the problem of hydrolyzing esters and amides. The α/β hydrolase fold appears to be a very effective solution

to the problem of constructing a skeleton on which to hang a catalytic triad. AChE, CPW, DLH, HAL and GLP contain more variation in catalytic triad residues than had previously been seen.

We have described a new family of enzymes, whose common feature, the α/β hydrolase fold, provides the scaffolding for the catalytic triad involved in its enzyme activity. This family provides the first clear example of divergent evolution of catalytic sites. We expect that additional enzymes, whose sequence will have given no clue that they belong to this family, will turn out to belong to it when their three-dimensional structure is elucidated.

Acknowledgements

We thank the Molecular Biology Computing Laboratory at the Waksman Institute, Rutgers University for the use of their facilities. This project was supported by NATO Grant CRG 900670, NSF Grant 89-04341, NSF Grant DMB-8817438, US Army Medical Research and Development Command Contract No. DAMD17-89-C-9063, the Minerva Foundation, the Revson Foundation, the United States–Israel Binational Science Foundation (BSF), Jerusalem, the Kimmelman Center for Biomolecular Structure and Assembly, Rehovot, and the Charles and Johanna Busch Memorial Fund. Adrian Goldman is a Markey fellow, and this work was supported in part by the Lucille P.Markey Charitable Trust.

References

Acharya,K.R., Stuart,D.I., Walker,N.P.C., Lewis,M. and Phillips,D.C. (1989) *J. Mol. Biol.*, **208**, 99–127.

Bernstein,F.C., Koetzle,T.F., Williams,G.J.B., Meyer,E.F., Brice,M.D., Rodgers,J.R., Kennard,O., Simanouchi,T. and Tasumi,M. (1977) *J. Mol. Biol.*, **112**, 535–542.

Blundell,T.L., Hodgkin,D.C., Dodson,G.G. and Mercola,D.A. (1972) *Adv. Prot. Chem.*, **26**, 279–402.

Brady,L., Brzozowski,A., Derewenda,Z.S., Dodson,E., Dodson,G., Tolley,S., Turkenburg,J.P., Christiansen,L., Huge-Jensen,B., Norskov,L., Thim,L. and Menge,U. (1990) *Nature*, **343**, 767–770.

Brändén,C.-I. (1980) *Q. Rev. Biophys.*, **13**, 317–338.

Brändén,C.-I. (1986) In Fletterick,R. and Zoller,M. (eds), *Computer Graphics and Molecular Modelling*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp. 45–51.

Breddam,K. (1986) *Carlsberg Res. Commun.*, **51**, 83–128.

Brenner,S. (1988) *Nature*, **334**, 528–530.

Clark,P.I. and Lowe,G. (1978) *Eur. J. Biochem.*, **84**, 293–299.

Cooper,A. and Bussey,H. (1989) *Mol. Cell Biol.*, **9**, 2706–2714.

Devereux,J., Haeberli,P. and Smithies,O. (1984) *Nucleic Acids Res.*, **12**, 387.

Dickerson,R.E. and Geis,I. (1983) *Hemoglobin: Structure, Function, Evolution and Pathology*. The Benjamin Cummings Publishing Company, Menlo Park, CA, pp. 65–112.

Douglas,K.T., Nakagawa,Y. and Kaiser,E.T. (1976) *J. Am. Chem. Soc.*, **98**, 8231–8236.

Drenth,J., Hol,W.G.J., Jansonius,J.N. and Koekoek,R. (1971) *Cold Spring Harbor Symp. Quant. Biol.*, **36**, 107–116.

Farber,G.K. and Petsko,G.A. (1990) *Trends Biochem. Sci.*, **15**, 228–234.

Fitch,W.M. and Margoliash,E. (1967) *Science*, **155**, 279–284.

Franken,S.M., Rozeboom,H.J., Kalk,K.H. and Dijkstra,B.W. (1991) *EMBO J.*, **10**, 1297–1302.

Garavito,R.M., Rossmann,M.G., Argos,P. and Eventoff,W. (1977) *Biochemistry*, **16**, 5065–5070.

Gilliland,G.L. and Quiocho,F.A. (1981) *J. Mol. Biol.*, **146**, 341–362.

Goldman,A., Ollis,D. and Steitz,T.A. (1987) *J. Mol. Biol.*, **194**, 143–153.

Henderson,R. (1970) *J. Mol. Biol.*, **54**, 341–346.

Higaki,J.N., Evnin,L.B. and Craik,C.S. (1989) *Biochemistry*, **28**, 9256–9263.

Jansen,D.B., Pries,F., van der Ploeg,J., Kazemier,B., Terpstra,P. and Witholt,B. (1989) *J. Bacteriol.*, **171**, 6791–6799.

Jones,T.A. (1978) *J. Appl. Crystallogr.*, **11**, 614–618.

Kimbara,K., Hashimoto,T., Fukuda,M., Koana,T., Takagi,M., Oishi,M. and Yano,K. (1989) *J. Bacteriol.*, **171**, 2740–2747.

Lesk,A., Brändén,C.-I. and Chothia,C. (1989) *Proteins: Struct., Funct. Genet.*, **5**, 139–148.

Liao,D.-I. and Remington,S.J. (1990) *J. Biol. Chem.*, **265**, 6528–6531.

Luger,K., Hommel,H., Herold,M., Hofsteenge,J. and Kirschner,K. (1989) *Science*, **243**, 206–210.

Marchionni,M. and Gilbert,W. (1986) *Cell*, **46**, 133–141.

Matthews,B.W. (1972) *Macromolecules*, **5**, 818–819.

Meyer,E., Cole,G., Radhakrishnan,R. and Epp,O. (1988) *Acta Crystallogr.*, **B44**, 26–38.

Neidhardt,D.J., Kenyon,G.L., Gerlt,J.A. and Petsko,G.A. (1990) *Nature*, **347**, 692–694.

Nordland,I. and Shingler,V. (1990) *Biochim. Biophys. Acta*, **1049**, 227–230.

Pathak,D. and Ollis,D. (1990) *J. Mol. Biol.*, **214**, 497–525.

Pathak,D., Ngai,K.-L. and Ollis,D. (1988) *J. Mol. Biol.*, **204**, 435–445.

Pathak,D., Ashley,G. and Ollis,D. (1991) *Proteins: Struct., Funct. Genet.*, **9**, 267–279.

Pflugrath,J.W., Saper,M.A. and Quiocho,F.A. (1984) In Hall,S. and Ashida,T. (eds), *Methods and Applications in Crystallographic Computing*. Clarendon Press, Oxford, pp. 404–407.

Quinn,D.M. (1987) *Chem. Rev.*, **87**, 955–979.

Rees,D.C., Lewis,M. and Lipscomb,W. (1983) *J. Mol. Biol.*, **168**, 367–387.

Rice,P.R., Goldman,A. and Steitz,T.A. (1990) *Proteins: Struct., Funct. Genet.*, **8**, 334–341.

Richardson,J.S. (1976) *Proc. Natl. Acad. Sci. USA*, **73**, 2619–2623.

Richardson,J.S. (1981) *Adv. Prot. Chem.*, **34**, 167–339.

Rossmann,M.G. and Argos,P. (1975) *J. Biol. Chem.*, **250**, 7525–7529.

Rossmann,M.G., Liljas,A., Brändén,C.-I. and Banaszak,L.J. (1975) In Boyer,P.D. (ed.), *The Enzymes*. Academic Press, New York, Vol. XIA, p. 61–102.

Schrag,J.D., Li,Y., Wu,S. and Cygler,M. (1991) *Nature*, **351**, 761–764.

Schumacher,M., Camp,S., Maulet,Y., Newton,M., MacPhee-Quigley,K., Taylor,S.S., Friedmann,T. and Taylor,P. (1986) *Nature*, **319**, 407–409.

Shimada,Y., Sugihara,A., Izumi,T. and Tominaga,Y. (1990) *J. Biochem.*, **107**, 703–707.

Slabas,A.R., Windust,J. and Sidebutton,C.M. (1990) *B.J. Lett.*, **269**, 279–280.

Sussman,J.L., Harel,M., Frolow,F., Oefner,C., Goldman,A., Toker,L. and Silman,I. (1991) *Science*, **253**, 872–879.

Thomas,T., Cooper,A., Bussey,H. and Thomas,G.J. (1990) *J. Biol. Chem.*, **265**, 10821–10824.

Winkler,F.K., D'Arcy,A. and Hunziker,W. (1990) *Nature*, **343**, 771–774.

Wright,C.S., Alden,R.A. and Kraut,J. (1969) *Nature*, **221**, 235–242.

Received on December 17, 1991; revised and accepted on January 21, 1992

Note added in proof

It should be noted that the handedness of papain and the enzymes of the α/β hydrolase fold is the same.