Structure, Vol. 10, 1435–1451, October, 2002, ©2002 Elsevier Science Ltd. All rights reserved. PII S0969-2126(02)00861-4

Sequence and Structural Differences between Enzyme and Nonenzyme Homologs

Annabel E. Todd,[1] Christine A. Orengo,[1] and Janet M. Thornton[1,2,3,4]
[1]Biochemistry and Molecular Biology Department
University College London
Gower Street
London WC1E 6BT
United Kingdom
[2]Crystallography Department
Birkbeck College
Malet Street
London WC1E 7HX
United Kingdom

Summary

To improve our understanding of the evolution of novel functions, we performed a sequence, structural, and functional analysis of homologous enzymes and nonenzymes of known three-dimensional structure. In most examples identified, the nonenzyme is derived from an ancestral catalytic precursor (as opposed to the reverse evolutionary scenario, nonenzyme to enzyme), and the active site pocket has been disrupted in some way, owing to the substitution of critical catalytic residues and/or steric interactions that impede substrate binding and catalysis. Pairwise sequence identity is typically insignificant, and almost one-half of the enzyme and nonenzyme pairs do not share any similarity in function. Heterooligomeric enzymes comprising homologous subunits in which one chain is catalytically inactive and enzyme polypeptides that contain internal catalytic and noncatalytic duplications of an ancient enzyme domain are also discussed.

Introduction

The wealth of biological data now available has revealed the prolific evolutionary adaptation of old proteins for new functions. Indeed, evidence suggests that there is a limited number of protein folds in nature, perhaps as few as one thousand [1]. Ancestral genes have been duplicated, mutated, and combined through evolution to generate the multitude of functions necessary for life. An understanding of these adaptations and their functional consequences is essential for both genome analysis and protein design.

The evolutionary origin of extant enzymes has been discussed in detail [2–7]. Many enzyme superfamilies are particularly promiscuous in terms of function [8–12], and, commonly, both the nature of bound substrates and the reaction catalyzed varies between family members [7]. Typically, these variations have evolved through incremental modifications of one or both of the catalytic and substrate binding sites and through changes in modular content. Discussion of homologous enzymes and nonenzymes, however, has been much more limited [7, 12–16]. How many families contain both enzymes and nonenzymes? Is it common? How have they evolved? Have these nonenzymes lost the catalytic properties of their enzyme homologs? Or have the enzymes evolved from nonenzymes through the fortuitous recruitment of catalytic groups? Do they share any similarity in function, such as the same ligand binding capacity?

Perhaps the best-known example of a homologous enzyme and nonenzyme pair is that of C-type lysozyme and α-lactalbumin. They share over 35% sequence identity [17, 18], but α-lactalbumin is catalytically inactive, owing to the mutation of critical catalytic residues in the active site. They differ completely in function; lysozyme hydrolyzes bacterial cell wall polysaccharides, while calcium binding α-lactalbumin regulates the substrate specificity of galactosyltransferase in milk production.

To understand the molecular basis of functional differences of enzymes and their catalytically inactive homologs, we must have knowledge of protein three-dimensional structures, since the function of a protein is critically defined by its fold. The fold reveals binding sites, interaction surfaces, and the precise spatial relationships of functional groups. In this work, we focus on homologous enzyme and nonenzyme proteins in the Protein Data Bank [19]. We do not attempt to provide a comprehensive list, but we have included all superfamilies in the Protein Data Bank that we know to contain both enzymes and nonenzymes. As far as is known, the nonenzymes lack enzyme activity, unless stated otherwise, and, for all protein pairs, strong sequence, structural, and/or functional evidence supports an evolutionary relationship between them.

Results

Multifunctional Genes

Gene recruitment, or gene sharing, refers to the acquisition of a new function by an existing gene product, rendering the protein multifunctional. This evolutionary strategy is exemplified by the recruitment of enzymes as crystallins, structural proteins in the eye lens [20], where this second, noncatalytic role has been acquired by modifications in gene expression. A number of other genes have both catalytic and noncatalytic roles, and, often, the two functions share absolutely no similarity. These so-called "moonlighting" proteins and the mechanisms for switching between functions have been reviewed elsewhere [21]. The multifunctionality of other genes may be attributed to posttranslational modifications, alternate splicing, and alternative translation initiation (see Table 1).

[3]Correspondence: thornton@ebi.ac.uk
[4]Present address: European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SD, United Kingdom.

Key words: diversity; evolution; function; superfamilies; three-dimensional structure

Table 1. Enzymes and Nonenzymes Derived from the Same Gene

Enzyme	Additional Nonenzyme Function	Mechanism	Reference
Lactate dehydrogenase	Lens crystallin	Differential expression	[58]
PutA proline dehydrogenase	Transcriptional repressor	Differential localization	[59]
Thymidine phosphorylase	Platelet-derived endothelial cell growth factor	Inside and outside the cell	[60]
Aconitase	Iron-responsive element binding protein	Changes in ligand concentration	[61]
Protein disulphide isomerase	Subunit of triglyceride transfer protein complex	Complex formation	[62]
Human α-enolase	MPB1, *c-myc* promoter binding protein	Alternative translation initiation	[63]
Sindbis virus serine proteinase	Viral capsid protein	Posttranslational modification (autocatalytic cleavage)	[64]

The penultimate column indicates the mechanism for changing between the two functions. Note that, in the first five examples, the proteins that perform the enzyme and nonenzyme functions are identical (this is referred to as gene recruitment; for an excellent review see [21]), whereas, in the last two examples, the functions are carried out by nonidentical proteins that are, nevertheless, derived from the same gene.

It is thought that the multifunctionality of genes results in constraints on adaptability. Subject to adaptive pressures, such genes may undergo gene duplication and divergence to allow for the independent specialization of each function. In the structural analysis that follows, multifunctional genes, as far as is known, have not been included, unless stated otherwise. We consider only pairs of related proteins that have evolved by gene duplication and divergence.

Enzyme and Nonenzyme Homologs

Figure 1 illustrates the distribution of the pairwise sequence identities of the closest enzyme relatives of nonenzymes. For 98% of those nonenzymes having one or more enzyme homologs, the sequence identity of the closest relative lies below 50%, and, for the majority, the sequence identity is below 20%. The existence of homologous enzymes and nonenzymes at high-sequence identity levels appears to be extremely uncommon.

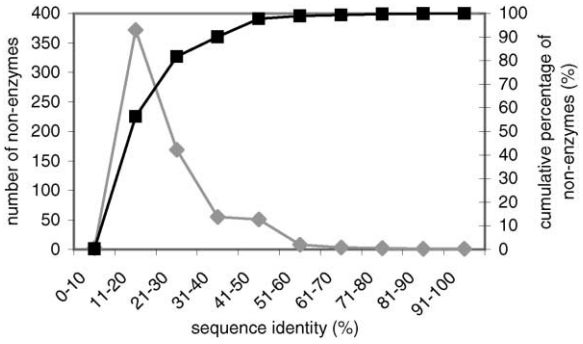

Figure 1. Distribution of the Pairwise Sequence Identities of the Closest Enzyme Relatives of Nonenzymes

Only single-domain proteins are considered. The gray line graph illustrates the number of closest enzyme relatives that fall within each level of sequence identity, and the black line graph is a cumulative percentage of closest relatives with increasing sequence identity. All single-domain proteins in the CATH structural classification [53, 54] were used as probes for PSI-BLAST [55] sequence searches. For each nonenzyme in the expanded classification (structural and SWISS-PROT [56] sequence data), its closest enzyme having one or more EC numbers assigned was identified. Of 3642 nonidentical nonenzymes in the classification, 664, contained in 29 homologous superfamilies, have one or more enzyme relatives.

Table 2 provides a list of homologous enzymes and nonenzymes of known three-dimensional structure. Table 3 indicates the frequency of occurrence of particular similarities and changes in these proteins, such as the loss of catalytic residues or the sharing of a common binding property.

Consistent with the results presented in Figure 1, sequence similarity is generally insignificant between the proteins listed. A notable exception is the 90% sequence identity shared between duck δ crystallin II, which has argininosuccinate lyase activity, and turkey δ crystalline, which lacks enzyme activity. This analysis is limited to the structural data, however. Some nonenzymes listed may be more closely related to enzymes not contained in the Protein Data Bank and vice versa but are nevertheless in the sequence databases. For example, human transferrin receptor protein of the Zn peptidase superfamily shares 24% sequence identity with membrane glutamate carboxypeptidase of the same species [22], but the latter has unknown structure.

How Have They Evolved?

Figure 2 illustrates two alternative scenarios for the evolution of these proteins; the nonenzymes have evolved from enzyme precursors, or, conversely, the enzymes have evolved from noncatalytic precursors. The direction of evolution is usually apparent from the nature of the superfamily and its members and through phylogenetic analyses. For example, fumarase/aspartase-like turkey δ crystallin is the only nonenzyme in a superfamily which otherwise contains a rich variety of enzymes. As such, it is the "odd one out" and it appears to have evolved from an ancestral enzyme. Evolutionary data and suppositions regarding the origin of the proteins considered in this analysis have been extracted from the literature, and the relevant papers are cited. We have not attempted to conduct phylogenetic analyses of our own. Given all the examples of homologous enzymes and nonenzymes that we identified in the Protein Data Bank, it appears that the loss of enzyme activity and the acquisition of a noncatalytic function (12 examples) is a more common scenario than the design of a catalytic function on a nonenzyme precursor (5 examples) [16].

Table 2. Enzyme and Nonenzyme Homologs

Family Name	Enzyme(s)	PDB	Nonenzyme(s)	PDB	ID	Basis for Loss/Gain of Activity	Functional Similarity
Enzyme to Nonenzyme							
Metal-dependent vicinal oxygen chelate superfamily	Glyoxalase I and other members	1qip	Bleomycin resistance protein (BRP)	1byl		Loss of metal binding site	*Substrate binding (putative)* BRP contains a colocated substrate binding site. BRP sequesters the antibiotic bleomycin, and it could have evolved from a metalloenzyme involved in the synthesis of bleomycin or a related compound [65].
Trypsin-like serine proteases	Neutrophil elastase	1ppf	Heparin binding protein (HBP)	1ae5	44%	Triad nucleophile: S195G Proton shuttle: H57S Charge relay: (D102)	*None* Human HBP is involved in host defence during infections and inflammations and it binds heparin and lipopolysaccharides.
TIM barrel (trans)glycosidases	Hevamine	2hvm	Concanavalin B	1cnv	41%	Putative intermediate stabilizing role: (D125) General acid: E127Q Disruption of substrate binding site	*Carbohydrate binding (putative)* A cluster of four aromatic residues, two of which are involved in *cis*-peptide bonds, a rarity for non-Pro residues, is present and is absolutely conserved across the family. This has been implicated in carbohydrate binding in hevamine, so concanavalin B may have a short saccharide binding function [26].
	Chitinase A	1edq	Lectin Ym1	1e9l	26%	General acid: E315Q Nucleophile: (D391)	*Carbohydrate binding*
	Chitinase A and other family 18 members[a]	1edq	Narbonin	1nar		General acid of catalytic family 18 members is present, but it is incorporated in a salt bridge, rendering the active site inaccessible to the oligosaccharides bound by its homologs [24–26].	*Carbohydrate binding (putative)* May have a short saccharide binding function (see above).
Lysozyme-like	Lysozyme	1jsf	α-lactalbumin	1b9o	37%	Acid/base: E35T Nucleophile: D53E	*None* Lysozyme hydrolyzes bacterial cell wall polysaccharides, and Ca binding α-lactalbumin, a regulatory subunit of lactose synthase, plays a substrate regulatory role in milk production.
Fumarase/aspartase	Duck argininosuccinate lyase/δ crystallin II	1auw	Turkey δ crystallin	—	90%	Disruption of substrate binding site Catalytic residues of bifunctional duck δ crystallin II are conserved in turkey δ crystalline, but the latter lacks enzyme activity. Most sequence variation occurs in the N-terminal region, which is involved in substrate binding in the catalytically active crystallin [66].	*None* (They have identical nonenzyme functions.)

(*continued*)

Table 2. Continued

Family Name	Enzyme(s)	PDB	Nonenzyme(s)	PDB	ID	Basis for Loss/Gain of Activity	Functional Similarity
Zn peptidases	Bacterial leucine aminopeptidase and other members	1amp	Transferrin receptor protein	1cx8		Lacks three of the five dinuclear Zn binding residues, as well as the catalytic base	*None*
His-Me finger homing endonuclease-like	Endonuclease I-*Ppo*I	1a73	Smad transcription regulator, MH1 domain	1mhd		R61- General base: H98A Mg^{2+} cofactor binding: N119-	*DNA binding* Both bind DNA by way of a β hairpin in the major groove.
Thioredoxin-like	Protein disulphide isomerase/ thioredoxin-like[b]	1mek	Calsequestrin	1a8y		Redox active CXXC motif is absent	*None* Calsequestrin has a Ca storage function in muscle.
			Phosducin	2trc		Redox active CXXC motif is absent	*None* Phosducin plays a regulatory role in dark/light adaptation and forms a complex with the β and γ components of transducin.
Ferritin-like di-iron carboxylate proteins	Ribonucleotide reductase β subunit/ferritin-like[b]	1xik	DNA protection during starvation protein	1dps		Enzyme di-iron site is absent	*None*
Nonenzyme to Enzyme							
GroES-like	Alcohol dehydrogenase (ADH), catalytic domain, and other members	1deh	GroES	1aon		GroES and the catalytic domain of ADH probably evolved from an ancient GroES-like protein [29]. Extant ADHs are derived from the fusion of this ancestral module to a nucleotide binding Rossmann domain.	*None* Topologically equivalent loops, however, are involved in ligand binding in all members. A mobile loop in GroES is involved in GroEL binding, while the equivalent loop in ADH forms part of the active site and contacts the Rossmann domain [29, 30].
Periplasmic binding proteins type II	Thiaminase I Porphobilinogen deaminase	3thi 1pda	Periplasmic binding proteins type II and transferrins	1ixh		Periplasmic binding domains bind a wide variety of water-soluble ligands, such as mono- and oligosaccharides, amino acids, oligopeptides, sulphate, and phosphate. These two enzymes are likely to have independently evolved from an ancestral noncatalytic periplasmic binding domain with similar substrate binding preferences, via gene duplication and recruitment of catalytic groups in the binding cleft [67]. Their noncatalytic precursors are yet to be identified, however.	*Substrate binding (putative)*

(*continued*)

Table 2. Continued

Family Name	Enzyme(s)	PDB	Nonenzyme(s)	PDB	ID	Basis for Loss/Gain of Activity	Functional Similarity
Homeodomain-like AraC and λ integrase families	Bacteriophage P1 Cre recombinase, C-terminal domains	1crx	MarA transcriptional activator	1bl0		These distantly related proteins [16, 68] share two repeats of the homeodomain-like module containing the DNA binding helix-turn-helix motif (HHTH fold). In Cre recombinase and related enzymes, the first repeat appears to be involved primarily in DNA binding, while the second, in which the Tyr nucleophile is located, is adapted to a catalytic function [16]. Several topoisomerases possess a mechanism similar to that of Cre recombinase, having a Tyr nucleophile in an HHTH fold, but the structural location of the Tyr is very variable. The catalytic properties of HHTH domains in DNA breaking/rejoining enzymes may have arisen independently several times [16].	*DNA binding*
Nuclear transport factor 2-like	Scytalone dehydratase (SDH) Steroid Δ-isomerase	1std 1qjg	Nuclear transport factor 2 (NTF2)[c]	1oun		The enzymes do not share one catalytically essential residue in common [35], but NTF2 contains residues equivalent to a few of the catalytic groups of each enzyme, including the D31-H85 dyad of SDH. SDH serves a unique, nonvital function in its pathogenic host, whereas NTF2 proteins are more widespread, and its activity could have evolved by exploiting the catalytic potential of the dyad in an ancient nuclear transport factor [29].[a]	*None* SDH and NTF2, however, bind aromatic groups in equivalent hydrophobic cavities. In NTF2, this group is a Phe residue derived from a protein interaction partner, Ras-like GTPase Ran [69].
Cupredoxins	Nitrite reductase, L-ascorbate oxidase, and other multicopper oxidases	1aoz	Azurin and other blue copper electron transfer agents	1nwp		Multicopper oxidases have evolved via duplication and fusion of the cupredoxin fold (up to six repeats) through the appearance of new and different types of Cu sites, which are necessary for catalysis [31].	*Cu type I site for single electron transfer*
Further Examples							
Barwin-like endoglucanases	Endoglucanase V	2eng	Barwin	1bw3		Catalytic base: D10Y	*Polysaccharide binding* Barwin is a putative plant defense protein.
Nuclear transport factor 2-like	Scytalone dehydratase (SDH) Steroid Δ-isomerase	1std 1qjg	Naphthalene 1,2-dioxygenase (NDO) noncatalytic β subunit	1ndo		The C terminus fills the region equivalent to the enzyme active site cavity [36].	*None* The β subunit probably plays a structural role in the NDO $\alpha_3\beta_3$ hexameric complex [36].

(*continued*)

Table 2. Continued

Family Name	Enzyme(s)	PDB	Nonenzyme(s)	PDB	ID	Basis for Loss/Gain of Activity	Functional Similarity
Thioredoxin-like	Protein disulphide isomerase (PDI)	1mek	Thioredoxin	2trx	27%		*Dithiol/disulphide redox activity* Redox active Cys residues in the CXXC motif contained in PDI are conserved. Thioredoxin oscillates between dithiol and disulphide states, but it is not classified as an enzyme as such.
Ferritin-like di-iron carboxylate proteins	Ribonucleotide reductase β subunit and other members	1xik	Ferritin	2fha		The enzyme di-iron site is present in ferritin heavy chain, but it is missing in the light chain.	*Ferroxidase activity of di-iron site* This is necessary for the uptake and storage of iron as a crystalline mineral in the oligomeric ferritin complex.
Dicopper center-containing domain	Catechol oxidase	1bug	Hemocyanin[d]	1oxy		Hemocyanin from *L. polyphemus* has two additional domains fused to either side of the dicopper domain. A Phe residue in the N-terminal domain structurally aligns perfectly with the aromatic ring of a bound inhibitor in catechol oxidase. This residue blocks the access of any substrates, allowing hemocyanin to function as an oxygen transport protein [27].	*O_2 binding and activation of dicopper site*

The direction of evolution of most examples is apparent (see text), and these examples have been classified under "Enzyme to Nonenzyme" and "Nonenzyme to Enzyme," accordingly. The evolution of a few enzyme and nonenzyme pairs and their superfamilies is more complicated or else not clear, and these have been listed under "Further Examples." Some nonenzymes listed share significant sequence similarity with a particular enzyme of the same superfamily, and, for those pairs sharing at least 20% pairwise identity, the sequence identity is given. In some examples, groups of proteins are referred to, such as "glyoxylase I and other members," since the nonenzyme (enzyme) does not share significant sequence similarity with any particular enzyme (nonenzyme) of the same superfamily. An explanation for the loss/gain of catalytic activity was obtained from the literature, where possible, and structural alignments were generated using SSAP [70]. For a few examples, reasons for the loss of activity can, at least in part, be attributed to the substitution of critical residues in the active site, while, for others, the basis for differences in activity is not so well understood. Where a particular pair of proteins, rather than groups of proteins, is listed and specific active site substitutions are given, the residue numbering is the same as that in the Protein Data Bank of the enzyme, and the substituted residue corresponds to that found in the nonenzyme. "-" denotes an insertion in the structural alignment. Catalytic residues listed in parentheses are conserved in the nonenzyme. The superfamily names are taken from SCOP [71] or else from the literature. Note that the thioredoxin superfamily, ferritin-like di-iron carboxylate proteins and the nuclear transport factor 2-like superfamily all contain entries in different sections of the table.

[a] Family 18 refers to a particular sequence family in the classification of glycosyl hydrolases by Henrissat and coworkers [37, 38].

[b] Some members of the thioredoxin-like superfamily contain a redox active CXXC motif, where the N-terminal Cys forms a mixed disulphide bond with the substrate. These proteins include protein disulphide isomerase, thioredoxin, and glutaredoxin, but the last two are not classified as enzymes as such. Similarly, ferritin of the di-iron carboxylate protein superfamily is not classed as an enzyme, but it has the ferroxidase activity of its enzyme homologs (see text for more details). Thioredoxin and ferritin have been listed as nonenzymes under "Further Examples."

[c] Nuclear transport factor 2 (NTF2) has no reported enzyme activity. Given the presence of an Asp-His dyad that plays a catalytic role in the distant relative scytalone dehydratase, however, it may have an as yet unidentified enzyme function that serves to modulate the interaction of NTF2 with Ran [69, 72].

[d] Hemocyanins have been shown to function as phenoloxidases under certain conditions [73] that distort the protein matrix and permit substrate access.

Table 3. Frequency of Occurrence of Similarities and Changes in Enzyme and Nonenzyme Homologs

	Number of Examples
Enzyme and Nonenzyme Homologs	22
Nonenzymes that have evolved from enzyme precursors	12
Loss of catalytic residues	10
Steric block/structural disruption of active site or substrate binding pocket	2
Common binding property	2
Common binding property (putative)	3
No similarity in function	7
Enzymes that have evolved from nonenzyme precursors	5
Common binding property	4
Common binding property (putative)	1
No similarity in function	2
Further examples	5
Conservation of catalytic residues	3
Steric block/structural disruption of active site or substrate binding pocket	2
Common binding property	4
No similarity in function	1
Heterooligomers: Homologous Catalytic and Inactivated Subunits	6
Loss of catalytic residues	5
Inactivated subunit(s) has a regulatory role	3
Internal Duplication: Homologous Catalytic and Inactivated Domains	7
Loss of catalytic residues	6
Steric block/structural disruption of active site or substrate binding pocket	4
Inactivated domain(s) has a regulatory role	1

Loss of catalytic residues includes the loss of metal binding residues if the metal is involved in catalysis. The common binding property may refer to the identical binding of a metal ion, O_2, carbohydrate, DNA, or other ligand or simply to the ligand binding loop in GroES and alcohol dehydrogenase (see text). Some proteins putatively have an ancestor with the same ligand binding capacity, but it has yet to be identified (e.g., porphobilinogen deaminase that has evolved from a periplasmic binding protein; see Table 2); these cases are considered separately in the table. The direction of evolution for five enzyme and nonenzyme pairs and their superfamilies is more complicated or else not clear, and these are considered under "Further Examples" (see Table 2).

Nonenzymes that Have Evolved from Enzyme Precursors: Why Are They Inactive?

Table 2 provides an indication of the conservation and variation of those residues that play a functional role in the enzyme homolog. Ten of the 12 enzyme-derived nonenzymes owe their catalytic inactivity to the mutation of critical residues in the active site. Heparin binding protein, α-lactalbumin, and concanavalin B share high sequence similarity with enzyme homologs (37% or more). Their inactivity could be predicted from sequence alignment alone, owing to the mutation of functional residues, which are otherwise conserved. Heparin binding protein, for instance, lacks the His and Ser residues in the well-known catalytic triad of neutrophil elastase and other enzymes of the trypsin-like serine protease family [23] (see Figure 3A).

A few nonenzymes owe their inactivity, at least in part, to the disruption of the substrate binding site or to the steric block of the active site cleft. As far as is known, the TIM barrel glycosyl hydrolase-like protein narbonin

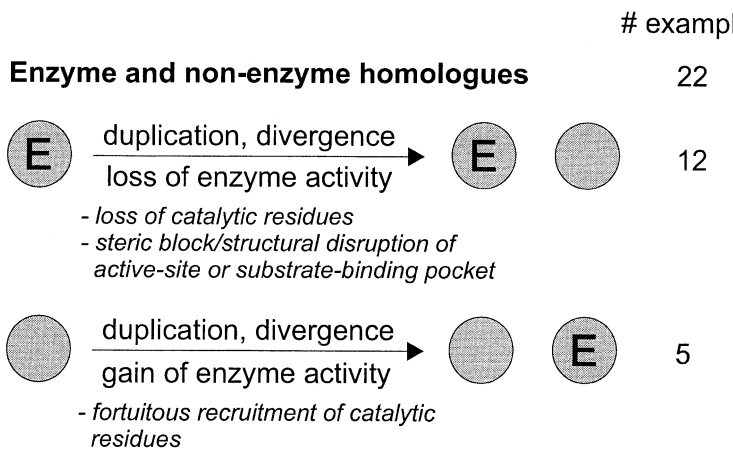

Figure 2. Homologous Enzyme and Nonenzyme Proteins and Evolutionary Paradigms

Circles labeled with and without the letter "E" denote enzymatic and nonenzymatic genes, respectively. These paradigms are discussed in the text. The direction of evolution for five enzyme and nonenzyme pairs and their superfamilies is more complicated or else not clear, explaining why the figures do not add up.

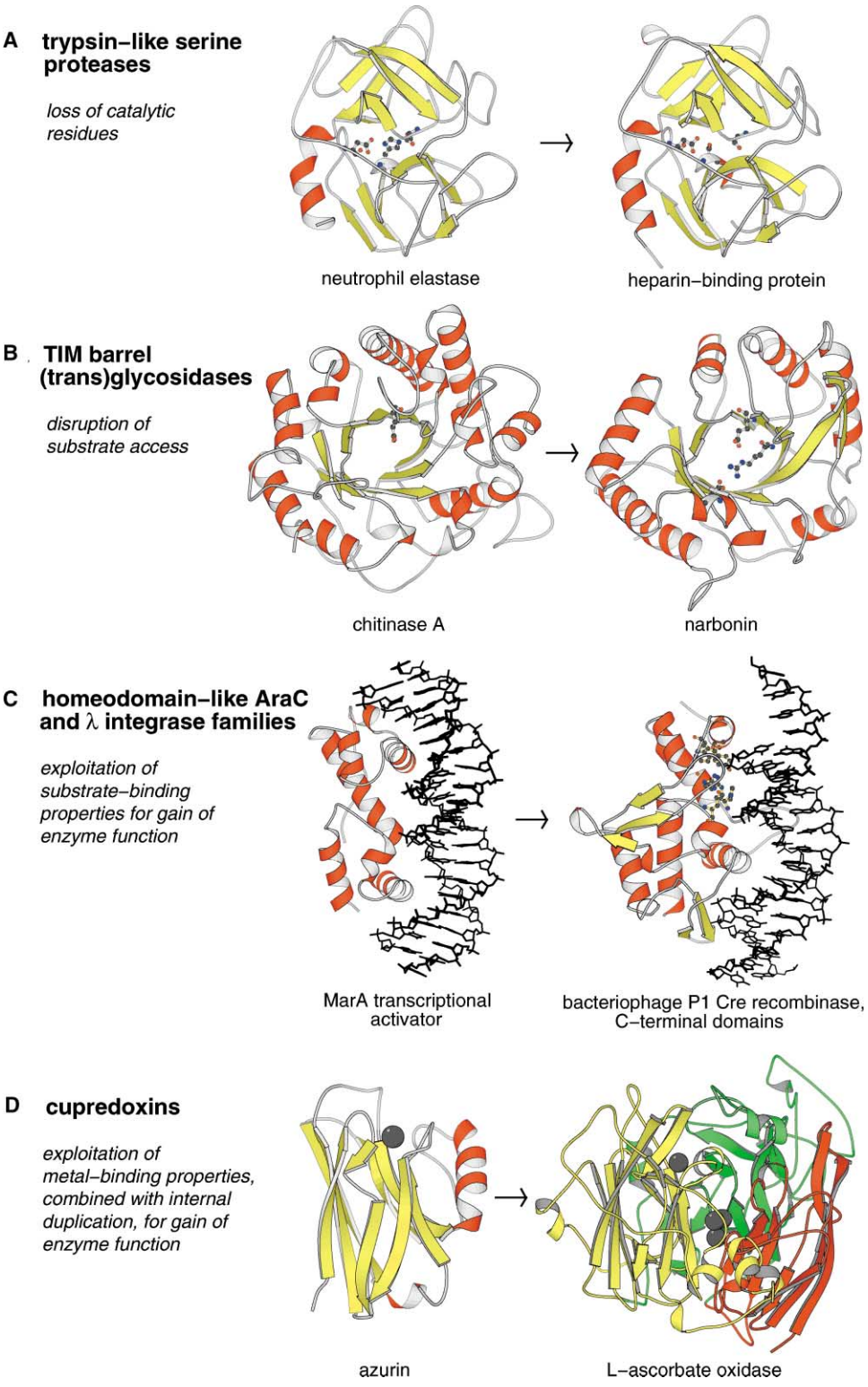

Figure 3. Examples of Homologous Enzyme and Nonenzyme Proteins

Very brief explanations for the basis of loss/gain of enzyme activity are provided. See Table 2 for further details. These diagrams were created with MolScript [57].

(A) Residues in the Ser-His-Asp catalytic triad in neutrophil elastase and the equivalent residues in heparin binding protein are shown in ball and stick representation.

(B) The Glu general acid in chitinase A and the equivalent Glu in narbonin, together with the Arg and Asp residues in narbonin with which the Glu is involved in a salt bridge, are shown.

(C) The Tyr nucleophile, Arg-His-Arg catalytic triad, and functional Trp in Cre recombinase are shown.

(D) The gray spheres represent bound Cu ions. The topmost Cu in L-ascorbate oxidase is equivalent to the single type I Cu ion in azurin. Each of the three cupredoxin domains in the oxidase is represented by a different color.

lacks catalytic activity, although the Glu general acid of catalytic family members is conserved. In narbonin, this Glu is incorporated in a salt bridge with a spatially adjacent Arg residue, rendering the active site inaccessible to the oligosaccharides bound by its homologs [24–26] (see Figure 3B). Another example is provided by hemocyanin. This protein is a distant homolog of catechol oxidase, and both bind and activate molecular oxygen in their dicopper centers. A Phe residue in the extra N-terminal domain of hemocyanin structurally aligns perfectly with the aromatic ring of a bound inhibitor in catechol oxidase. This residue blocks the access of any substrates, allowing hemocyanin to function as an oxygen transport protein [27].

What Is the Function of the Odd Ones Out? Is There Any Similarity to the Function of Other Family Members?

In only 12 of the 22 examples of enzyme and nonenzyme homologs, a similarity in their functions was identified (see Tables 2 and 3). Up to 14 examples share a common binding property, but, in two, this property is minor, so it is not included as a functional similarity (e.g., GroES and alcohol dehydrogenase).

Nonenzymes tend to have a binding role, interacting with metal ions, small ligands, proteins, or other biological macromolecules. In all five "nonenzyme to enzyme" examples in Table 2, nature appears to have exploited the specific binding properties of the catalytically inactive precursor (see Figures 3C and 3D).

Many enzymes have a nucleotide binding Rossmann domain in common [28]. By way of fusion of this ancient module to a variety of catalytic domains during evolution, new and specialized enzyme functions have evolved. The medium chain alcohol dehydrogenase-like enzymes, for example, have a Rossmann fold with a complex all-β structure fused to their N termini. The origin of this catalytic domain was unclear until the structural determination of GroES, with which it shares a distant evolutionary relationship [29, 30], and it has probably evolved from an ancient GroES-like protein [29]. Evidence for a common ancestry is provided by the conservation of a Gly-Asp dipeptide within the core and by other structural features. In particular, while these proteins share no similarity in function, they all use equivalent loops for ligand interactions; a mobile loop in GroES is involved in GroEL binding, while the equivalent loop in alcohol dehydrogenase forms part of the active site and contacts the Rossmann domain [29, 30].

The single-domain single-electron transfer agents, such as azurin and plastocyanin, are likely to most-closely represent the ancestor of the huge cupredoxin superfamily of Cu binding proteins [31]. Their Cu binding and electron transfer properties have been exploited for gain of enzyme function. The multicopper oxidases of this superfamily, such as L-ascorbate oxidase, have evolved through multiple gene duplication and fusion events of this ancestral cupredoxin domain and through the appearance of new and different types of Cu sites in the domain interfaces that are necessary for catalysis [31] (see Figure 3D).

Similarly, up to 5 of the 12 nonenzymes that are derived from ancient enzymatic domains have retained the ligand binding capacity of their catalytic precursors. For example, the DNA binding modes of endonuclease I-PpoI and the MH1 domain of a Smad transcription regulator are identical [32]; this appears to be one of several examples where an enzyme has been recruited to function as a eukaryotic transcription factor [13, 33]. In the evolution of nonenzymes from enzymes, however, it may be that the binding properties are more likely to change (7/12 examples) than in the reverse evolutionary event (0/5).

Some nonenzymes have the chemical functionality associated with the enzyme activity of their catalytic homologs, although they are not classified as enzymes as such, and some or all of the residues associated with this functionality are preserved. Hemocyanin and catechol oxidase and their oxygen activation function have been discussed. The iron storage protein ferritin of the di-iron carboxylate protein superfamily presents another example. Like ribonucleotide reductase and other oxidoreductase members, ferritin has a di-iron site within a four-helix bundle core and uses the ferroxidase activity of this site to allow storage of iron as an insoluble oxide in the central cavity of its oligomeric structure. The third example is thioredoxin and protein disulphide isomerase, both of which have dithiol/disulphide redox activity associated with a CXXC motif, and all three examples have been listed under "Further Examples" in Table 2.

In 10 of the 22 examples of homologous enzymes and nonenzymes, no similarity in function could be identified. For example, the biochemical function of most members of the thioredoxin superfamily involves sulfur redox chemistry. Redox-inactive phosducin and calsequestrin are exceptions. The former plays a regulatory role in dark/light adaptation and forms a complex with the β and γ components of transducin. The latter contains three thioredoxin motifs [34] and has a high-capacity Ca^{2+} binding function in muscle, which requires aggregation into a polymeric state.

The nuclear transport factor 2 (NTF2)-like superfamily also supports a variety of functions. NTF2 itself facilitates protein transport into the nucleus. The two catalytic members scytalone dehydratase and steroid Δ-isomerase bind different substrates, differ in activity, and do not share one catalytically essential residue in common [35]. The β subunit of naphthalene 1,2-dioxygenase (NDO) is a fourth family member. It plays no role in dioxygenase activity, and its C-terminal tail fills the region equivalent to the active site cavity of its enzyme homologs [36]. Instead, it probably plays a structural role in the NDO $\alpha_3\beta_3$ hexameric complex [36]. Clearly, the identification of other family members is necessary to shed light on the evolution of function within this superfamily.

Heterooligomers: Homologous Catalytic and Inactivated Subunits

During analysis we identified six enzymes that exist as heterooligomers comprising homologous subunits, but in which one or more subunits are catalytically inactive

Table 4. Heterooligomers: Homologous Catalytic and Inactivated Subunits

Family Name	Enzyme	PDB	Subunit Assembly and Activity	ID	Basis for Inactivity	Function of Inactivated Subunit(s)
N-terminal nucleophile hydrolases	20S proteasome	1pma	$(\alpha_7)_2(\beta_7)_2$ heterooligomer assembly comprises four seven-membered rings; β and α subunits form the two inner and outer rings, respectively; each subunit is a single-domain protein that adopts the Ntn hydrolase fold; α subunits are inactive.	26%	Like other family members, the β subunit uses its N-terminal residue (Thr) as a catalytic nucleophile and base in amide hydrolysis. The α subunit is inactive, owing to a 35-residue N-terminal extension.	The N-terminal extension in the α subunit acts as a molecular gate, blocking access of the protein substrate to the central channel of the oligomeric structure where catalysis takes place. Binding of a regulatory particle relieves the inhibition by opening the channel, to give the proteolytically active form of the complex [74, 75].
Phospholipase A2 (PLA2)	Vipoxin	1aok	Heterodimer Toxic PLA2 and nontoxic inactive Inhibitor	62%	Catalytic base: H48Q (Y52) (D99)	The inhibitor reduces enzyme activity and toxicity of PLA2, but, in its absence, PLA2 becomes unstable, losing both its activity and toxicity. The inhibitor provides stability to maintain toxicity at the expense of phospholipase activity [76, 77].
Bacterial luciferase-like	Luciferase	1brl	αβ heterodimer Active site exists only in the α chain.	32%	Phosphate ($FMNH_2$) binding: R107D Putative base, essential: H44N	The β chain is essential for high catalytic efficiency and may stabilize the α chain for catalysis [78], but its mechanism of function is not completely understood.
Cobalamin (vitamin B12)-dependent enzymes	Methylmalonyl-CoA mutase	1req	αβ heterodimer The α chain is catalytic, but the β chain is inactive; both chains contain two domains, a TIM barrel, and an α/β domain that bind CoA and the adenosylcobalamin cofactor, respectively, in the α chain.	24%	The β subunit lacks most of the residues involved in cofactor and substrate binding in the α subunit, its C terminal domain is swung away from the TIM barrel, no active site is formed [40].	The β chain has no obvious function, although it appears to be essential for activity and contributes one residue to substrate binding. It may be an evolutionary relic or else may have some as yet unidentified function [40].
Class II aminoacyl-tRNA synthetases (aaRS) and biotin synthetases	Phenylalanyl-tRNA synthetase	1b70	$(\alpha\beta)_2$ heterotetramer The α subunit contains only one catalytic domain, while the β subunit contains six domains; the fifth domain is homologous to the catalytic domain of the α subunit but lacks activity.	23%	Functional residues invariant in the α chain are mutated in the β subunit.	The β chain is required for catalysis. The extra domains of this subunit have been recruited for tRNA recognition and binding [39].
Ferritin-like di-iron carboxylate proteins	Methane monooxygenase (MMO), hydroxylase component	1mhy	$\alpha_2\beta_2\gamma_2$ hexamer The α and β subunits are homologous, but only the α chain is catalytic.		The β subunit lacks four of the six residues in the α chain di-iron site.	MMO comprises three protein components, the hydroxylase, reductase, and B components. The reductase component forms a specific complex with the hydroxylase β subunit [79], and this affects the relative redox potentials of the two iron atoms, so that two-electron transfer is favored [80]. Thus, together, these two proteins exert a regulatory effect on catalysis.

The table lists those enzyme complexes comprising nonidentical, but homologous, subunits in which one or more subunits are catalytically inactive. In all cases identified, evidence strongly suggests that the inactive subunits have evolved from enzymatic ancestors and not the reverse scenario, where the catalytic subunits have gained activity during the course of evolution. The sequence identity of pairs of subunits that share more than 20% sequence identity is given. An explanation for the loss of catalytic activity was obtained from the literature, where possible; for a few examples, the reasons for the loss of activity are obvious, such as the substitution of critical residues in the active site, while, for others, the basis for differences in activity is more subtle or else are not as well understood. Where specific active site substitions are given, the residue numbering is the same as that in the Protein Data Bank of the active subunit, and the substituted residue corresponds to that found in the inactivated subunit. Catalytic residues listed in brackets are conserved in the inactivated subunit.

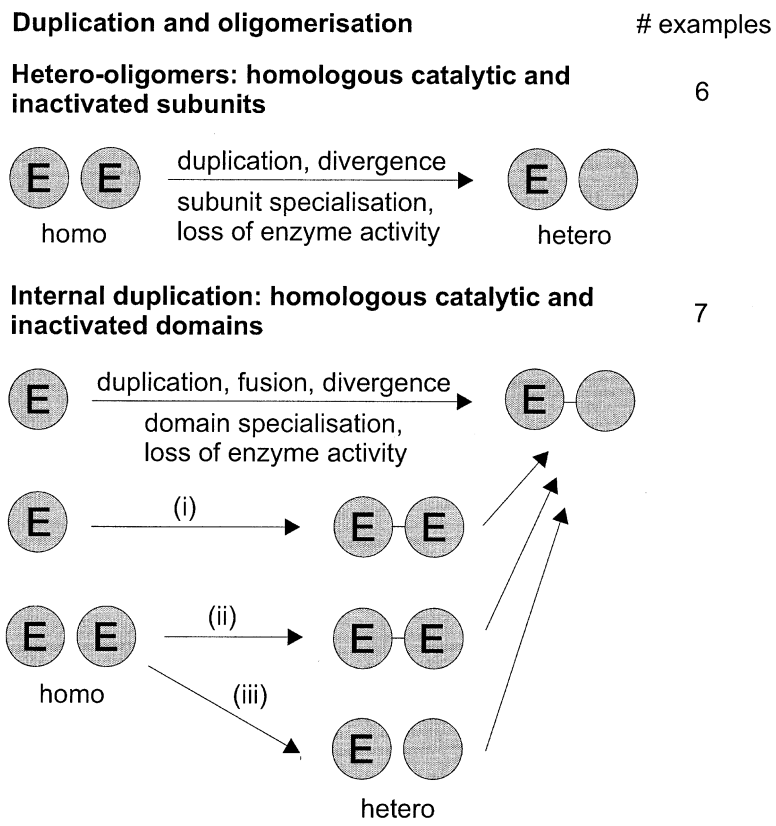

Figure 4. Oligomerization, Internal Domain Duplication, and Evolutionary Paradigms

Circles labeled with and without the letter "E" denote enzymatic and nonenzymatic genes, respectively. These paradigms are discussed in the text.

(see Table 4). Interestingly, two of the six examples listed correspond to nonhomologous TIM barrel proteins, bacterial luciferase and bacterial methylmalonyl-CoA mutase. The TIM barrel motif is the most common fold identified in enzymes [14], and only very rarely does it assume a nonenzymatic role. The few other noncatalytic TIM barrels include narbonin, concanavalin B, and lectin Ym1 of the glycosyl hydrolase family 18 [37, 38] (see Table 2). Sequence similarity is significant between all homologous subunits, with the exception of the α and β subunits of the hydroxylase component of methane monooxygenase.

In all cases identified, the evidence strongly suggests that the inactive subunits have evolved from enzymatic ancestors and not from the reverse scenario, where the catalytic subunits have gained activity during the course of evolution. For at least three examples, there is evidence to indicate that the heterooligomeric enzymes have evolved from homooligomeric precursors (see Figure 4). This direction of evolution (homo to hetero) seems more likely, as more complex and specialized systems have evolved.

Vipoxin from the Bulgarian sand viper (*Vipera ammodytes meridionalis*) functions as a heterodimer, one subunit having phospholipase A2 activity while the other acts as its inhibitor. Phospholipase A2 enzymes of evolutionarily "older" snakes function as homodimers. Bacterial methylmalonyl-CoA mutase functions as an $\alpha\beta$ heterodimer, whereas the human enzyme is an α_2 homodimer and shares high sequence identity with the bacterial catalytic α chain. Lastly, class II aminoacyl-tRNA synthetases function as $(\alpha\beta)_2$ heterotetramers, α_4 homotetramers, or α_2 homodimers. The heterodimeric interface formed by the catalytic-like domain of the inactive β and catalytic α chains in phenylalanyl-tRNA synthetase is similar to that observed in the α_2 homodimers of other class II aminoacyl-tRNA synthetases [39].

For two examples the function of the inactive subunit is not completely understood. However, in almost, if not, all cases, the heterooligomeric enzyme complex, containing both the catalytic and noncatalytic subunits, is required for full activity, although the inactive subunit does not contribute to the active site in any way (with the exception of methylmalonyl-CoA mutase, in which the β subunit contributes just one residue to the substrate binding site [40]). Presumably, the individual components within each heterooligomer have coevolved to optimize the interactions at their subunit interfaces. If the oligomeric state is necessary for full structural stability and, therefore, catalytic activity, all components, whether catalytic or inactive, will be required for optimal catalysis. Interestingly, in three of the six examples, the inactivated subunit has evolved a regulatory role in the enzyme complex: 20S proteasome, vipoxin, and methane monooxygenase (see Table 4).

Internal Duplication: Homologous Catalytic and Inactivated Domains

We also identified seven enzyme complexes that contain two or more domain repeats within a single polypeptide

Table 5. Internal Duplication: Homologous Catalytic and Inactivated Domains

Family Name	Enzyme	PDB	Domain Organization and Activity	ID	Basis for Inactivity	Function of Inactivated domain(s)
Vicinal oxygen chelate superfamily	Extradiol dioxygenase	1han	Two duplicated domains The N domain is inactive.		Fe binding: H146Y, H210F, (E260) Two of the three Fe binding residues are lost. The third ligand is involved in a salt bridge, which, together with large side chains in the vestigial active site, blocks substrate access [42].	Unknown
Actin-like ATPase domain	Hexokinase I	1qha	Two duplicated domains The N domain is inactive.	30%	Residues implicated in catalysis are conserved in the inactive N domain, and this domain is able to bind the substrate and product of the reaction. Its inactivity is probably due to the unproductive binding of the phosphoryl donor, ATP, in a site distant from the vestigial active site.	Hexokinase I uses ATP to phosphorylate glucose and is strongly inhibited by the product, glucose 6-phosphate. The N domain is involved in the allosteric phosphate-induced relief of product inhibition [47].
Cytidine deaminase	Cytidine deaminase	1af2	Two duplicated domains The C domain is inactive.		Zn binding: H102P, C129L, C132A Multiple catalytic roles: E104L An extra loop in the C terminus disrupts the region that structurally corresponds to the cofactor Zn binding site in the N domain and precludes access by ligands [41].	The C domain of one subunit acts as a lid on the active site of the second subunit in the dimer, sequestering the substrate from solvent [41].
Rhodanese/cell cycle control phosphatase	Rhodanese	1e0c	Two duplicated domains The N domain is inactive.	21%	Persulphide linkage: C230D	The N domain does not directly contribute to the active site, but its presence in all known rhodanese and rhodanese-like proteins stresses its significance [81].
ADP ribosylation	ADP-ribosylating toxin	1qs1	Heterooligomer of VIP1 and VIP2 VIP2 contains two duplicated domains. The N domain is inactive.	22%	Catalytic H bonding role: E428K "STS motif," which is conserved in other toxins and stabilizes the NAD binding pocket in the C domain, is missing in the N domain. Two strands in the C domain block the central cleft of the first domain, thus impeding NAD binding [43].	The N domain interacts with the nonenzyme VIP1 component [82], which targets insect cells.

(*continued*)

Table 5. Continued

Family Name	Enzyme	PDB	Domain Organization and Activity	ID	Basis for Inactivity	Function of Inactivated domain(s)
Protein tyrosine phosphatases (PTP)	Protein tyrosine phosphatase LAR	1lar	14 domains in total, including 2 duplicated domains (D1 and D2) at the C terminus D1 is active, but D2 is inactive.	40%	General acid: D1490E The substitution of two highly conserved residues introduces subtle structural variations within the active site and probably accounts for the inactivity of D2. Active PTPs utilize an Asp general acid located in a WPD loop. Upon substrate binding the loop undergoes a large conformational change, moving towards the active site, to create a catalytically competent enzyme [83]. In D2 the Asp is conservatively substituted for Glu, but this forms a hydrogen bond to a main chain amide group of a Met in an adjacent, less-conserved loop. Combined with the steric hindrance of the Met side chain, the energy required for the substrate-induced shift of the Glu-containing loop is probably much higher in inactive LAR D2 [44]. The second substitution is in the KNRY loop, which recognizes the phospho-Tyr group of the substrate, where the Tyr is substituted for Leu. The substitution of Tyr affects activity, even though this residue does not play a direct catalytic role. In the active Tyr-containing D1 domain, the Tyr is stabilized by a hydrogen bond with a Ser residue that is adjacent to the essential thiol-phosphate-forming Cys nucleophile. The Leu substitution causes a shift in the position of Ser, which blocks the access of phospho-Tyr substrates to the Cys [44]. Reversing the two substitutions (Glu to Asp and Leu to Tyr) gives a mutant protein having the same level of activity as that of PTP LAR D1.	The function of D2 is unknown, but it may affect substrate specificity [46].
Thioredoxin-like	Protein disulphide isomerase	1mek 1bjx[a]	Four duplicated domains The central two domains are inactive.		Redox active: C36E C39S	The two central domains stimulate catalytic activity of the protein, and this may result from the binding of protein substrates [45].

The table lists those enzymes containing domain repeats within a single subunit and in which one or more of these domains are catalytically active, while the remainder are inactive. In all cases identified, the evidence strongly suggests that the inactive domains have evolved from enzymatic ones and not the reverse scenario, where the catalytic domains have gained activity during the course of evolution. See the legend to Table 4 for more details on the contents of this table. N domain, the N-terminal domain; C domain, the C-terminal domain.

[a]1mek and 1bjx are the Protein Data Bank entries for the first (catalytic) and second (noncatalytic) domains, respectively, of human protein disulphide isomerase. The structures of its two C-terminal domains are unknown.

chain. One or more domain duplicates harbor the active site, while the remainder has lost catalytic capacity (see Table 5).

Figure 4 illustrates three alternative routes in the evolution of these enzymes. In routes "i" and "ii," with two (or more) identical domains fused onto the same subunit, one duplicate is free to explore other functional possibilities through incremental mutations, while the catalytic apparatus of another remains intact. Alternatively, they could have evolved from heterooligomeric (and homooligomeric) precursors by route "iii." The fusion of two different, nevertheless homologous, subunits onto the same polypeptide chain ensures that they are transcribed together and function in tandem. Cytidine deaminase from *E. coli* may have evolved from a homooligomeric precursor [41]. It exists as a homodimer, where each subunit contains two duplicated domains, one catalytic and one inactive. The enzyme from *B. subtilis*, meanwhile, functions as a homotetramer, where each subunit contains just one catalytic domain.

In addition to the substitution of critical residues, the noncatalytic domains in four of the seven enzymes are inactive for steric reasons—a recurring theme in enzyme inactivation. In extradiol dioxygenase, substrate access is blocked, owing to the participation of a potential metal ligand in a salt bridge and to large side chains in the vestigial active site [42]. In cytidine deaminase, ligand access is precluded by an extra loop in the inactive C-terminal domain [41]. In ADP-ribosylating toxin, two strands in the second domain block the central cleft of the first domain, thus impeding NAD binding [43]. In protein tyrosine phosphatase LAR, the substitution of two highly conserved residues (one of them conservative, Asp to Glu) in the C-terminal domain introduces very subtle structural variations in the active site and probably accounts for its inactivity [44].

Although the divergence of function is accompanied with a loss of catalytic activity in one or more duplicated domains in these proteins, substitutions may amplify or refine the activity of the catalytic domain by enhancing substrate binding, for example. The central two thioredoxin-like motifs in protein disulphide isomerase lack the catalytic CXXC motif; nevertheless, they stimulate the isomerase activity of the protein, and this may result from the binding of protein substrates [45]. The function of the C-terminal domain of protein tyrosine phosphatase LAR is unknown, but it may affect substrate specificity [46]. The N-terminal domain of hexokinase I plays a regulatory role [47].

Discussion

An understanding of protein structure and function relationships is central to the success of the structural genomics initiatives, which aim to provide a structural representative for all homologous protein families. Given an uncharacterized structure and its relatives in the Protein Data Bank, structural biologists must anticipate the functional significance of structural similarities and variations between these proteins. In this work we have sought to understand the molecular basis for the functional differences observed in homologous enzymes and nonenzymes of known three-dimensional structure, to identify any functional attribute shared by these homologs and to provide explanations for their evolutionary origins. The analysis provides valuable insights, but, of course, a more complete catalog of protein structures (within these superfamilies and in others) is essential to confirm the trends observed and preliminary conclusions drawn in this study.

While the evolution of enzymes is well documented, there has been much less focus on homologous enzymes and nonenzymes. Our work with both structural and sequence data [7] has indicated that the existence of enzymes and noncatalytic proteins within the same superfamily is quite common, but the sequence identity between them is typically very low. Other interesting examples not considered in this work, owing to lack of structural data, include *Drosophila* cell-cell adhesion neurotactin, which is related to the ubiquitous α/β hydrolases, but the catalytic triad is disrupted [48], quinate repressor of the TIM barrel aldolase superfamily [49], the DNA repair protein RAD4, which adopts the transglutaminase fold [33], and polyketide cyclases of the start-related lipid transfer (START) domain superfamily [50], in which most members have a noncatalytic ligand binding function.

The examples presented suggest that the evolution of a nonenzyme from a catalytic precursor is more common than the reverse scenario, that is, the design of a catalytic function on an ancient nonenzyme domain. This analysis is restricted to the structural data, however, and it is possible that this conclusion is, to some extent, a reflection of the bias in the Protein Data Bank toward enzyme superfamilies [14]. Most nonenzymes derived from enzyme ancestors have lost one or more of the critical catalytic residues within the active site. However, it is well to remember that, for distantly related proteins, the substitution of critical residues may present an oversimplified picture of the basis of inactivity, and focusing on these alone may even be inappropriate. Small conformational effects of residues distributed throughout the fold play a role in shaping the active site for substrate complementarity and efficient catalysis [51, 52]. The orientations of secondary structure elements can vary extensively between distant homologs, owing to multiple insertions, deletions, and substitutions throughout the structure, and a vestigial active site in an evolving nonenzyme can thus change in shape and size, whether the catalytic residues are in place or not.

A few nonenzymes illustrate particularly well the importance of structural data for the provision of clues to their catalytic inactivity. For instance, the participation of a potential active site residue in a salt bridge or the blocking of substrate access are just two ways in which enzyme functions have been inactivated during the course of evolution. Several heterooligomeric enzymes contain two types of homologous subunits, where one is inactive, and, similarly, several enzyme polypeptides contain internal catalytic and noncatalytic duplications of an ancient enzyme domain. Often these inactive domains or subunits contribute to the overall activity of the molecule, and a few serve an important regulatory function within the enzyme complex.

Many nonenzymes (enzymes) bear no similarity in

function to their ancestral enzyme (nonenzyme) precursors, and such relationships present a major challenge in the inference of function from sequence and structure. Indeed, with the proliferation of biological data over the last few years, the extent of evolutionary and functional diversification is only just being realized [7, 35]. The genome sequencing and structural genomics projects herald many more new and exciting discoveries of unanticipated evolutionary kinships between enzyme and nonenzyme proteins.

Biological Implications

The recent growth in biological data has revealed the functional versatility of many protein superfamilies. It is not uncommon for a single fold to support a variety of both enzyme and nonenzyme functions. Understanding the molecular basis for these functional variations is essential for protein design and for the success of the structural genomics projects in which we hope to derive functional information from the structures of uncharacterized proteins.

This sequence, structural, and functional analysis of enzyme and nonenzyme homologs provides new insights into nature's ways of adapting old proteins for new functions. Typically, active sites are disrupted by point mutations and more gross structural rearrangements in the adaptation of catalytic folds for nonenzyme functions; ancestral binding sites are usually exploited for gain of enzyme function; in some enzyme complexes having multiple copies of a catalytic ancestral domain, one or more duplicates have lost activity to nevertheless enhance or refine the overall activity of the protein. The functional similarities and differences in these proteins are often discernible only with the structural data, regardless of pairwise sequence identity. These results have implications in genome analysis in the prediction of function from sequence and structure.

Acknowledgments

A.E.T. acknowledges financial support from the Midwest Centre for Structural Genomics. This is a publication from the Bloomsbury Centre for Structural Biology. Owing to space restrictions, it has not been possible to cite many of the articles from which data was extracted for this analysis, and we acknowledge the experimentalists who have contributed to these publications.

Received: December 28, 2001
Revised: May 3, 2002
Accepted: May 14, 2002

References

1. Chothia, C. (1992). Proteins—1000 families for the molecular biologist. Nature *357*, 543–544.
2. Jensen, R.A. (1976). Enzyme recruitment in evolution of new function. Annu. Rev. Microbiol. *30*, 409–425.
3. Petsko, G.A., Kenyon, G.L., Gerlt, J.A., Ringe, D., and Kozarich, J.W. (1993). On the origin of enzymatic species. Trends Biochem. Sci. *18*, 372–376.
4. 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.
5. 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.
6. O'Brien, P.J., and Herschlag, D. (1999). Catalytic promiscuity and the evolution of new enzymatic activities. Chem. Biol. *6*, R91–R105.
7. Todd, A.E., Orengo, C.A., and Thornton, J.M. (2001). Evolution of function in protein superfamilies, from a structural perspective. J. Mol. Biol. *307*, 1113–1143.
8. Ollis, D.L., Cheah, E., Cygler, M., Dijkstra, B., Frolow, F., Franken, S.M., Harel, M., Remington, S.J., Silman, I., Schrag, J., et al. (1992). The α/β hydrolase fold. Protein Eng. *5*, 197–211.
9. Babbitt, P.C., Hasson, M.S., Wedekind, J.E., Palmer, D.R., Barrett, W.C., Reed, G.H., Rayment, I., Ringe, D., Kenyon, G.L., and Gerlt, J.A. (1996). The enolase superfamily: a general strategy for enzyme-catalyzed abstraction of the α-protons of carboxylic acids. Biochemistry *35*, 16489–16501.
10. Holm, L., and Sander, C. (1997). An evolutionary treasure: unification of a broad set of amidohydrolases related to urease. Proteins *28*, 72–82.
11. Galperin, M.Y., and Koonin, E.V. (1997). A diverse superfamily of enzymes with ATP-dependent carboxylate-amine/thiol ligase activity. Protein Sci. *6*, 2639–2643.
12. Cove, D.J. (1974). Evolutionary significance of autogenous regulation. Nature *251*, 256.
13. Aravind, L., and Koonin, E.V. (1998). Eukaryotic transcription regulators derive from ancient enzymatic domains. Curr. Biol. *8*, R111–R113.
14. 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.
15. Devos, D., and Valencia, A. (2000). Practical limits of function prediction. Proteins *41*, 98–107.
16. Grishin, N.V. (2000). Two tricks in one bundle: helix-turn-helix gains enzymatic activity. Nucleic Acids Res. *28*, 2229–2233.
17. Brew, K., Castellino, F.J., Vanaman, T.C., and Hill, R.L. (1970). The complete amino acid sequence of bovine α-lactalbumin. J. Biol. Chem. *245*, 4570–4582.
18. Findlay, J.B.C., and Brew, K. (1972). The complete amino acid sequence of human α-lactalbumin. Eur. J. Biochem. *27*, 65–86.
19. Bernstein, F.C., Koetzle, T.F., Williams, G.J., Meyer, E.F., Jr., Brice, M.D., Rodgers, J.R., Kennard, O., Shimanouchi, T., and Tasumi, M. (1977). The Protein Data Bank: a computer-based archival file for macromolecular structures. J. Mol. Biol. *112*, 535–542.
20. Wistow, G., and Piatigorsky, J. (1987). Recruitment of enzymes as lens structural proteins. Science *236*, 1554–1556.
21. Jeffery, C.J. (1999). Moonlighting proteins. Trends Biochem. Sci. *24*, 8–11.
22. Rawlings, N.D., and Barrett, A.J. (1997). Structure of membrane glutamate carboxypeptidase. Biochim. Biophys. Acta *1339*, 247–252.
23. Rasmussen, B., Wilberg, F.C., Flodgaard, H.J., and Larsen, I.K. (1997). Structure of HBP, a multifunctional protein with a serine proteinase fold. Nat. Struct. Biol. *4*, 265–268.
24. Hennig, M., Schlesier, B., Dauter, Z., Pfeffer, S., Betzel, C., Hohne, W.E., and Wilson, K.S. (1992). A TIM barrel protein without enzymatic activity? FEBS Lett. *306*, 80–84.
25. Hennig, M., Schlesier, B., Pfeffer-Hennig, S., Dauter, Z., Nong, V.H., and Wilson, K.S. (1995). Crystal structure of narbonin at 1.8 Å resolution. Acta Crystallogr. D Biol. Crystallogr. *51*, 177–189.
26. Hennig, M., Jansonius, J.N., Terwisscha van Scheltinga, A.C., Dijkstra, B.W., and Schlesier, B. (1995). Crystal structure of concanavalin B at 1.65 Å resolution, an 'inactivated' chitinase from seeds of Canavalia ensiformis. J. Mol. Biol. *254*, 237–246.
27. Klabunde, T., Eicken, C., Sacchettini, J.C., and Krebs, B. (1998). Crystal structure of a plant catachol oxidase containing a dicopper center. Nat. Struct. Biol. *5*, 1084–1090.
28. Rossman, M.G., Moras, D., and Olsen, K.W. (1974). Chemical and biological evolution of a nucleotide-binding protein. Nature *250*, 194–199.
29. Murzin, A.G. (1996). Structural classification of proteins: new superfamilies. Curr. Opin. Struct. Biol. *6*, 386–394.

30. Taneja, B., and Mande, S.C. (1999). Conserved structural features ad sequence patterns in the GroES fold family. Protein Eng. *12*, 815–818.
31. Rydén, L.G., and Hunt, L.T. (1993). Evolution of protein complexity—the blue copper-containing oxidases and related proteins. J. Mol. Evol. *36*, 41–66.
32. Grishin, N.V. (2001). MH1 domain of Smad is a degraded homing endonuclease. J. Mol. Biol. *307*, 31–37.
33. Anantharaman, V., Koonin, E.V., and Aravind, L. (2001). Peptide-N-glycanases and DNA repair proteins, Xp-C/Rad4, are, respectively, active and inactivated enzymes sharing a common transglutaminase fold. Hum. Mol. Genet. *10*, 1627–1630.
34. Wang, S., Trumble, W.R., Liao, H., Wesson, C.R., Dunker, A.K., and Kang, C. (1998). Crystal structure of calsequestrin from rabbit skeletal muscle sarcoplasmic reticulum. Nat. Struct. Biol. *5*, 476–482.
35. Murzin, A.G. (1998). How far divergent evolution goes in proteins. Curr. Opin. Struct. Biol. *8*, 380–387.
36. Kauppi, B., Lee, K., Carredano, E., Parales, R.E., Gibson, D.T., Eklund, H., and Ramaswamy, S. (1998). Structure of an aromatic-ring-hydroxylating dioxygenase-naphthalene 1,2-dioxygenase. Structure *6*, 571–586.
37. Henrissat, B. (1991). A classification of glycosyl hydrolases based on amino-acid-sequence similarity. Biochem. J. *280*, 309–316.
38. Henrissat, B., Callebaut, I., Fabrega, S., Lehn, P., Mornon, J.J., and Davies, G. (1995). Conserved catalytic machinery and the prediction of a common fold for several families of glycosyl hydrolases. Proc. Natl. Acad. Sci. USA *92*, 7090–7094.
39. Mosyak, L., Reshetnikova, L., Goldgur, Y., Delarue, M., and Safro, M.G. (1995). Structure of phenylalanyl-tRNA synthetase from Thermus thermophilus. Nat. Struct. Biol. *2*, 537–547.
40. Mancia, F., Keep, N.H., Nakagawa, A., Leadlay, P.F., McSweeney, S., Rasmussen, B., Bosecke, P., Diat, O., and Evans, P.R. (1996). How coenzyme B_{12} radicals are generated: the crystal structure of methylmalonyl-coenzyme A mutase at 2 Å resolution. Structure *4*, 339–350.
41. Betts, L., Xiang, S., Short, S.A., Wolfenden, R., and Carter, C.W.J. (1994). Cytidine deaminase. The 2.3 Å crystal structure of an enzyme: transition-state analog complex. J. Mol. Biol. *235*, 635–656.
42. Han, S., Eltis, L.D., Timmis, J.N., Muchmore, S.W., and Bolin, J.T. (1995). Crystal structure of the biphenyl-cleaving extradiol dioxygenase from a PCB-degrading Pseudomonad. Science *270*, 976–980.
43. Han, S., Craig, J.A., Putnam, C.D., Carozzi, N.B., and Tainer, J.A. (1999). Evolution and mechanism from structures of an ADP-ribosylating toxin and NAD complex. Nat. Struct. Biol. *6*, 932–936.
44. Nam, H.-J., Poy, F., Krueger, N.X., Saito, H., and Frederick, C.A. (1999). Crystal structure of the tandem phosphatase domains of RPTP LAR. Cell *97*, 449–457.
45. Kemmink, J., Darby, N.J., Dijkstra, K., Nilges, M., and Creighton, T.E. (1997). The folding catalyst protein disulphide isomerase is constructed of active and inactive thioredoxin modules. Curr. Biol. *7*, 239–245.
46. Streuli, M., Krueger, N.X., Thai, T., Tang, M., and Saito, H. (1990). Distinct functional roles of the two intracellular phosphatase-like domains of the receptor-linked protein tyrosine phosphatases LCA and LAR. EMBO J. *9*, 2399–2407.
47. Aleshin, A.E., Kirby, C., Liu, X., Bourenkov, G.P., Bartunik, H.D., Fromm, H.J., and Honzatko, R.B. (2000). Crystal structures of mutant monomeric hexokinase I reveal multiple ADP binding sites and conformational changes relevant to allosteric regulation. J. Mol. Biol. *296*, 1001–1015.
48. Hortsch, M., Patel, N.H., Bieber, A.J., Traquina, Z.R., and Goodman, C.S. (1990). Drosophila neurotactin, a surface glycoprotein with homology to serine esterases, is dynamically expressed during embryogenesis. Development *110*, 1327–1340.
49. Hawkins, A.R., Lamb, H.K., Moore, J.D., Charles, I.G., and Roberts, C.F. (1993). The pre-chorismate (shikimate) and quinate pathways in filamentous fungi—theoretical and practical aspects. J. Gen. Microbiol. *139*, 2891–2899.
50. Iyer, L.M., Koonin, E.V., and Aravind, L. (2001). Adaptations of the helix-grip fold for ligand binding and catalysis in the START domain superfamily. Proteins *43*, 134–144.
51. Baker, P.J., Waugh, M.L., Wang, X.G., Stillman, T.J., Turnbull, A.P., Engel, P.C., and Rice, D.W. (1997). Determinants of substrate specificity in the superfamily of amino acid dehydrogenases. Biochemistry *36*, 16109–16115.
52. Oue, S., Okamoto, A., Yano, T., and Kagamiyama, H. (1999). Redesigning the substrate specificity of an enzyme by cumulative effects of the mutations of non-active site residues. J. Biol. Chem. *274*, 2344–2349.
53. Orengo, C.A., Michie, A.D., Jones, S., Jones, D.T., Swindells, M.B., and Thornton, J.M. (1997). CATH—a hierarchic classification of protein domain structures. Structure *5*, 1093–1108.
54. Pearl, F.M.G., Martin, N., Bray, J.E., Buchan, D.W.A., Harrison, A.P., Lee, D., Reeves, G.A., Shepherd, A.J., Sillitoe, I., Todd, A.E., et al. (2001). A rapid classification protocol for the CATH domain database to support structural genomics. Nucleic Acids Res. *29*, 223–227.
55. Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J.H., Zhang, Z., Miller, W., and Lipman, D.J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. *25*, 3389–3402.
56. Bairoch, A., and Apweiler, R. (2000). The SWISS-PROT protein sequence database and its supplement trEMBL in 2000. Nucleic Acids Res. *28*, 45–48.
57. Kraulis, P.J. (1991). MOLSCRIPT—a program to produce both detailed and schematic plots of protein structures. J. Appl. Crystallogr. *24*, 946–950.
58. Wistow, G.J., Mulders, J.W.M., and Dejong, W.W. (1987). The enzyme lactate-dehydrogenase as a structural protein in avian and crocodilian lenses. Nature *326*, 622–624.
59. Ostrovsky de Spicer, P., and Maloy, S. (1993). PutA protein, a membrane-associated flavin dehydrogenase, acts as a redox-dependent transcriptional regulator. Proc. Natl. Acad. Sci. USA *90*, 4295–4298.
60. Furukawa, T., Yoshimura, A., Sumizawa, T., Harraguchi, M., Akiyama, S., Fukui, K., Ishizawa, M., and Yamada, Y. (1992). Angiogenic factor. Nature *356*, 668.
61. Kennedy, M.C., Mendemueller, L., Blondin, G.A., and Beinert, H. (1992). Purification and characterization of cytosolic aconitase from beef-liver and its relationship to the iron-responsive element binding-protein. Proc. Natl. Acad. Sci. USA *89*, 11730–11734.
62. Wetterau, J.R., Combs, K.A., Spinner, S.N., and Joiner, B.J. (1990). Protein disulphide isomerase is a component of the microsomal triglyceride transfer protein complex. J. Biol. Chem. *265*, 9800–9807.
63. Subramanian, A., and Miller, D.M. (2000). Structural analysis of α-enolase. J. Biol. Chem. *275*, 5958–5965.
64. Choi, H.-K., Tong, L., Minor, W., Dumas, P., Boege, U., Rossmann, M.G., and Wengler, G. (1991). Structure of Sindbis virus core protein reveals a chymotrypsin-like serine proteinase and the organization of the virion. Nature *354*, 37–43.
65. Bergdoll, M., Eltis, L.D., Cameron, A.D., Dumas, P., and Bolin, J.T. (1998). All in the family: structural and evolutionary relationships among three modular proteins with diverse functions and variable assembly. Protein Sci. *7*, 1661–1670.
66. Vallée, F., Turner, M.A., Lindley, P.L., and Howell, P.L. (1999). Crystal structure of an inactive duck δ II crystallin mutant with bound argininosuccinate. Biochemistry *38*, 2425–2434.
67. Louie, G.V., Brownlie, P.D., Lambert, R., Cooper, J.B., Blundell, T.L., Wood, S.P., Warren, M.J., Woodcock, S.C., and Jordan, P.M. (1992). Structure of porphobilinogen deaminase reveals a flexible multidomain polymerase with a single catalytic site. Nature *359*, 33–39.
68. Gillette, W.K., Rhee, S., Rosner, J.L., and Martin, R.G. (2000). Structural homology between MarC of the AraC family of transcriptional activators and the integrase family of site-specific recombinases. Mol. Microbiol. *35*, 1582–1583.
69. Stewart, M., Kent, H.M., and McCoy, A.J. (1998). Structural basis for molecular recognition between nuclear transport factor 2 (NTF2) and the GDP-bound form of the Ras-family GTPase Ran. J. Mol. Biol. *277*, 635–646.

70. Taylor, W.R., and Orengo, C.A. (1989). Protein structure alignment. J. Mol. Biol. *208*, 208–229.
71. Murzin, A.G., Brenner, S.E., Hubbard, T., and Chothia, C. (1995). SCOP—a structural classification of proteins database for the investigation of sequences and structures. J. Mol. Biol. *247*, 536–540.
72. Bullock, T.L., Clarkson, W.D., Kent, H.M., and Stewart, M. (1996). The 1.6 Å resolution crystal structure of nuclear transport factor 2 (NTF2). J. Mol. Biol. *260*, 422–431.
73. Zlateva, T., di Muro, P., Salvato, B., and Beltramini, M. (1996). The *o*-diphenol oxidase activity of arthropod hemocyanin. FEBS Lett. *384*, 251–254.
74. Groll, M., Bajorek, M., Kohler, A., Moroder, L., Rubin, D.M., Huber, R., Glickman, M.H., and Finley, D. (2000). A gated channel into the proteasome core. Nat. Struct. Biol. *7*, 1062–1067.
75. Whitby, F.G., Masters, E.I., Kramer, L., Knowlton, J.R., Yao, Y., Wang, C.C., and Hill, C.P. (2000). Structural basis for the activation of 20S proteasomes by 11S regulators. Nature *408*, 115–120.
76. Perbandt, M., Wilson, J.C., Eschenburg, S., Mancheva, I., Aleksiev, B., Genov, N., Willingmann, P., Weber, W., Singh, T.P., and Betzel, Ch. (1997). Crystal structure of vipoxin at 2.0 Å: an example of regulation of a toxic function generated by molecular evolution. FEBS Lett. *412*, 573–577.
77. Betzel, C., Genov, N., Rajashankar, K.R., and Singh, T.P. (1999). Modulation of phospholipase A_2 activity generated by molecular evolution. Cell. Mol. Life Sci. *56*, 384–397.
78. Fisher, A.J., Raushel, F.M., Baldwin, T.O., and Rayment, I. (1995). Three-dimensional structure of bacterial luciferase from *Vibrio harveyi* at 2.4 Å resolution. Biochemistry *34*, 6581–6586.
79. Fox, B.G., Liu, Y., Dege, J.E., and Lipscomb, J.D. (1991). Complex formation between the protein components of methane monooxygenase from *Methylosinus trichosporium* OB3b. Identification of sites of component interaction. J. Biol. Chem. *266*, 540–550.
80. Liu, Y., Nesheim, J.C., Paulsen, K.E., Stankovich, M.T., and Lipscomb, J.D. (1997). Roles of the methane monooxygenase reductase component in the regulation of catalysis. Biochemistry *36*, 5223–5233.
81. Bordo, D., Deriu, D., Colnaghi, R., Carpen, A., Pagani, S., and Bolognesi, M. (2000). The crystal structure of a sulfurtransferase from *Azotobacter vinelandii* highlights the evolutionary relationship between the rhodanese and phosphatase enzyme families. J. Mol. Biol. *298*, 691–704.
82. Barth, H., Hofmann, F., Olenik, C., Just, I., and Aktories, K. (1998). The N-terminal part of the enzyme component (C2I) of the binary *Clostridium botulinum* C2 toxin interacts with the binding component C2II and functions as a carrier system for a Rho ADP-ribosylating C3-like fusion toxin. Infect. Immun. *66*, 1364–1369.
83. Jia, Z., Barford, D., Flint, A.J., and Tonks, N.K. (1995). Structural basis for phosphotyrosine peptide recognition by protein tyrosine phosphatase 1B. Science *268*, 1754–1758.