Plasticity of enzyme active sites

Annabel E. Todd, Christine A. Orengo and Janet M. Thornton

The expectation is that any similarity in reaction chemistry shared by enzyme homologues is mediated by common functional groups conserved through evolution. However, detailed enzyme studies have revealed the flexibility of many active sites, in that *different* functional groups, unconserved with respect to position in the primary sequence, mediate the same mechanistic role. Nevertheless, the catalytic atoms might be spatially equivalent. More rarely, the active sites have completely different locations in the protein scaffold. This variability could result from: (1) the hopping of functional groups from one position to another to optimize catalysis; (2) the independent specialization of a low-activity primordial enzyme in different phylogenetic lineages; (3) functional convergence after evolutionary divergence; or (4) circular permutation events.

Annabel E. Todd
Christine A. Orengo
Biochemistry and Molecular Biology Dept, University College London, Gower Street, London, UK WC1E 6BT.

Janet M. Thornton
European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, UK CB10 1SD.
e-mail: thornton@ebi.ac.uk

Protein structures are resilient to evolutionary amino acid changes. Related sequences can diverge to such an extent that their common ancestry, obscured at the sequence level, is evident only in their structural similarity [1]. Although divergence in function has accompanied divergence in sequence within many protein superfamilies, members of a family often share a characteristic functional feature, such as a common binding property or mechanistic strategy in catalysis [2,3]. The expectation is that a conserved functional feature is associated with one or more key residues that are invariant across a family of proteins; owing to their important functional role, these amino acids are subject to evolutionary constraints and their loss during evolution would be deleterious to function. Meanwhile, residues at other positions are freer to mutate. Invariant amino acids could also correspond to residues of structural significance, in that they maintain the integrity of the protein fold.

Indeed, there are many superfamilies in which absolutely conserved functional residues have been identified. For example, members of the type I pyridoxal 5′-phosphate (PLP)-dependent aspartate aminotransferase-like superfamily have diverged considerably in sequence; nevertheless, they all have invariant Asp and Lys residues in their active sites and these are involved in PLP-binding and catalysis. Similarly, an absolutely conserved Cys residue in the cytochrome P450s forms the fifth ligand of the haem iron.

Owing to the recent and rapid increase in structural data, many enzyme superfamilies in the Protein Data Bank (PDB) [4] are now well populated. Individual superfamily analyses have revealed how nature has optimized binding and catalysis, and re-structured 'old' proteins for new activities through gene duplication and mutation [3]. Increasingly, we are finding that active sites are variable, confounding the expectation of functional residue conservation in protein homologues. In fact, the seminal work of Babbitt, Gerlt and many other investigators over the past decade, has shown how homologous enzymes can catalyse different overall reactions and reaction chemistries, sometimes with different catalytic mechanisms [2,5,6]. Often, these functional modifications are achieved through changes in structural characteristics, either in the neighbourhood of the active site or through more remote structural changes that can affect interactions with other domains or subunits [3,7].

However, examples of the converse situation, in which similar reaction chemistries are achieved in different ways, are more rare, although Grishin recently presented some interesting cases involving structural rearrangements in the region of the active site [7]. Here, we discuss examples of related enzymes in which *different* functional groups, unconserved with respect to position in the primary sequence, mediate the same mechanistic role in catalysis. That is, two or more solutions to the same catalytic conundrum have evolved within a single superfamily. These examples are discussed in the context of protein structure–function relationships and the evolution of catalysis.

Examples

Table 1 lists 17 examples, identified in 16 superfamilies, of enzyme homologues in which residues that play the same role in catalysis are located at non-equivalent positions in the structural scaffold. Non-equivalent residues are those that do not align in the structure-based sequence alignment. This type of active-site variation within the α/β hydrolase, enolase and thioredoxin superfamilies has been discussed from an evolutionary perspective [8–10], and the ideas and hypotheses raised by the authors of these articles have been extended in this review.

In recent work, Todd *et al.* [3] reported a collective analysis of 31 enzyme superfamilies in the PDB. Their work, with further analyses after publication, revealed that as many as 15 of the 31 superfamilies studied exhibit active-site variation of this nature. This suggests that the poor conservation of functional groups in related enzymes could be widespread. Many of these 15 superfamilies are included in Table 1, in which a brief description of each variation is provided, and the examples are collectively summarized below.

Similarity in overall enzyme function

Seven out of 17 examples in Table 1 involve proteins with apparently identical biochemical functions, as defined by their Enzyme Commission (EC) numbers [11] (a heirarchical classification of enzyme reactions,

 0968-0004/02/$ – see front matter PII: S0968-0004(02)02158-8

Table 1. Enzyme relatives in which the same catalytic role is mediated by residues that are non-equivalent in the structure-based sequence alignment

Superfamily	Catalytic role	Protein	EC[a]	PDB	Residue(s)	Comment[b]
Enolase	Stabilization of enolate intermediate	Enolase	4.2.1.11	1one	K396	In 1one the Lys electrophile is on β8 but in 2mnr it is on β2 of the $(\beta\alpha)_8$ barrel. The catalytic atoms are co-located.
		Mandelate racemase	5.1.2.2	2mnr	K164	
Hexapeptide repeat proteins	Putative base for deprotonation of substrate hydroxyl	Chloramphenicol acetyltransferase	2.3.1.28	1xat	H79	In 1lxa the essential His is located in the core of the β-solenoid structure. In 1xat it is situated in a long loop.
		UDP-*N*-acetylglucosamine acyltransferase	2.3.1.129	1lxa	H125	
α/β hydrolases	Acid in nucleophile–His–acid catalytic triad	Triacylglycerol lipase	3.1.1.3	1thg	E354	1thg is similar to typical superfamily members, with the acid in a loop after β7. In 1lpb it is after β6. (Fig.2a) [8,17].
		Pancreatic triacylglycerol lipase	3.1.1.3	1lpb	D176	
Phosphoglycerate mutase-like	Acid for protonation of oxygen in substrate scissile P–O bond in initial enzyme phosphorylation step	Fructose-2,6-bisphosphatase	3.1.3.46	1bif	E325	In 1bif and phosphoglycerate mutase, the acid is a Glu located in a long loop after β3, but in 1dkq and acid phosphatases it is an Asp located just after β4. Nevertheless, the relative disposition of the acid and the conserved catalytic His nucleophile in all proteins is similar.
		3-phytase	3.1.3.8	1dkq	D304	
Medium-chain alcohol dehydrogenases	Proton shuttle system for removal of hydroxyl proton of alcohol substrate to bulk solvent	Human alcohol dehydrogenase class I β	1.1.1.1	1deh	T48,H51	62% sequence identity. One His and one Thr probably constitute the proton shuttle in these enzymes and the His functions also as the base catalyst. The Thr residues are equivalent, but the His preceeds Thr in 1teh whereas it is located three positions after Thr in 1deh.
		Human alcohol dehydrogenase class III χ	1.1.1.1	1teh	H47,T48	
Creatinase/ aminopeptidase	Activation of water and protonation of nitrogen in scissile bond in tetrahedral intermediate	Creatinase	3.5.3.3	1chm	H232	The domain comprises a pair of related α/β modules. H232 (1chm) is located on a long loop after β1 in the first module, whereas E204 (1c24) is on a strand in the second module.
		Prokaryotic methionine aminopeptidase	3.4.11.18	1c24	E204	
Aldehyde reductase (dehydrogenase)	General base	Aldehyde dehydrogenase class II	1.2.1.3	1cw3	E268	The two Glu residues are conserved in both enzymes. They are located in different domains.
		Aldehyde dehydrogenase class III	1.2.1.5	1ad3	E333	
DNA breaking and/or rejoining enzymes, λ integrase-like	Tyr nucleophile involved in 3′-phosphotyrosine DNA-enzyme formation	DNA topoisomerase I	5.99.1.2	1ecl	Y319	The Tyr is located on a different secondary structure element in each of the four enzymes, in their helix-turn-helix-containing domain.
		DNA topoisomerase II	5.99.1.3	1bgw	Y783	
		DNA topoisomerase IV	5.99.1.3	1d3y	Y103	
		Cre recombinase	-	1crx	Y324	
Restriction endonuclease-like	Mg^{2+}-binding	Restriction endonuclease *Eco*RI	3.1.21.4	1ckq	D91,E111	The Asp residues are equivalent, but in 1ckq the Glu is on a β strand whereas, in 1cfr, it is at the N terminus of a helix. The catalytic atoms in the Glu residues are co-located [13].
		Restriction endonuclease *Cfr*10I	3.1.21.4	1cfr	D134, E204	
Thioredoxin	Redox-active (seleno)cysteine, oxidized to X–OH during catalysis	Glutathione peroxidase	1.11.1.9	1gp1	Se45[c]	Se45 (1gp1) is equivalent to the first Cys of the redox-active CxxC motif in the homologue thioredoxin but C47 (1prx) is equivalent to the second Cys in the motif.
		Peroxidase	1.11.1.7	1prx	C47	
	Activation of glutathione by promoting thiolate formation	Glutathione S-transferase class α	2.5.1.18	1b48	Y9	These functional groups are separated by three residues along the primary sequence. The catalytic atoms are co-located. The β class has a different, but unknown, mechanism of glutathione activation. (Fig. 2a) [10,12].
		Glutathione S-transferase class θ	2.5.1.18	1ljr	S11	
DD-peptidase/ β-lactamase	Activation of Ser nucleophile in enzyme acylation and activation of water in deacylation	β-lactamase class A	3.5.2.6	1dja	K73 or E166	The acid–base catalysts are located on different secondary structure elements in these two β-lactamase classes.
		β-lactamase class C	3.5.2.6	1gce	Y150	

contd…

in which each reaction is assigned a four-digit number). Note that enzymes with the same EC classification, however, can differ slightly in their substrate-specificity profiles. Classical mammalian ferritins and *Listeria innocua* ferritin of the di-iron carboxylate protein superfamily probably have analogous iron-storage functions. They are not classified as enzymes. Nevertheless, they have ferroxidase activity that is necessary for this storage function. In a further two examples, the enzymes differ only in their cofactor requirements. In total, therefore, in over one-half of the examples listed, the proteins in question do not share only a simple mechanistic strategy that is used in diverse reactions. Instead, the proteins have identical, or very similar, overall biochemical functions.

Table 1. Enzyme relatives in which the same catalytic role is mediated by residues that are non-equivalent in the structure-based sequence alignment contd

Superfamily	Catalytic role	Protein	EC[a]	PDB	Residue(s)	Comment[b]
FMN-dependent oxidoreductases	H-bond with N5 atom of flavin	Flavocytochrome b_2	1.1.2.3	1fcb	A198 backbone amide	In 1fcb and other homologues the hydrogen-bond donor is a backbone amide located in the loop that follows β1 of the $(\beta\alpha)_8$ barrel, but in 1dor the donor is a nitrogen in a Lys sidechain located after β2. The donor atoms are co-located.
		Dihydroorotate dehydrogenase A	1.3.3.1	1dor	K43 sidechain	
Ferritin-like di-iron carboxylate proteins	Ferroxidase di-iron centre	Human ferritin	Non-enzyme	2fha	E27,Y34, E61,E62, H65,E107, Q141	In classical ferritins the ferroxidase di-iron site, necessary for their iron-storage function, is located in the core of the four-helix bundle structure. In 1qgh, the site is located at the interface between two subunits in the dodecamer [16].
		Listeria innocua ferritin	Non-enzyme	1qgh	H31,H43, D47,D58′, E62′[d]	
Aldolase	Formation of Schiff base intermediate with the substrate	Fructose-bisphosphate aldolase class I	4.1.2.13	1fba	K229	In 1fba the Lys is situated on β6 whereas in 1onr it is on β4. (Fig. 2e) [29].
		Transaldolase	2.2.1.2	1onr	K132	
Zn peptidases	Activation of water	Carboxypeptidase A	3.4.17.1	2ctc	E270	In 2ctc the Glu base is in the C-terminal region of the fold but in 1amp it is in the middle of the polypeptide chain. (Fig. 2d) [24].
		Bacterial Leu aminopeptidase	3.4.11.10	1amp	E151	
Pyridine nucleotide disulfide oxidoreductases	Two redox-active Cys that oscillate between dithiol and disulfide redox states during catalysis; one interacts with the substrate	High MW thioredoxin reductase	1.6.4.5	1h6v	C59xxxxC64[e]	In 1h6v, the redox-active disulfide is located in the FAD-binding domain, in 1tde it is in the NAD(P)(H)-binding module, and in 1gcd the two redox-active Cys are located on two distinct domains and lie almost 200 residues apart. None of the Cys residues is equivalent. (Fig. 2d) [25,26].
		Low MW thioredoxin reductase	1.6.4.5	1tde	C135xxC138	
		Flavocytochrome disulfide dehydrogenase	1.8.2.-	1fcd	C161x_nC337	

[a]Enzyme Commission (EC) [11] is a hierarchical classification of enzyme reactions. A reaction is assigned a four-digit EC number, in which the first digit denotes the class of reaction (1, oxidoreductases; 2, transferases; 3, hydrolases; 4, lyases; 5, isomerases; 6, ligases). Subsequent levels or nodes in the hierarchy become increasingly specific. For example, enzymes that vary only in the fourth digit in their EC assignments could differ in their substrate and/or product specificity, or cofactor dependency. Some enzyme reactions lack an EC classification and are marked with '-'. [b]In the vast majority of cases, the origin of the variation in active-site residues is unclear without sufficient sequence and structural data, and detailed phylogenetic analyses that were beyond the scope of this review. For a few cases, however, suppositions regarding the origin of the variation are provided in the literature. These have been marked with reference to Fig. 2 and references are given. [c]Se represents a selenocysteine residue. [d]A prime symbol (′) denotes a different subunit in the dodecamer. [e]x denotes an intervening residue.

Conservation of residue identity

In nine cases, the non-equivalent, but functionally analogous, catalytic residues are identical. It might be that, for these enzymes, there is an optimum amino acid type for the chemical role in question. In a further three cases, there is a conservative difference in amino acid identity (Asp–Glu and Ser–Tyr).

Relative location in the primary sequence

In just three examples, the equivalent functional residues in the enzyme homologues are separated by three or fewer amino acids (or positions) in the structure-based sequence alignment. More commonly, the equivalent functional groups are situated on non-equivalent secondary structure elements and they are distant with respect to location in the primary sequence. This is consistent with the observation that active sites are formed usually from several different regions of the polypeptide chain. In two cases, the equivalent catalytic groups lie in different structural domains.

Spatial equivalence of catalytic atoms

In a few cases, the position of the specific atoms involved in catalysis is preserved, whereas the residues to which they belong lie at different points in the protein scaffold [3,8,9,12,13]. For these active sites at least, this spatial conservation implies that there is an optimum disposition of functional atoms for catalysis, although there is a degree of flexibility with respect to the locations of the residues that contain them.

Active sites in completely different locations in the fold

For almost all examples in Table 1, the active sites in the enzyme homologues are co-located. That is, they are situated in equivalent pockets in the structural scaffold, as implied by the spatial equivalence of catalytic atoms already discussed.

For the ferritins, however, the active sites are in completely different locations in the fold. These iron-storage proteins have a di-iron site with an associated ferroxidase activity that is responsible for the formation of a di-ferric species, an intermediate in the iron oxidation and uptake process [14]. In classical ferritins, and in other superfamily members such as ribonucleotide reductase [15], the ferroxidase di-iron centre is located in a four-helix bundle core [14]. Ferritin from *L. innocua* is unusual in that despite having ferroxidase activity, it lacks all the iron-binding residues in the classical ferroxidase centre [16]. In this protein, the ferroxidase di-iron site is located at the interface between two subunits related

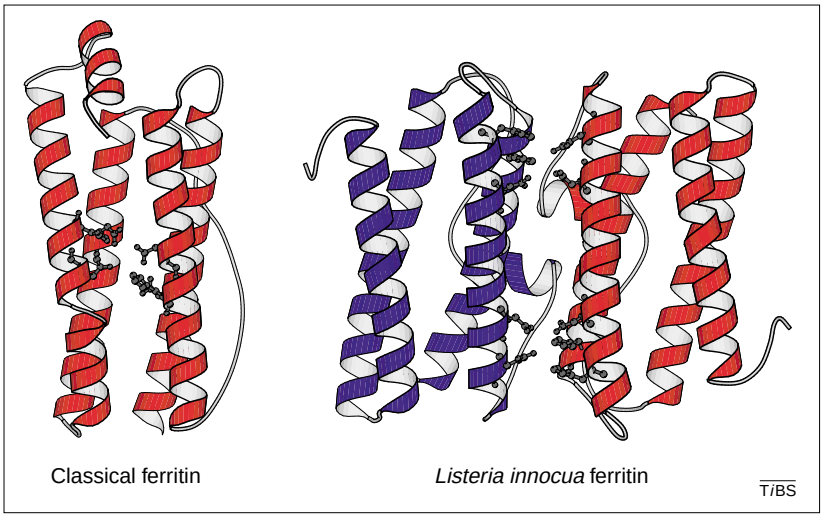

Fig. 1. The difference in location of the ferroxidase di-iron site in ferritins. Iron-binding residues are shown in ball-and-stick representation. In classical ferritins, the di-iron site is located within the four-helix bundle core, which is typical of members of this superfamily. In *Listeria innocua* ferritin, the di-iron site is located at the interface between two subunits, but it is chemically and structurally similar to that in classical ferritins [16]. The figure was generated using Molscript [35].

by twofold symmetry (Fig. 1). Despite the difference in location, this centre is both chemically and structurally similar to that in classical ferritins [16].

Possible origins of the variability in active-site residues

The non-equivalence of functional residues that play the same catalytic role in related proteins is unexpected. Several possible origins for this variability are outlined here and in Fig. 2. For almost all superfamilies, the origin is unclear. Only with a more complete catalogue of sequences and structures for a particular superfamily, and with detailed phylogenetic analyses, can the evolution of enzyme function and catalysis be fully understood.

Evolutionary optimization by functional residue hopping

Active-site variation might reflect the evolutionary optimization of the catalytic efficiency of these enzymes. As new groups fortuitously evolve in the active site they might take over the roles of 'old' residues if they are better suited to the job. This could be referred to as 'functional residue migration' [8] or 'hopping' (Fig. 2a). Structural reorganization of the active site in this way might also be favoured if it improves protein stability. Alternatively, hopping could have occurred without the aim of optimizing catalysis or stability. It might result simply from the loss of an 'old' catalytic residue by mutation and, before the deletion of the gene from the organism can occur, a 'new' residue is recruited, which has fortuitously evolved to play the same catalytic role.

Evidence for the hopping of a functional group is provided by the presence of the 'old' residue as an evolutionary relic, which no longer plays a role in catalysis. For example, in some lipases of the α/β hydrolase superfamily, the acid in the canonical nucleophile–His–acid catalytic triad has migrated from strand seven to strand six in the main β sheet. In human pancreatic lipase, the ancestral acid on strand seven remains as an evolutionary relic, and points away from the catalytic His, unable to make hydrogen-bonding contact [8] (Fig. 3a).

The facile redesign of the *Geotrichum candidum* lipase active site by site-directed mutagenesis provides further evidence for this hopping mechanism in the α/β hydrolases [17]. Enzyme activity was maintained when the acid in the catalytic triad of this protein was moved from strand seven to strand six. Analogously, the catalytic metal-binding site of restriction endonuclease Cfr10I (Table 1) was successfully changed to the canonical active site, without loss of activity [13].

Alternatively, the 'new' residue might exist in the 'old' enzyme as an evolutionary precursor to catalysis. The glutathione-S-transferases (GSTs) have been classified into several distinct subfamilies on the basis of primary sequence and substrate specificity. Members of most classes have an essential Tyr residue involved in glutathione activation. This is present in a few θ class GSTs, but it is not part of the active site [12] and, instead, members of this class use a Ser for the same job, located four positions downstream (Fig. 3b). Because the θ class probably predates the other GST subfamilies, the non-catalytic Tyr might have been recruited for catalysis in the evolution of other subfamilies [10].

Independent optimization in different phylogenetic lineages

Alternatively, variability in the active sites of proteins with the same activity might reflect independent optimization of a less efficient primordial enzyme in different phylogenetic lineages, rather than an actual hopping of residues from one position to another (Fig. 2c). During evolution, different amino acids could be recruited in distinct protein families to, nevertheless, evolve the same type of high-activity enzyme. This might have occurred in the evolution of the α/β-fold uracil DNA glycosylases [18].

Evolution of a new activity by functional residue hopping

The fortuitous appearance of alternative residues that have the necessary functionality within the active site might, itself, be a method by which completely *new* activities have evolved, rather than being simply a means by which a function already in place is optimized (Fig. 2b). This is well illustrated by a directed evolution experiment in the $(\beta\alpha)_8$ barrel protein 2-keto-3-deoxy-6-phosphogluconate aldolase. Repositioning of the key catalytic Schiff base-forming Lys residue from strand six to strand seven in the central β-barrel significantly broadens the substrate profile of this enzyme [19].

Functional convergence after evolutionary divergence

Another possibility is that a particular enzyme activity has evolved two or more times during evolution within a single superfamily by completely independent means. That is, through functional convergence after evolutionary divergence (Fig. 2d); two homologous primordial enzymes with distinct functions are both duplicated, and the two copies

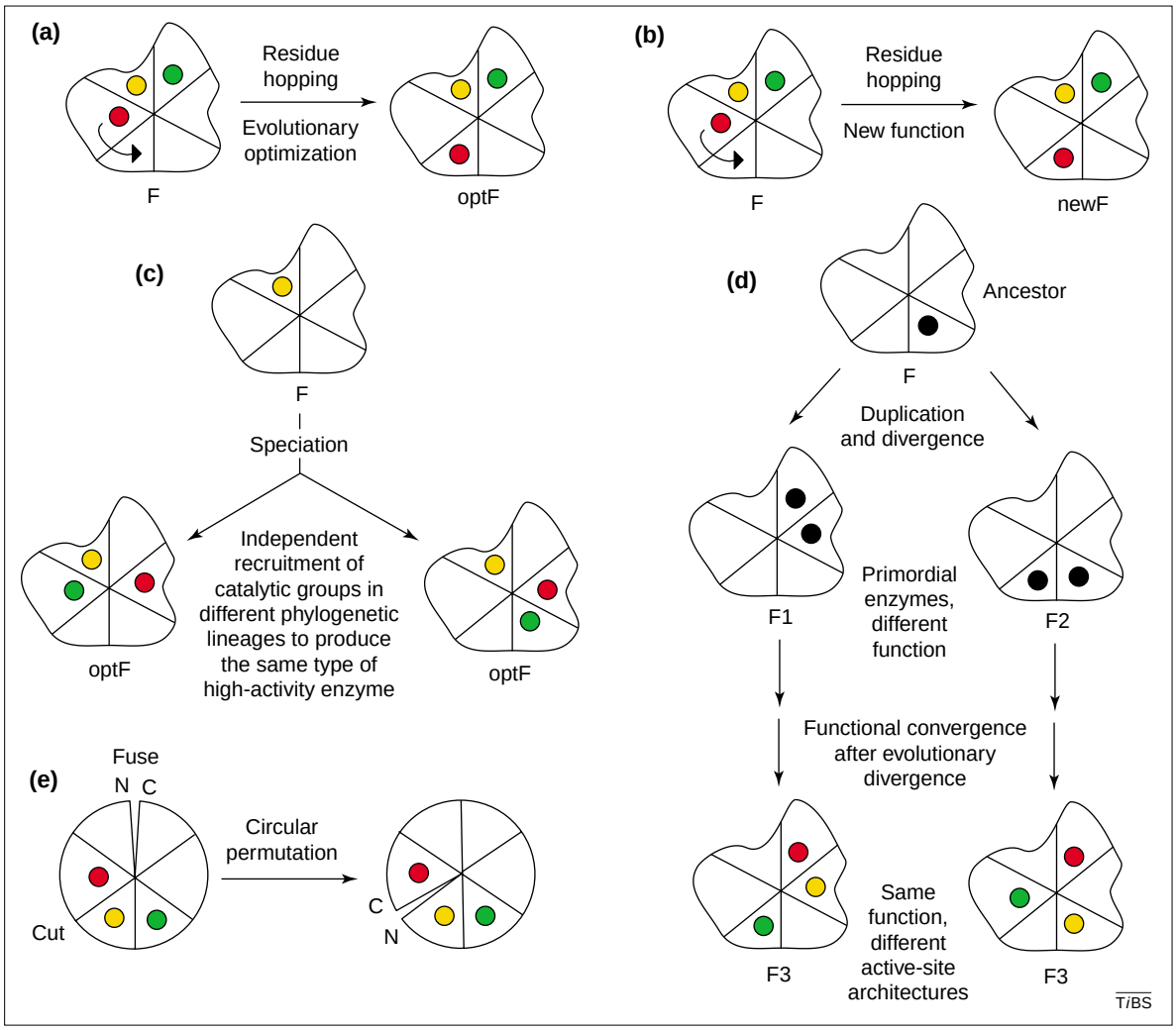

Fig. 2. Hypotheses regarding the non-equivalence of functional residues that play the same catalytic role in enzyme homologues. Each shape represents a protein structure, and each 'segment' in the protein structure represents a position in the fold at which a functional residue can be situated. Coloured or black circles denote functional residues, and those with the same colour have analogous roles in catalysis. 'F' or 'F1', for example, denote particular functions, and 'optF' denotes the same, but optimized, function. 'newF' denotes a new function, and 'N' and 'C' denote the N- and C termini, respectively.

diverge to evolve new functions that are, nevertheless, identical to each other. The independent occurrence of the same enzyme activity in different structural scaffolds, typified by subtilisin and chymotrypsin, is also known [20–23]. Evidence points to independent evolution of the same function in the Zn peptidase [24] and the pyridine nucleotide disulfide oxidoreductase superfamilies [25,26] (Table 1).

This second enzyme superfamily provides an amazing example of active-site variation (Fig. 4). All members contain a FAD-binding domain and an NAD(P)(H)-binding region. Typical members, including high molecular weight mammalian thioredoxin reductases, have a third interface domain at the C terminus, and a catalytic redox-active disulfide located in the FAD domain. By contrast, plant and bacterial low molecular weight thioredoxin reductases lack the interface domain, and the disulfide is located in the NAD(P)(H)-binding module [25], in a catalytically incompetent position on the *re* face of the isoalloxazine ring of FAD; this necessitates a large conformational change before catalysis. The same function has probably evolved on two independent occasions within the superfamily [25]. Interestingly, flavocytochrome *c* : sulfide dehydrogenase has yet another active-site arrangement. Whereas the redox-active Cys residues are separated by just several amino acids along the primary sequence in other members, in this enzyme they are located on two distinct domains and lie almost 200 residues apart [27]. Another disulfide oxidoreductase belonging to this family, coenzyme A disulfide reductase from *Staphylococcus aureus*, has been sequenced [28]. This protein is more similar to NADH peroxidase, which acts on hydrogen peroxide rather than disulfide substrates, than to other disulfide oxidoreductases in the family. These enzymes have only one redox-active Cys residue, not two, located in the FAD-binding domain. Thus, members of this superfamily use similar sulfur redox chemistries in one of four distinct active sites.

Circular permutation

Lastly, genetic rearrangements can account for the apparent migration of catalytic residues (Fig. 2e). Circular permutation involves fusion of the N and C termini of the protein, and the polypeptide chain is cut in another place. An excellent example is provided by the aldolase superfamily, members of which have the $(\beta\alpha)_8$ barrel fold, also known as the

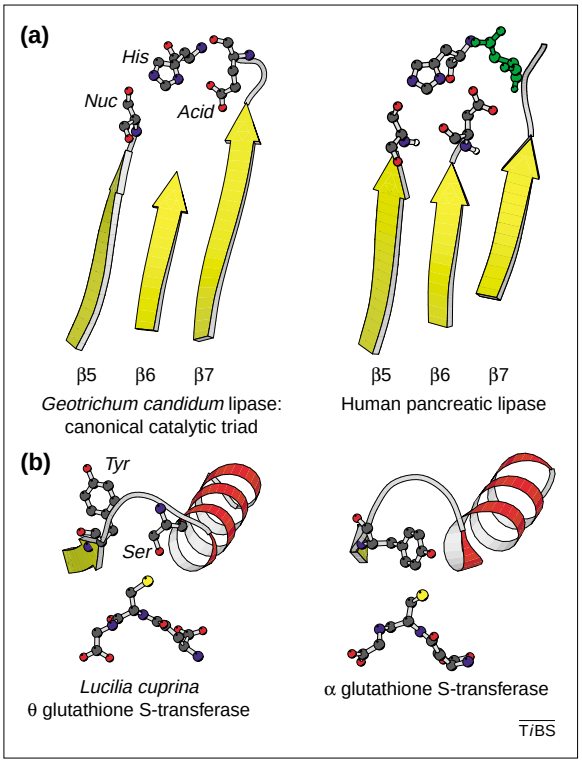

Fig. 3. (a) Migration of the catalytic acid from strand seven to strand six in the triacylglycerol lipases of the α/β hydrolase superfamily [8,17]. Enzymes in this family have a nucleophile–His–acid catalytic triad. In *Geotrichum candidum* lipase and other typical family members, the acid is located in a loop after strand seven, but in human pancreatic lipase (HPL) it follows strand six. The ancestral acid after strand seven (green) remains as an evolutionary relic and points away from the catalytic His, unable to make hydrogen-bonding contact. Lipoprotein lipases are similar to HPL in that the acid follows strand six, but the ancestral acid has gone (not shown). Thus, HPL represents an intermediate in the movement of the acid from strand seven to strand six [8]. 'Nuc' denotes the catalytic nucleophile. (b) Evolution of glutathione interactions in the glutathione S-transferases (GSTs) [10]. θ class enzymes use a Ser for glutathione activation, but members of other subfamilies use a non-equivalent Tyr residue. In *Lucilia cuprina* θ GST, and in a few other θ GSTs, the same Tyr is present but it is not part of the active site. As the θ class probably predates the other GST subfamilies, this non-catalytic Tyr might have been recruited for catalysis in the evolution of other classes. In α GST, the active site Tyr can be seen close to the thiol group of glutathione and is closely involved in catalysis. Figure is adapted, with permission, from [36], and was generated using Molscript [35].

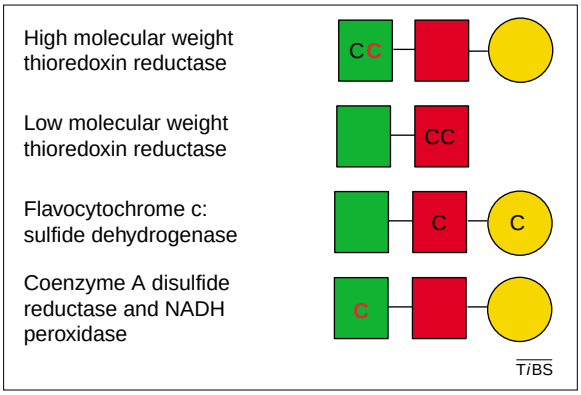

Fig. 4. Variation in the active-site location in members of the pyridine nucleotide disulfide oxidoreductase superfamily [25–28]. Each shape corresponds to a structural domain, and domains of the same shape are homologous. The green square represents the FAD-binding domain, the red square represents the NAD(P)(H)-binding domain and the circle represents the interface domain. The NAD(P)(H)-binding domain interrupts the first domain in all proteins, but it is not indicated in the diagram for clarity. Each redox-active Cys is represented by 'C' to illustrate its domain location; only those cysteines in red are structurally equivalent.

triosephosphate isomerase (TIM) barrel. This structure comprises eight βα motifs, assembled in a circular arrangement, such that there is an eight-stranded β-barrel in the core.

Fructose-bisphosphate aldolase class I and several other mechanistically similar homologues, including transaldolase, have an active-site Lys that forms a Schiff base intermediate with the substrate. Typically, this Lys is located on strand six of the TIM barrel. In transaldolase, however, it is on strand four, and optimal sequence and structural alignment with aldolase class I requires circular permutation of the β strands in the barrel [29]. The strong similarity of these two enzymes in their substrate specificities and reaction chemistries supports the evolution of transaldolase from class I aldolase; such a relationship would require movement of the first two β strands to the C terminus after other members of the family diverged [29] (Fig. 5). Given its high degree of symmetry, a genetic permutation in the TIM barrel fold yields a protein with topological connections analogous to those of its precursor, and it appears that the active-site residues have collectively 'hopped along' one or more strands.

Properties of the fold and functional diversification

A greater degree of active-site flexibility must facilitate functional diversification. It is possible that folds offering several points from which catalytic groups can be recruited are those that are particularly susceptible to adaptation for new functions. This might account, at least in part, for the functional diversity of the TIM barrel, a fold that supports an enormous variety of enzyme functions (>60 different enzyme classifications are recorded) [30]. In enzymes with this fold, active-site residues are located invariably at the C termini of the β strands, or in the subsequent loops [31]. Because of its cylindrical shape, the TIM barrel can recruit catalytic residues from up to eight possible βα motifs [2]. For example, in the enolase superfamily, functional groups are donated from as many as seven of the eight β strands. Furthermore, directed evolution experiments have illustrated the ease with which the functions of proteins adopting this fold can be altered, not only with respect to a change in substrate specificity [32] but also to a change in catalytic activity [33].

Conclusions

When the first examples of functional equivalence but positional inequivalence of active-site residues were identified in protein homologues, it was unknown whether the evolutionary processes at work within these families were peculiar to these families, or whether they

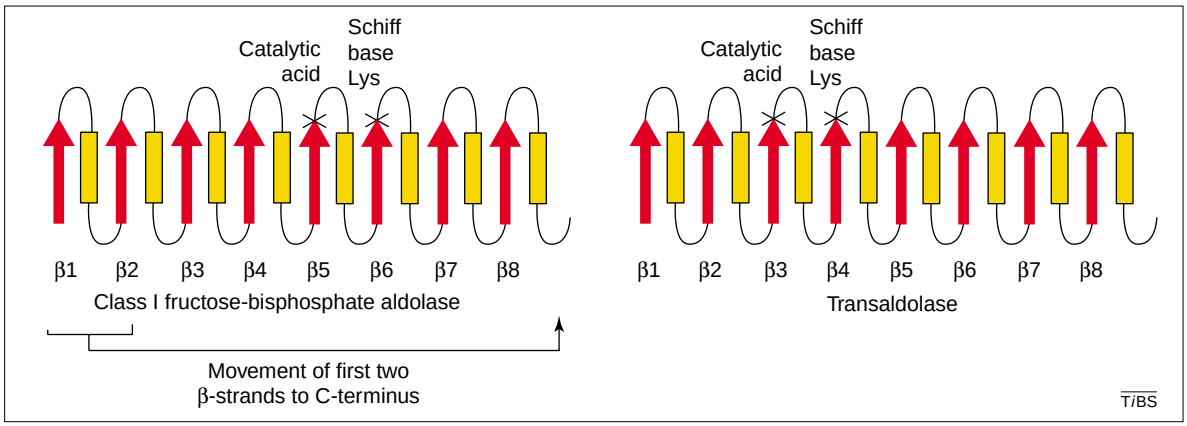

Fig. 5. Circular permutation in the evolution of transaldolase from aldolase class I [29]. The triosephosphate isomerase (TIM) barrel structures of these proteins are 'rolled' out for clarity; arrows and rectangles represent strands and helices, respectively. Fructose-bisphosphate aldolase class I and other family members have an active-site Lys that forms a Schiff base intermediate with the substrate. Typically, this Lys is located on strand six of the TIM barrel. In transaldolase, however, it is on strand four, and optimal sequence and structural alignment with aldolase class I requires circular permutation of the β strands in the barrel. The strong functional similarities shared by these two enzymes supports the evolution of transaldolase from class I aldolase; such a relationship would require movement of the first two β strands to the C terminus after other members of this family diverged.

were a more widespread phenomena [9]. We propose that these peculiarities are not so unusual after all and that many active sites are actually quite 'malleable'. We expect that there are other examples hidden in the literature that have not been reported in this article, and that many more cases will be identified with the increase in structural and biochemical data over the next few years. Other active-site 'anomalies' besides those discussed here are also known. For example, several metal-dependent enzymes have catalytically active cofactor-independent homologues that do not bind metal ions, implying a difference in reaction chemistry [3].

Genome sequencing has seen the advent of structural genomics initiatives that aim to derive a structural representative for all homologous protein families. As such, biologists are presented with a complete reversal of the classical approach to protein characterization. That is, the three-dimensional structure of a protein is determined before having any knowledge of its function. Inference of function from structure is not straightforward, but through the identification of structural relatives and conserved active-site architectures in enzyme families, functional features can be inferred [34]. The plasticity of active sites presents a challenge to the success of these projects. The absence of one or more catalytic residues in a newly determined structure does not necessarily imply a lack of, or difference in, enzyme activity. Another major challenge is to engineer proteins with novel functions. On a far more optimistic note, that proteins can offer alternative points for catalytic groups is promising for protein design.

Acknowledgements
A.E.T. acknowledges financial support from the Midwest Centre for Structural Genomics. This is a publication from the Bloomsbury Centre for Structural Biology. We thank Michael W. Parker for the coordinates of *Lucilia cuprina* θ class glutathione S-transferase. The data in Table 1 have been collected from numerous structural studies and biochemical experiments. Owing to space restrictions, it has not been possible to cite many of the articles from which biochemical data were extracted, and we acknowledge the numerous experimentalists who have contributed to these publications. Many references can be found in the PDB [4].

References

1 Chothia, C. and Lesk, A. (1986) The relation between divergence of sequence and structure in proteins. *EMBO J.* 5, 823–826
2 Gerlt, J.A. and Babbitt, P.C. (1998) Mechanistically diverse enzyme superfamilies: the importance of chemistry in the evolution of catalysis. *Curr. Opin. Chem. Biol.* 2, 607–612
3 Todd, A.E. *et al.* (2001) Evolution of function in protein superfamilies, from a structural perspective. *J. Mol. Biol.* 307, 1113–1143
4 Berman, H.M. *et al.* (2000) The Protein Data Bank. *Nucleic Acids Res.* 28, 235–242
5 Petsko, G.A. *et al.* (1993) On the origin of enzymatic species. *Trends Biochem. Sci.* 18, 372–376
6 Babbitt, P.C. and Gerlt, J.A. (1997) Understanding enzyme superfamilies – chemistry as the fundamental determinant in the evolution of new catalytic activities. *J. Biol. Chem.* 272, 30591–30594
7 Grishin, N.V. (2001) Fold change in evolution of protein structures. *J. Struct. Biol.* 134, 167–185
8 Schrag, J.D. *et al.* (1992) Pancreatic lipases – evolutionary intermediates in a positional change of catalytic carboxylates. *J. Biol. Chem.* 267, 4300–4303
9 Hasson, M.S. *et al.* (1998) Evolution of an enzyme active site: the structure of a new crystal form of muconate lactonizing enzyme compared with mandelate racemase and enolase. *Proc. Natl. Acad. Sci. U. S. A.* 95, 10396–10401
10 Rossjohn, J. *et al.* (1998) A mixed disulphide bond in bacterial glutathione transferase: functional and evolutionary implications. *Structure* 6, 721–734
11 Nomenclature Committee of the International Union of Biochemistry and Molecular Biology (NC-IUBMB) (1992) *Enzyme Nomclature*, Academic Press
12 Wilce, M.C.J. *et al.* (1995) Crystal structure of a θ-class glutathione transferase. *EMBO J.* 14, 2133–2143
13 Skirgaila, R. *et al.* (1998) Structure-based redesign of the catalytic/metal binding site of Cfr10I restriction endonuclease reveals importance of spatial rather than sequence conservation of active centre residues. *J. Mol. Biol.* 279, 473–481
14 Harrison, P.M. and Arosio, P. (1996) Ferritins: molecular properties, iron storage function and cellular regulation. *Biochim. Biophys. Acta* 1275, 161–203
15 Nordlund, P. *et al.* (1990) Three-dimensional structure of the free radical protein of ribonucleotide reductase. *Nature* 345, 593–598
16 Ilari, A. *et al.* (2000) The dodecameric ferritin from *L. innocua* contains a novel intersubunit iron-binding site. *Nat. Struct. Biol.* 7, 38–43
17 Schrag, J.D. *et al.* (1995) Redesigning the active site of *Geotrichum candidum* lipase. *Protein Eng.* 8, 835–842
18 Aravind, L. and Koonin, E.V. (2000) The α/β fold uracil DNA glycosylases: a common origin with diverse fates. *Genome Biol.* 1, 0007.1–0007.8
19 Wymer, N. *et al.* (2001) Directed evolution of a new catalytic site in 2-keto-3-deoxy-6-phosphogluconate aldolase from *Escherichia coli*. *Structure* 9, 1–9
20 Blow, D.M. *et al.* (1969) Role of a buried acid group in the mechanism of action of chymotrypsin. *Nature* 221, 337–340
21 Wright, C.S. *et al.* (1969) Structure of subtilisin BPN′ at 2.5Å resolution. *Nature* 221, 235–242
22 Galperin, M.Y. *et al.* (1998) Analogous enzymes: independent inventions in enzyme evolution. *Genome Res.* 8, 779–790
23 Hegyi, H. and Gerstein, M. (1999) The relationship between protein structure and function: a comprehensive survey with application to the yeast genome. *J. Mol. Biol.* 288, 147–164
24 Makarova, K.S. and Grishin, N.V. (1999) The Zn-peptidase superfamily: functional convergence after evolutionary divergence. *J. Mol. Biol.* 292, 11–17

25 Kuriyan, J. *et al.* (1991) Convergent evolution of similar function in two structurally divergent enzymes. *Nature* 352, 172–174
26 Petsko, G.A. (1991) Déjà vu all over again. *Nature* 352, 104–105
27 Chen, Z.W. *et al.* (1994) The structure of flavocytochrome-c sulfide dehydrogenase from a purple phototrophic bacterium. *Science* 266, 430–432
28 del Cardayre, S.B. and Davies, J.E. (1998) *Staphylococcus aureus* coenzyme A disulphide reductase, a new subfamily of pyridine nucleotide-disulphide oxidoreductase. *J. Biol. Chem.* 273, 5752–5757
29 Jia, J. *et al.* (1996) Crystal structure of transaldolase B from *Escherichia coli* suggests a circular permutation of the α/β barrel within the class I aldolase family. *Structure* 4, 715–724
30 Pujadas, G. and Palau, J. (1999) TIM barrel fold: structural, functional and evolutionary characteristics in natural and designed molecules. *Biologia (Bratisl.)* 54, 231–253
31 Farber, G.K. and Petsko, G.A. (1990) The evolution of α/β barrel enzymes. *Trends Biochem. Sci.* 15, 228–234
32 Jurgens, C. *et al.* (2000) Directed evolution of a $(\beta\alpha)_8$-barrel enzyme to catalyze related reactions in two different metabolic pathways. *Proc. Natl. Acad. Sci. U. S. A.* 97, 9925–9930
33 Altamirano, M.M. *et al.* (2000) Directed evolution of new catalytic activity using the α/β-barrel scaffold. *Nature* 403, 617–621
34 Thornton, J.M. *et al.* (2000) From structure to function: approaches and limitations. *Nat. Struct. Biol.* (Suppl.), 991–994
35 Kraulis, P.J. (1991) MOLSCRIPT – a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.* 24, 946–950
36 Armstrong, R.N. (1997) Structure, catalytic mechanism, and evolution of the glutathione transferases. *Chem. Res. Toxicol.* 10, 2–18

Structural insight into substrate specificity and regulatory mechanisms of phosphoinositide 3-kinases

Snezana Djordjevic and Paul C. Driscoll

Phosphoinositide 3-kinases (PI3Ks) are implicated in a variety of fundamental cellular processes. These enzymes catalyse phosphorylation of the 3′-OH position of *myo*-inositol lipids that serve as secondary messengers. The catalytic subunit for one of the family members, PI3Kγ, has been structurally characterized, independently, in complexes with kinase inhibitors and with the p21^{Ras} GTPase. These atomic structures provide a basis for the rationalization of some PI3K substrate specificities and regulatory mechanisms, establishing links to functional and cellular data. Ongoing comprehensive structural and functional studies are essential to realize the promise of PI3K isozyme-specific therapeutic agents.

Snezana Djordjevic*
Bloomsbury Centre for Structural Biology and Dept of Biochemistry and Molecular Biology, University College London, London, UK WC1E 6BT.
*e-mail: snezana@biochem.ucl.ac.uk

Paul C. Driscoll
Ludwig Institute for Cancer Research, 91 Riding House Street, London, UK W1W 7BS; and Bloomsbury Centre for Structural Biology and Dept of Biochemistry and Molecular Biology, University College London, London, UK WC1E 6BT.

Looking at the three-dimensional structure of a famous biological macromolecule for the first time is akin to associating a face with the name of somebody one has heard so much about, but never met in person. Therefore, it was with great excitement that biologists working in the field of signal transduction gained the opportunity to see the structure of a key signalling enzyme that had, for a long time, eluded experimental description at this level: phosphoinositide 3-kinase (PI3K). This review focuses on three separate reports from the laboratory of Roger Williams and his co-workers in Cambridge [1–3]. These reports describe specific structural aspects of the catalytic subunit of the γ isoform of PI3K (p110γ), and its functional properties, providing long-awaited insights into the *modus operandi* of this pivotal lipid kinase. At the same time, these advances have highlighted several old and some new questions pertaining to the lipid-substrate specificity of different classes of PI3K, and to the nature of the subtle protein- and lipid-borne regulatory plasticity of these signalling molecules.

There are three major classes of PI3Ks, denoted Class I–III; Class I is also subdivided into Ia and Ib subsets. [An additional set of more distantly related molecules constitutes a fourth class in the PI3K 'superfamily', denoted Class IV. These molecules are not known to possess lipid-kinase activity but are Ser/Thr kinases. Examples include the mammalian target of rapamycin (mTOR/FRAP) and the catalytic subunit of DNA-dependent protein kinase]. The different classes of PI3K catalyse phosphorylation of the 3′-OH position of phosphatidyl *myo*-inositol (PtdIns) lipids, generating different 3′-phosphorylated lipid products that act as secondary messengers. PI3Ks are implicated in fundamental biological processes, including cell survival and proliferation, cell motility and adhesion, cytoskeletal rearrangement and vesicle trafficking [4,5]. There is increasing evidence that the different PI3K isoforms have cell-specific functions. For example, genetic manipulation of Class Ia PI3K in the fruit fly *Drosophila melangaster* has revealed a specific role for the enzyme in the control of growth [6,7]. Overexpression of PI3K leads to increased organ (e.g. wing) size, whereas expression of dominant negative forms leads to the diminution of organ size. Similarly, overexpression of a constitutively active Class Ia PI3K in the developing mouse embryo leads to an enlarged heart, whereas a dominant negative variant yields a small heart [8]. Macrophages and neutrophils from Class Ib PI3K knockout mice show reduced motility in *in vitro* and *in vivo* models of inflammation, as well as an impaired respiratory burst [9–11]. Class Ib PI3K-deficient mice also show resistance to thromboembolism [12].In addition, it has been noted that the lipid products resulting from PI3K

 0968-0004/02/$ – see front matter PII: S0968-0004(02)02136-9