All in the family: Structural and evolutionary relationships among three modular proteins with diverse functions and variable assembly

MARC BERGDOLL,¹ LINDSAY D. ELTIS,² ALEXANDER D. CAMERON,³
PHILIPPE DUMAS,⁴ AND JEFFREY T. BOLIN⁵

¹Laboratoire de Biologie Structurale, Institut de Génétique et de Biologie Moléculaire et Cellulaire du CNRS,
1 rue Laurent Fries, B.P. 163, 67404 Illkirch Cedex, France

²Department of Biochemistry, Pavillon Marchand, Université Laval, Québec City, PQ Canada G1K 7P4

³Department of Molecular Biology, Uppsala University, Biomedical Center, Box 590, S-751 24, Uppsala, Sweden

⁴UPR 9002, Institut de Biologie Moléculaire et Cellulaire du CNRS, 15 rue René Descartes, 67084 Strasbourg Cedex, France

⁵Department of Biological Sciences/LILY, Purdue University, West Lafayette, Indiana 47907-1392

(RECEIVED March 19, 1998; ACCEPTED May 18, 1998)

Abstract

The crystal structures of three proteins of diverse function and low sequence similarity were analyzed to evaluate structural and evolutionary relationships. The proteins include a bacterial bleomycin resistance protein, a bacterial extradiol dioxygenase, and human glyoxalase I. Structural comparisons, as well as phylogenetic analyses, strongly indicate that the modern family of proteins represented by these structures arose through a rich evolutionary history that includes multiple gene duplication and fusion events. These events appear to be historically shared in some cases, but parallel and historically independent in others. A significant early event is proposed to be the establishment of metal-binding in an oligomeric ancestor prior to the first gene fusion. Variations in the spatial arrangements of homologous modules are observed that are consistent with the structural principles of three-dimensional domain swapping, but in the unusual context of the formation of larger monomers from smaller dimers or tetramers. The comparisons support a general mechanism for metalloprotein evolution that exploits the symmetry of a homooligomeric protein to originate a metal binding site and relies upon the relaxation of symmetry, as enabled by gene duplication, to establish and refine specific functions.

Keywords: bleomycin resistance protein; extradiol dioxygenase; glyoxalase I; metalloproteins; modular proteins; protein evolution; three-dimensional domain swapping

Gene duplication is recognized as a primary mechanism for the origination of new functions during the course of evolution. Indeed, numerous analyses of the sequences and tertiary structures of proteins have revealed the duplication, and, in some cases the subsequent fusion, of genes that code for protein modules. Recognized families of proteins derived from duplicated genes tend to have shared or closely related functions because identification by sequence analysis is limited to cases where the primary structures

have not diverged beyond a certain point. However, comparisons of tertiary structure offer the possibility of detecting relationships between proteins of more divergent primary structures and functions (Bajaj & Blundell, 1984).

Here we report comparisons between independently determined crystal structures of three modular proteins that evolved through gene duplication and fusion, but possess low sequence similarity and diverse functions: bleomycin resistance protein (BRP) from *Streptoalloteichus hindustanus* (Dumas et al., 1994); 2,3-dihydroxybiphenyl 1,2-dioxygenase (DHBD) from *Burkholderia cepacia* LB400 (Han et al., 1995); and human glyoxalase I (GLO) (Cameron et al., 1997). Phylogenetic analyses based on the structural comparisons were also performed at several levels of structural hierarchy to investigate evolutionary relationships among these proteins and their modern homologues.

BRP (Fig. 1A) binds and sequesters bleomycin and related compounds without degrading or transforming them. In organisms that produce these antibiotics, BRP homologues are thought to be in-

Reprint requests to: Jeffrey T. Bolin, Department of Biological Sciences/LILY, Purdue University, West Lafayette, Indiana 47907-1392; e-mail: jtb@cc.purdue.edu.

Abbreviations: BRP, bleomycin resistance protein from *Streptoalloteichus hindustanus* and other proteins with the same apparent function; DHBD, 2,3-dihydroxybiphenyl 1,2-dioxygenase from *Burkholderia cepacia* strain LB400 and other proteins with the same apparent function; GLO, human glyoxalase I and other proteins with the same apparent function; DIOX, extradiol dioxygenase; PDB, Protein Databank (Bernstein et al., 1977); GenBank, GenBank DNA sequence databank (Benson et al., 1997).

Fig. 1. The similar backbone structures of (A) the BRP homodimer, (B) the DHBD monomer, and (C) the GLO homodimer. Each molecule is illustrated by a backbone ribbon drawing prepared with MOLSCRIPT (Kraulis, 1991). In each case a true or pseudo twofold axis of symmetry is approximately perpendicular to the page. Metal atoms are drawn as green spheres. For the BRP and GLO dimers, the first and second $\beta\alpha\beta\beta\beta$ modules in each monomer are colored red and blue, respectively; the monomers are distinguished by saturation. Modules 1, 2, 3, and 4 of DHBD are colored blue, red, cyan, and magenta, respectively. For all molecules, parts of the backbone not in a $\beta\alpha\beta\beta\beta$ module are represented by narrow, light gray ribbons or coils.

involved in self-protection and/or exportation, whereas in other organisms they are clearly resistance proteins. Amino acid sequences are available for at least seven BRPs. All are from fungal or bacterial sources and have monomer molecular weights of ~ 14 kDa. The BRP from *S. hindustanus* is a homodimer and binds two bleomycin molecules; the monomer comprises 124 amino acid residues.

DHBD (EC 1.13.11.39, Fig. 1B) is an extradiol dioxygenase. Enzymes in this family exploit non-heme Fe(II) or Mn(II) (rarely) to catalyze the opening of aromatic rings, a critical step in the microbial degradation of aromatic compounds. The DHBD from *B. cepacia* LB400 is Fe(II)-dependent and acts in a pathway that degrades biphenyl and polychlorinated biphenyls (PCBs). Amino acid sequences are available for over 40 extradiol dioxygenases homologous to DHBD (DIOXs); all are from bacterial sources. Within this group, enzymes with large (~ 33 kDa) and small (~ 21 kDa) monomers are known. *B. cepacia* LB400 DHBD has a large monomer comprising 298 amino acids, and it binds one Fe(II) in a single active site. This enzyme is octameric, but the oligomeric state varies among DIOXs and is not a significant issue in comparisons of DHBD with GLO and BRP. *B. cepacia* LB400 is often identified as *Pseudomonas* sp. strain LB400 in prior publications.

GLO (EC 4.4.1.5) is also known as lactoylglutathione lyase and utilizes Zn to catalyze the formation of an acylthioester during the first of two steps in the glutathione-dependent inactivation of toxic methylglyoxal. At least 11 sequences from eubacterial, fungal, plant, and animal sources are known. As observed for the DIOXs, the GLOs can also be divided into two groups according to monomer size. The large and small GLOs have 32–37 kDa and 15–21 kDa monomers, respectively. Human GLO (Fig. 1C) has a small monomer of 183 amino acid residues, is homodimeric, and binds two Zn ions in its two active sites.

Results and discussion

Structural equivalence with variation in assembly

Although BRP, DHBD, and GLO possess less than 20% sequence identity as well as notable differences in monomer size, oligomeric state, and function, the overall three-dimensional structures of the BRP and GLO dimers and the DHBD monomer are remarkably similar (Fig. 1). All three proteins belong to the α/β class (Levitt & Chothia, 1976), and each structure includes four superimposable copies of a $\beta\alpha\beta\beta\beta$ module (Figs. 2 and 3) spanning approximately 50–60 amino acid residues. Within a module, the sequential order of the secondary structure elements is $\beta_1\alpha\beta_2\beta_3\beta_4$, whereas the spatial order and polarity of the strands are $\beta_1^+\beta_4^+\beta_3^-\beta_2^+$. The four modules of each protein have comparable spatial positions and thus the BRP or GLO dimer is grossly equivalent to the DHBD monomer. Moreover, the functional units of BRP and GLO are the dimers, and the functional unit of DHBD is the monomer.

Despite these similarities, there is a fundamental difference between the arrangement of modules observed in the BRP and GLO dimers and that found in the DHBD monomer (Cameron et al.,

Fig. 2. The structure of the $\beta\alpha\beta\beta\beta$ module. The first $\beta\alpha\beta\beta\beta$ module from BRP is shown as a backbone ribbon drawing prepared with MOLSCRIPT (Kraulis, 1991). The secondary structure of this module is representative of the common secondary structure of all modules.

Fig. 3. Superposition of the eight independent $\beta\alpha\beta\beta$ modules in BRP, DHBD, and GLO. In this stereo diagram, the odd- and even-numbered modules are in darker and lighter shades, respectively. The superposition of the β_1 , β_4 , and β_3 strands (from left to right) as well as the helices is clearly demonstrated. The structural distinction between the odd and even modules at the C-terminal end of the helix, as noted in the text, is also apparent. The figure was prepared with MOLSCRIPT (Kraulis, 1991).

1997). The different modes of assembly may be understood in terms of the contacts between modules. As illustrated in Figure 4, each $\beta\alpha\beta\beta$ module makes two major contacts with neighboring modules. In an edge-to-edge contact, the β sheets from two mod-

Fig. 4. Schematic drawing of the arrangement of $\beta\alpha\beta\beta$ modules in the DHBD monomer and GLO dimer. Each module is represented by four shaded rectangles and one large, shaded circle; the rectangles represent β strands and the circles represent helices. Connecting segments, indicated by arrows and lines, may include regular secondary structure external to the $\beta\alpha\beta\beta$ modules as well as loops. Large numbers near the corners of the molecules indicate the sequence of the modules along the polypeptide chains; for GLO, the superscripts A and B distinguish the monomers. Numbers on each rectangle indicate the order of β strands along the chains. The active sites are marked by small circles representing the Fe^{2+} ion in DHBD and the two Zn^{2+} ions in GLO. The modules of the BRP dimer are arranged as those in GLO.

ules interact through their β_1 edge strands to form an extended β structure. In a back to back contact, two modules are arranged such that their α helices as well as β sheets interact.

In BRP and GLO, the two $\beta\alpha\beta\beta$ modules within each subunit are involved in a back-to-back contact: module 1^A with 2^A, and 1^B with 2^B, where 1 and 2 indicate the first and second modules and A and B designate the subunits. Edge-to-edge contacts are formed between modules from different subunits, 1^A with 2^B and 1^B with 2^A. Both BRP and GLO have two active sites related by the symmetry of the dimer. In both cases, the active sites are situated at the subunit-subunit interface and lie wholly within boundaries established by edge-to-edge association. In DHBD, the sequentially adjacent modules 1 and 2 have an edge-to-edge contact, as do modules 3 and 4. Back-to-back contacts are formed between modules 1 and 4, and between modules 2 and 3, across an interface that separates the N- and C-terminal halves of the molecule. The single active site is surrounded by modules 3 and 4. It is in the same relative location as one of the active sites of the BRP and GLO dimers and is similarly established by an edge-to-edge contact.

The difference between the arrangement of modules in BRP and GLO versus that in DHBD is readily embraced by the concepts of three-dimensional domain swapping (Bennett et al., 1995), which entail a special relationship between the interactions observed among domains of a monomeric protein and those observed in an oligomeric form of the same protein or a homologue. In a domain-swapped oligomer, a domain of each monomer is replaced in its three-dimensional environment by an identical domain from another monomer: interactions between domains that are found in the monomer are also observed in the oligomer, but involve domains from different monomers. Thus it has been proposed that domain swapping provides a mechanism whereby dimers or higher oligomers could arise from monomers

by exploiting existing, stable interactions (Bax et al., 1990; Bennett et al., 1994).

In this case, swapped interactions are observed at the level of the $\beta\alpha\beta\beta$ modules inasmuch as the positions of sequentially adjacent modules are exchanged in BRP and GLO relative to DHBD. However, the apparent swapping is observed in an atypical context. Here it is found in the different arrangement of modules in homodimers of smaller subunits when compared with a larger monomer. Moreover, as will be discussed further below, the monomeric descendent arose from a dimeric ancestor.

Structure and evolution at the level of the $\beta\alpha\beta\beta$ module

The modern structures and evolution of these proteins may be analyzed at the level of the single modules or the covalently linked two-module units by the use of structural superpositions and sequence alignments derived from them. The superpositions include 28 comparisons among the eight modules of known structure, whereas sequence comparisons and phylogenetic analyses involve larger sets of homologous proteins, or portions of these proteins, as described in the section on Methods.

At the level of the $\beta\alpha\beta\beta$ module, the results of pair-wise least-squares superpositions of $C\alpha$ coordinates, documented in Table 1, establish that the tertiary structures of the eight crystallographically characterized modules are extremely similar: on average, 39 $C\alpha$ atoms are superimposed with a root-mean-square (RMS) difference of 1.35 Å. Sequence alignments based on the superpositions demonstrate the weak conservation of primary structure among modules of known structure (Fig. 5).

It is noteworthy that the tertiary structural relationships are stronger among odd-numbered modules (module 1 of BRP or GLO and modules 1 and 3 of DHBD) or among even-numbered modules (module 2 of BRP or GLO and modules 2 and 4 of DHBD) than between odd- and even-numbered modules. For example, the range of RMS differences between the superimposed $C\alpha$ models is 0.88–1.20 Å (average = 1.07 Å) for all comparisons among odd-numbered modules, and the range is 1.12–1.35 Å (average = 1.27 Å) for all comparisons among even-numbered modules. In contrast, the range is 1.46–1.78 Å (average = 1.57 Å) for comparisons between odd and even modules. Structural features that distinguish odd and even modules include the connections between

the α helix and the preceding and following β strands as well as the connection between β_3 and β_4 . The former are very consistent within the set of odd modules, whereas the latter is consistent within the even modules. In all comparisons, the most variable elements of secondary structure are associated with the outer β strand, β_2 , and its connections to the preceding helix and the following strand.

The most conserved residues in the odd and even modules occur in distinct signature sequences, S_{odd} and S_{even} , that overlap the distinguishing tertiary structural features. In the odd modules, the motif $S_{\text{odd}} = \text{DX}_6\text{FXTX}_2\text{LG}(\text{F,M,L})\text{X}_6\text{D}$ includes residues within the α helix or in the connection that follows it, whereas the signature of the even modules, $S_{\text{even}} = (\text{D,T})\text{PXGX}_2(\text{L,V,I})(\text{E,H})$, corresponds to residues in the loop between β_3 and β_4 , or in the N-terminal end of β_4 itself; the final (E,H) residue is a metal ligand in all metal-binding modules. Although the sequence conservation among larger sets of modules derived from all available sequences is not sufficiently strong to establish meaningful phylogenetic relationships, the LG(F,M,L) helix termination in S_{odd} as well as the turn sequence PXG in S_{even} persist and remain distinctive.

On the basis of the structural relationships between the modules, and their presence in multiple copies within these proteins, we postulate that divergent evolution from a common, ancestral gene for the $\beta\alpha\beta\beta$ module is very much more probable than convergent evolution. As this module could be a stable folding unit, one might expect to find additional examples within libraries of known folds, such as the FSSP database (Holm & Sander, 1994). We were unable to identify a single case where a $\beta\alpha\beta\beta$ module is present as an independent domain, or any further cases where two (or more) modules are found in tandem. At least five known folds do contain $\beta\alpha\beta\beta$ substructures, but these occur within larger domains that bear little or no resemblance to the two- or four-module assemblies considered here. Moreover, the signature sequences are not observed in these cases. The identified examples are the proof-reading domain of *Escherichia coli* DNA polymerase I (PDB:1kfd) (Beese et al., 1993), the palm domains of *Bacillus subtilis* polymerase I and of rat polymerase β (PDB:1bpb) (Sawaya et al., 1994), the single domain of human ADP-ribosylation factor 1 (PDB:1hur) (Amor et al., 1994), the B chain of wheat serine carboxypeptidase II (PDB:1wht) (Bullock et al., 1994), and the single domain of *E. coli* peptidyl-tRNA hydrolase (PDB:2pth) (Schmitt et al., 1997). It seems likely that the $\beta\alpha\beta\beta$ substructures of these

Table 1. Results of least-squares superposition of $\beta\alpha\beta\beta$ modules^a

	BRP1	GLO1	DHBD1	DHBD3	DHBD2	DHBD4	GLO2	BRP2
BRP1		1.15 Å	1.06 Å	1.19 Å	1.54 Å	1.53 Å	1.60 Å	1.65 Å
GLO1	41		0.88 Å	1.20 Å	1.55 Å	1.50 Å	1.50 Å	1.46 Å
DHBD1	41	40		0.94 Å	1.66 Å	1.55 Å	1.36 Å	1.53 Å
DHBD3	41	46	48		1.42 Å	1.77 Å	1.57 Å	1.57 Å
DHBD2	39	41	38	34		1.28 Å	1.26 Å	1.35 Å
DHBD4	37	38	33	33	46		1.34 Å	1.25 Å
GLO2	43	39	36	38	42	42		1.12 Å
BRP2	36	34	32	30	45	43	38	

^aEntries above the diagonal to the right list the RMS difference between $C\alpha$ coordinates following superposition of the two modules using the program O (Jones et al., 1991) as described in the text. Entries at the lower left contain the number of $C\alpha$ pairs included in the comparison. Data pertaining to comparisons of odd vs. odd modules and even vs. even modules are emboldened.

Fig. 5. Structure-based alignment of the sequences of $\beta\alpha\beta\beta$ modules. The limits of the modules, the superposition methods, and the alignment criteria are described in the text. The sequence labels specify the protein and the position of the module within the protein; thus BRP1 is module one of BRP. Sequences of odd- and even-numbered modules are separated into two blocks, with the odd modules at the top. The secondary structure indicated in the top line is that of the BRP1 module, but is representative of all modules. The solid bars in the line between the blocks mark positions used to initiate the pairwise superpositions. Asterisks, "*", in the first line below the sequences mark the positions where metals ligands may be found; residues that act as ligands are in bold type. Shaded symbols highlight residues that are strongly conserved within odd- or even-modules. Boxed symbols highlight residues that are strongly conserved in both odd- or even-modules. The lines labeled "Odd," "Even," and "All" indicate consensus sequences for corresponding sets, with a capital letter indicating complete conservation within the set and a lower case letter indicating strong conservation ($\geq 75\%$ conserved for the odd- or even-blocks, $\geq 50\%$ conserved for all sequences). The figure was prepared with ALSCRIPT (Barton, 1993).

proteins are simply analogous features contained within the folds of more complex α/β proteins.

Structure and evolution at the levels of the two-module unit and four-module assemblies

Based on the structural similarities of the modules, it is reasonable to conclude that the gene for a two-module structure arose from the ancestral $\beta\alpha\beta\beta$ gene by duplication and fusion. These events might have occurred independently for any subset of the known proteins. However, the stronger similarities observed within the sets of even and odd modules, together with the consistent distinctions between even and odd modules, suggest that all two-module units in the three structurally characterized proteins and their known homologues share a common two-module ancestor.

Phylogenetic analyses were thus undertaken at the level of the two-module units to evaluate evolutionary relationships beyond the origination of this ancestor. The N- and C-terminal halves of four-module chains were treated as independent units. A phylogenetic graph and bootstrap analysis are illustrated in Figure 6, which shows division of the family into subfamilies distinguished by function. The apparent relationships among and between the DIOXs and GLOs are of particular interest because both enzymes bind metal ions, and both have modern representatives with small (two-module) and large (four-module) monomers.

Within the DIOX subfamily, as previously noted (Eltis & Bolin, 1996), the two-module monomers of the small-chain class are cladistically segregated from the two-module N- and C-terminal domains of the large-chain class. The N- and C-domains also segregate into distinct clades, and the small-chain monomers have the same relationship to the clade of N-terminal domains as they have to the clade of C-terminal domains. These observations imply that all two-module units in small- and large-chain DIOXs arose from a unique two-module ancestor, and that the small- and large-chain lineages diverged at the two-module stage. For the large-chain class, similar phylogenetic relationships exist among almost all

species within the clades of N- and C-terminal domains. This result implies evolution from a unique four-module DIOX, and that the domains co-evolved for almost all species. The few exceptions to the general pattern do not indicate a second origination of a four-module enzyme, but rather have relationships consistent with genetic exchange at the domain level (Benjamin et al., 1991; Harayama & Rekik, 1993; Eltis & Bolin, 1996).

In contrast, the domains of the large GLOs do not merge with or conform to this pattern: the N- and C-terminal domains do not group with the corresponding domains of the DIOXs, are not segregated from each other, and are not segregated from the small GLO monomers. Instead, all GLO two-module units belong to a single clade that is separated from all DIOX units, and is divisible into two groups, both containing N- and C-terminal domains of large monomers as well as small monomers. These results suggest that all known GLO two-module units evolved from a common two-module ancestor after divergence of the GLO and DIOX lineages. Thus it appears that the large GLOs and large DIOXs do not share a common large ancestor, but arose by independent duplication and fusion events. Moreover, the group relationships within the GLO clade suggest that large GLOs arose from smaller ancestors at two or more independent points.

A proposed scheme for modular evolution of the genes

Figure 7 summarizes a scheme for the development of the genes associated with the known members of this family of proteins, which include the two- and four-module DIOXs and GLOs, BRP, and a Mn-dependent glutathione transferase, FosA, that imparts resistance to the antibiotic fosfomycin (Bernat et al., 1997). The latter is a dimer of 16 kDa subunits, possesses the S_{odd} and S_{even} signature sequences, and has potential metal-binding side chains at appropriate positions, but the metal ligands are not known and the three-dimensional structure has not been determined.

We postulate the existence of an ancestral gene for a $\beta\alpha\beta\beta$ module, which probably coded for a corresponding one-module

Fig. 6. Cladogram of majority-rule, consensus tree. This tree was obtained via PAUP (Swofford, 1993) bootstrap analysis using covalent two-module domains as the operational taxonomic unit. Bootstrap values refer to the closest node. The sequences BRP SH and BRP KP were used as outgroups. Other details about the calculation, including a description of the input alignment, are provided in the text. Long dashed lines mark boundaries between proteins of different function, whereas shorter dashed lines indicate the segregation of the DIOXs into three groups as described in the text. A leading GLO or BRP in a sequence abbreviation indicates a member of the glyoxalase I (lactoylglutathione lyase) group or bleomycin resistance protein group, respectively. A trailing N or C indicates the N- or C-terminal domain of a four-module protein. Otherwise, the abbreviations represent the species and functions as follows, with references and/or GenBank accession numbers provided in parentheses: FOSA, *Serratia marcescens* transposon Tn2921 fosfomycin resistance protein (Navas et al., 1990) (M85195); FOSB, *Staphylococcus epidermidis* fosfomycin resistance protein (Zilhao & Courvalin, 1990) (X54227); GLO SH, human (Ranganathan et al., 1993) (L07837); GLO SC, *Saccharomyces cerevisiae* (unpublished) (Z49810); GLO SP, *Schizosaccharomyces pombe* (unpublished) (Z54140); GLO EC, *E. coli* (unpublished, U57363); GLO SPO, *Sporolobus stapfianus* (unpublished) (Y10782); GLO BO, *Brassica oleracea* (unpublished) (Z74962); PS400, *B. cepacia* LB400 DHBD (Hofer et al., 1993) (X66122); PSF1, *Pseudomonas putida* F1 3-methylcatechol 2,3-dioxygenase (Zylstra & Gibson, 1989) (J04996); RGP6C1, *Rhodococcus globulus* P6 DHBD-I (Asturias et al., 1994) (X75633); PSQ1, *Sphingomonas paucimobilis* Q1 DHBD (Taira et al., 1988) (M20640); PSG7, *P. putida* PpG71,2-dihydroxynaphthalene dioxygenase (Harayama & Rekik, 1989) (J04994); MT-2, *P. putida* MT-2 catechol 2,3-dioxygenase (Nakai et al., 1983) (V01161); PS35X, *P. putida* P35X catechol 2,3-dioxygenase (unpublished) (X77856); PSHV3, *Pseudomonas* sp. HV3 catechol 2,3-dioxygenase (Yrjälä et al., 1994) (L10655); RRCTM, *Rhodococcus rhodochrous* CTM catechol 2,3-dioxygenase (Candidus et al., 1994) (X69504); BSTE, *Bacillus steaerothermophilus* FFTP-3 catechol 2,3-dioxygenase (unpublished) (X67860); BphC2_RGP6, *R. globulus* P6 DHBD-II (Asturias et al., 1994) (X75634); BphC_BN6, naphthalenesulfonate-degrading bacterium BN6 DHBD (Heiss et al., 1995) (U22355); BRP SH, *S. hindustanus* (Drocourt et al., 1990) (X52869); BRP KP, *E. coli* (Mazodier et al., 1985) (X01702).

Fig. 7. An evolutionary scheme leading to genes that code for the known modern proteins homologous to BRP, DHBD, and GLO. Each filled rectangle symbolizes the gene for a $\beta\alpha\beta\beta$ module. The distinctions between the odd- and even-numbered modules in the modern proteins are symbolized by the use of blue patterns for all odd modules and red patterns for all even modules. Patterns with lines in the same orientation are used for all modules in proteins with the same or closely similar functions. The shading of the large GLO gene indicates that genes for different large GLOs were apparently created by independent duplication and fusion events. A more extensive explanation of the scheme is given in the text.

protein. Gene duplication and fusion coupled with point mutations gave rise to the gene for a two-module protein with distinctive first and second modules. This gene is the ancestor of all modern genes in the family. Genes for all modern functions are proposed to have arisen from this gene by mutation coupled with elements of gene duplication and speciation, but not gene fusion; that is, all modern functions originated in two-module proteins. In the DIOX lineage, a subsequent gene duplication and fusion event gave rise to the ancestor of the four-module monomers, which are the dominant modern group (see below). In the GLO lineage, independent duplication and fusion events gave rise to four-module enzymes at two or more distinct points. There is no evidence for additional gene fusion in the evolution of BRP or FosA homologues, but undiscovered large-chain versions of these proteins might exist.

Structure and evolution of the metal binding and active sites

Despite diversity among the functions and substrates/ligands of these proteins, the active sites are comparable in location and structure. This comparison is particularly striking and informative for the case of DHBD and GLO. After superposition of the six central β strands of the C-terminal domain of DHBD with corresponding segments of the GLO dimer, the active site Fe atom bound by DHBD lies 2.5 Å from one of the two Zn^{2+} ions of the GLO dimer. Furthermore, the three residues that serve as Fe ligands in DHBD superimpose with three of the four Zn ligands in GLO (Table 2): residues His146, His210, and Glu260 from DHBD correspond to residues Gln33^A, His126^B, and Glu172^B from GLO, where the superscripts distinguish the monomers of GLO. The fourth Zn ligand in GLO is residue Glu99^A, which corresponds to a conserved alanine residue in the DIOXs, Ala198 in DHBD. A similar metal binding site has been predicted for FosA from sequence data (Bernat et al., 1997).

It is significant that the ligands occur in pairs at equivalent positions along strands β_1 and β_4 of two $\beta\alpha\beta\beta$ modules in edge-to-edge contact, and that these positions are related by pseudo

twofold symmetry. Using GLO as an example, if modules 1^A and 2^B are superimposed, the Zn ligands Gln33^A and His126^B superimpose along the β_1 strands, whereas ligands Glu99^A and Glu172^B superimpose along the β_4 strands. The locations of the ligands suggest that the formation of a symmetric, oligomeric protein with the ability to bind a metal ion via four ligands was a crucial step in the evolution of this family of proteins. Thus the ultimate evolutionary link among these proteins seems to be ancestral metal binding.

The hypothesis of ancestral metal binding is challenged by the structure of BRP, which lacks a metal binding site; in fact, the sequence of only one BRP includes as many as two typical ligands for transition metals at the four positions indicated by comparisons with GLO and DHBD. Rather than binding a metal, it is likely that BRP binds bleomycin in the same active-site cavity (Dumas et al., 1994). It is intriguing that bleomycin itself chelates Fe(II). Moreover, the observation that genes involved in the biosynthesis of antibiotics are often linked to genes that impart resistance (Martin & Liras, 1989) offers the possibility that BRP arose from a metalloenzyme involved in the synthesis of bleomycin or a similar compound. At this time there is no evidence for the involvement of such a metalloenzyme. However, the pathway for bleomycin biosynthesis is incompletely characterized at the genetic or enzymatic level, and the gene that codes for BRP in the producing organism *Streptomyces verticillus* is located within a cluster that includes at least one gene that may play a role in bleomycin biosynthesis as well as several genes likely to be involved in transport (Calcutt & Schmidt, 1994).

As will be discussed further below, the first metal-binding ancestor could have been assembled with symmetry from $\beta\alpha\beta\beta$ modules in several different ways. Regardless of the original mode of protein assembly, the first gene duplication would have facilitated the introduction of asymmetry and specialization of the active site for different functions. Analysis of the structure of DHBD suggests several results of such specialization. For example, the loss by mutation of one metal-binding side chain must have been a crucial step in the origination of a functional extradiol dioxygenase inasmuch as current mechanistic proposals (Shu et al., 1995; Spence et al., 1996) require three accessible Fe ligand sites to bind

both hydroxyl groups of its catecholic substrate and dioxygen. In essence, the replacement of the fourth ligand by alanine establishes the necessary third, readily available, coordination site. Thus, the crystal structure of DHBD in the absence of substrates shows that the Fe atom has only three protein ligands, two easily displaced water ligands, and one empty coordination site (Han et al., 1995; Senda et al., 1996). When catecholic substrates bind, the empty coordination site is occupied by one hydroxyl group of the substrate, and one water is displaced by the other hydroxyl group (Bolin et al., 1996; Senda et al., 1996). The other water ligand, which binds near Ala198, is probably displaced by dioxygen in the ternary complex.

The evolution of the DIOXs produced a second specialization in the establishment of a conserved histidine residue adjacent to the Fe binding site, His195 in DHBD. His195 is the second residue within a β bulge found exclusively in the DIOXs and only in their odd-numbered modules. It is located on the N-terminal end of strand β_4 , just prior to the position of the discarded ligand. Although loss of a ligand enabled the extradiol dioxygenase reaction, it would have increased the lability of the Fe(II) site. His195 appears to stabilize Fe binding through a hydrogen bond with one of the water ligands, and thus we speculate that the introduction of a hydrogen-bonding side chain at this location provided compensation for the absence of the ligand, which would be sufficient to explain its exclusive presence in the dioxygenases. Although a number of hydrogen-bonding side chains could serve this function, the ultimate selection of a histidine residue may also reflect improved catalytic function through acid-base chemistry, and/or through electrostatic stabilization of an anionic Fe-bound, activated dioxygen species by a histidinium residue.

A final example of specialization in the extradiol dioxygenases is the apparent loss of a second Fe binding site in most of the known enzymes. The crystal structures of DHBD established that the capacity for Fe binding exists only in one active site located in the C-terminal domain of the enzyme. In fact, a very large majority of the more than 40 characterized DIOXs (ca. 90%) are similar to DHBD in that they have four-motif monomers and possess less than three potential Fe ligands at the expected positions in the N-terminal domain. The dominance of the large-chain, one-Fe, enzymes might again be associated with a selective advantage related to low availability of Fe and/or its weak binding in sites that provide only three protein ligands. Certain mutations that stabilize Fe-binding in one site at the expense of binding in the other could be tolerated and advantageous in the case of the large-monomer enzymes, but would not be possible for a symmetric dimer of small monomers. Such mutations could occur at the interface between the N- and C-terminal domains of the large monomer, or occur within the N-domain and exert their influence through the interface.

The analysis by Ohlendorf and coworkers (1988, 1994) of the crystal structure of a nonhomologous Fe(III)-dependent intradiol dioxygenase, protocatechuate 3,4-dioxygenase, 3,4-PCD, from *Pseudomonas aeruginosa*, suggested an evolutionary scenario that is similar in the effective result but different in mechanism. 3,4-PCD is a dodecamer of $\alpha\beta$ dimers, and the structurally similar α and β chains appear to have arisen by gene duplication and subsequent divergence. Each dimer possesses one Fe(III)-binding, functional active site and one Fe-free, vestigial active site; access to the vestigial site is limited by the dodecameric assembly. The existence of an ancestral, symmetric homodimer with two functional active sites was inferred. In this case, it was proposed that loss of metal binding in one of the sites was promoted by mutations that stabi-

Table 2. Correspondence of metal binding sites after superposition^a

Dioxygenase residue and atom	Glyoxalase residue and atom	Positional difference (Å) between	
		Ligand/metal atoms	C α atoms
His 146 Ne2	Gln A33 Oe1	2.3	0.7
Ala 198 —	Glu A99 Oe1	—	1.4
His 210 Ne2	His B126 Ne2	2.1	1.4
Glu 260 Oe1	Glu B172 Oe1	1.7	0.9
Fe	Zn	2.5	—

^aThe superposition of DHBD and GLO was performed with the lsq_explcit option of the program O (Jones et al., 1991) using 30 pairs of equivalent C α atoms from the six β strands closest to the metal binding site. The segments used were (DHBD, GLO): 146–150, A33–A37; 185–189, A69–A73; 197–201, A98–A102; 210–214, B126–B130; 248–251, B160–B164; and 259–263, B171–B175.

lized the dodecamer, which enhanced the stability of the enzymatic function, but reduced access to one active site to the point where its metal-dependent catalytic activity was not a significant advantage.

In the case of the extradiol dioxygenases, it is unlikely that aggregation of four-module monomers is closely associated with the loss of Fe binding. Although the DHBD from *B. cepacia* is octameric, the accessibilities of the functional and vestigial sites appear to be equivalent in the octamer, and both the N- and C-terminal domains are involved in interactions that stabilize the oligomer. Moreover, there exist monomeric, large-monomer extradiol dioxygenases that, on the basis of sequence, possess only one Fe-binding site in the C-terminal domain; an example is 2,2',3-trihydroxybiphenyl 1,2-dioxygenase from *Sphingomonas* strain RW1 (Happe et al., 1993).

Further considerations of symmetry, ancestors, and missing cousins

The evolutionary scheme in Figure 7 indicates the existence of ancestral one- and two-module genes but places no restrictions on the oligomeric state or mode of assembly of the corresponding proteins. However, if a critical early event in the evolution of this family was indeed the generation of a metal binding site, then the existence of a one-module ancestral protein that assembled into a symmetric edge-to-edge homodimer, or a higher oligomer that included a comparable structure, would have been advantageous. In this case, creation of a metal binding site with four ligands would have required fixation of particular amino acid residues at only two positions in the sequence; in the final step, only a single point mutation might have been needed to add two ligands. To exploit symmetry in this fashion, the minimal assembly is an edge-to-edge dimer. However, given that the functional units of all known modern proteins possess four modules, it is plausible that the one-module ancestor assembled as a symmetric homotetramer. Possible routes from metal-free dimers or tetramers of a 1-module ancestor to plausible forms of the 2-module ancestor of all members of the family are illustrated in Figure 8A. Although the figure and discussion above imply the prior assembly of stable, metal-free oligomers, this is not a requirement: generation of the metal binding site(s) could contribute to the stabilization of the oligomer.

In the subsequent evolution of the glyoxalases and dioxygenases, gene duplication and fusion events occurred at two stages with maintenance of the ancestral metal binding function. At both stages, it is presumed that generation of a larger monomer was a successful alternative because relaxation of the constraints of symmetry provided some advantage. In the case of the extradiol dioxygenases, for example, we suggest that the asymmetric loss of a ligand after the first fusion was necessary to establish the catalytic function. Thus the existence of symmetry was an essential factor in the early evolutionary events, whereas reduction of symmetry has been a significant factor thereafter.

Such an evolutionary scheme should be a common one for metalloenzymes: the origination of a metal-binding site with multiple ligands can occur at symmetric sites with mutational economy, whereas the subsequent development of specific catalytic functions in a biological context is more consistent with asymmetric metal environments. Gene duplication can play a significant role thereafter through two mechanisms that foster reduction of symmetry (Roderick & Matthews, 1993). In one mechanism, gene duplication followed by divergence of distinct genes for independent subunits leads to metal-binding sites at, or near, a pseudo symmetric interface of a heterooligomer. Here, the mechanism has involved gene duplica-

Fig. 8. Schematic representations of possible assembly patterns at various evolutionary stages. The color scheme corresponds to that used in Figure 7. **A:** Pathways leading to three possible forms of the two-module ancestor of the BRP, DIOX, and GLO proteins. A key feature in this scheme is the creation of a metal-binding site in a noncovalent oligomer that includes edge-to-edge-associated $\beta\alpha\beta\beta$ modules. Gene duplication and fusion, followed by mutation, results in a two-module protein with distinctive N- and C-terminal modules. **B:** Structural relationships among possible forms of the extradiol dioxygenases are shown. All forms might exist among modern proteins, but only the form at the bottom of the panel is confirmed by a three-dimensional structure.

tion and fusion to generate larger asymmetric monomers from smaller, symmetric oligomers. Structural evidence for this mechanism is also apparent in other families of metalloproteins such as the bacterial ferredoxins (Adman et al., 1973; Fukuyama et al., 1988) and oxygen carriers with binuclear metal sites such as hemocyanin and hemerythrin (Volbeda & Hol, 1989). In the case of the extradiol dioxygenases and the glyoxalases, this mechanism may have had a significant inherent advantage: in the evolution of these metalloproteins it has occurred at two stages, and at the second stage it has occurred independently in multiple instances.

In consideration of the two-module ancestral protein, it is interesting that the known two- and four-module structures express alternative modes of assembly characteristic of three-dimensional domain swapping. Based on proposed mechanisms for the evolution of oligomeric proteins via three-dimensional domain swapping (Bennett et al., 1995), it is reasonable to predict the existence of an edge-to-edge-fused, two-module monomeric protein equivalent to modules 1 and 2, or 3 and 4, in DHBD. The two-module ancestor may have been such a protein, and modern examples

might also exist. Moreover, the family may also include a modern or ancestral homodimer of such a protein, which would use the back-to-back interface for dimerization instead of the edge-to-edge dimerization found in BRP and GLO.

The modern, large DHBDs could have arisen from such monomeric or dimeric proteins, or from an ancestor assembled like BRP and GLO, as illustrated in Figure 8B. Thus the edge-to-edge association found between covalently linked modules of DHBD could reflect either persistence of an ancestral mode of association or reversion to it. In the latter case, the evolution of this protein family might have included the origination of a dimer of two-module monomers by three-dimensional domain swapping, and the origination of a four-module monomer through reversion of the swapped interactions.

Materials and methods

Superpositions of structures were performed using the *lsq_explicit* and *lsq_improve* options of O (Jones et al., 1991). All reported quantitative results were achieved with a distance maximum of 3.2 Å and a minimum segment length of three residues. For superposition of motifs, a consistent set of 26 paired C α atoms was used in the *lsq_explicit* stage and *lsq_improve* was used to extend the superposition; the residues used in the first stage are indicated in Figure 5. The boundaries of the motifs were consistently defined as extending from three sequence positions prior to the position of the ligand (or the structurally equivalent residue) in β_1 to four sequence positions after the ligand (or the structurally equivalent residue) in β_4 . For the superposition of the metal sites in DHBD and GLO, only the *lsq_explicit* option was used; the paired segments are listed in Table 2.

For groups of proteins with the same or similar functions, multiple sequence alignments were obtained by the use of the program CLUSTALW (Thompson et al., 1994) with minor alterations in the placement of gaps by extrapolation from the known crystal structures. Thus separate, initial alignments were prepared at the level of the $\beta\alpha\beta\beta\beta$ module and the two-modules units for the bleomycin resistance proteins, the extradiol dioxygenases, the proteins functionally related to human glyoxalase I, and the fosfomycin resistance proteins.

Three-dimensional superpositions and CLUSTALW sequence alignments were used to obtain an alignment of the sequences of the eight structurally defined $\beta\alpha\beta\beta\beta$ modules, which is reported in Figure 5. Additional sequences of modules from each functional group were added to the alignment of $\beta\alpha\beta\beta\beta$ modules by mapping the CLUSTALW alignment of the group onto its representative in the structure-based alignment. For example, sequences of the first modules of extradiol dioxygenases were added according to their alignment with the first module of DHBD. This procedure produced a master alignment of $\beta\alpha\beta\beta\beta$ modules from proteins functionally related to BRP, DHBD, and GLO. A selection of sequences was made within each functional group at this point so as to maintain the diversity of sequences within the group but reduce the total number of sequences; the result was an alignment of 68 modules. For example, the dioxygenase sequences used in the present study are all divergent members of the three major classes of type I extradiol dioxygenases previously identified (Eltis & Bolin, 1996). Type II extradiol dioxygenases, such as protocatechuate 4,5-dioxygenase, were not included in the current analyses because these enzymes do not possess the identified signature sequences, and it seems probable that they do not possess the $\beta\alpha\beta\beta\beta$ motif.

The sequences of covalently adjacent modules were merged to produce an alignment of 34 two-module units, but with the more variable sequence segments found between modules uniformly deleted. For the small-monomer proteins, the resulting, conjoined sequences correspond to the full sequence of the monomer. For the large-monomer proteins, separate two-module sequences were created for the N- and C-terminal domains. This alignment was further modified by deletion of sequence positions corresponding to: (1) a large insertion found only in some odd-numbered modules of the GLO family; and (2) a shorter insertion found only in some even-numbered modules of the BRP family. The former corresponds to 19 residues from Asp76 through Arg94 in human glyoxalase I, and the latter corresponds to 10 residues from Ser83 through Pro92 of *S. hindustanus* BRP (Fig. 4). Finally, the sequences of FosA and FosB (see Figure 5 for references) were added on the basis of alignments with the DHBD and GLO families to produce an aligned set of 36 two-module units. This alignment is available from the corresponding author on request.

Phylogenetic analyses were performed on the one- and two-module alignments using both the neighbor-joining algorithm of PHYLIP (version 3.5c) (Felsenstein, 1993) and the maximum parsimony approach of PAUP (version 3.1.1) (Swofford, 1993). Only the results from the two-module alignment are reported (see text). For the PHYLIP analyses, distance matrices were generated with the PROTDIST program from the PHYLIP package (PAM001 matrix). Neighbor-joining was performed using a random order of sequences on input. For bootstrap analyses, the SEQBOOT program of the PHYLIP package was used to generate 100 data sets, which were analyzed as described above. PAUP trees were obtained by a heuristic search using the tree bisection-reconnection branch-swapping routine. To assure that the minimum length tree was identified, 10 replicate searches were performed with different initial trees created by random variation of the order of the sequences. In each case, no new minimum length trees were identified in the five final searches. In the PAUP analyses, at least one gap in each run of gaps in the input alignment was treated as a new character state while the other gaps were treated as missing data to minimize the number of evolutionary events represented by a given run of gaps.

Acknowledgments

This work was supported in part by a grant from the National Institutes of Health, R01 GM52381, to JTB. LDE acknowledges support from the Natural Sciences and Engineering Research Council of Canada. MB and PD acknowledge support from CNRS/INSERM/ULP, and ADC was supported by the Göran Gustafsson Stiftelse.

References

- Adman ET, Sieker LC, Jensen LH. 1973. Structure of a bacterial ferredoxin. *J Biol Chem* 248:3987–3996.
- Amor JC, Harrison DH, Kahn RA, Ringe D. 1994. Structure of the human ADP-ribosylation factor 1 complexed with GDP. *Nature* 372:704–708.
- Asturias JA, Eltis LD, Prucha M, Timmis KN. 1994. Analysis of three 2,3-dihydroxybiphenyl 1,2-dioxygenases found in *Rhodococcus globularius* P6. Identification of a new family of extradiol dioxygenases. *J Biol Chem* 269:7807–7815.
- Bajaj M, Blundell T. 1984. Evolution and the tertiary structure of proteins. *Annu Rev Biophys Bioeng* 13:453–492.
- Barton GJ. 1993. ALSCRIPT: A tool to format multiple sequence alignments. *Protein Eng* 6:37–40.
- Bax B, Lapatto R, Nalini V, Driessen H, Lindley PF, Mahadevan D, Blundell TL, Slingsby C. 1990. X-ray analysis of $\beta\beta$ 2-crystallin and evolution of oligomeric lens proteins. *Nature* 347:776–780.
- Beese LS, Friedman JM, Steitz TA. 1993. Crystal structures of the Klenow fragment of DNA polymerase I complexed with deoxynucleoside triphosphate and pyrophosphate. *Biochemistry* 32:14095–14101.

- Benjamin RC, Voss JA, Kunz DA. 1991. Nucleotide sequence of *xylE* from the TOL pDK1 plasmid and structural comparison with isofunctional catechol-2,3-dioxygenase genes from TOL, pWW0 and NAH7. *J Bacteriol* 173:2724–2728.
- Bennett MJ, Choe S, Eisenberg D. 1994. Domain swapping: Entangling alliances between proteins. *Proc Natl Acad Sci USA* 91:3127–3131.
- Bennett MJ, Schlunegger MP, Eisenberg D. 1995. 3D domain swapping: A mechanism for oligomer assembly. *Protein Sci* 4:2455–2468.
- Benson DA, Boguski MS, Lipman DJ, Ostell J. 1997. GenBank. *Nucleic Acids Res* 25:1–6.
- Bernat BA, Laughlin LT, Armstrong RN. 1997. Fosfomycin resistance protein (Fosa) is a manganese metalloglutathione transferase related to glyoxalase I and the extradiol dioxygenases. *Biochemistry* 36:3050–3055.
- Bernstein FC, Koetzle TF, Williams GJ, Meyer EE Jr, Brice MD, Rodgers JR, Kennard O, Shimanouchi T, Tasumi M. 1977. The Protein Data Bank: A computer-based archival file for macromolecular structures. *J Mol Biol* 112:535–542.
- Bolin JT, Han S, Eltis LD. 1996. High resolution crystal structures of 2,3-dihydroxybiphenyl 1,2-dioxygenase from a PCB-degrading bacterium complexed with substrates. *Acta Crystallogr Sect A* 52 (suppl.):C129.
- Bullock TL, Branchaud B, Remington SJ. 1994. Structure of the complex of *L*-benzylsuccinate with wheat serine carboxypeptidase II at 2.0-Å resolution. *Biochemistry* 33:11127–11234.
- Calcutt MJ, Schmidt FJ. 1994. Gene organization in the bleomycin-resistance region of the producer organism *Streptomyces verticillus*. *Gene* 151:17–21.
- Cameron AD, Olin B, Ridderstrom M, Mannervik B, Jones TA. 1997. Crystal structure of human glyoxalase I—Evidence for gene duplication and 3D domain swapping. *EMBO J* 16:3386–3395.
- Candidus S, van Pée KH, Lings F. 1994. The catechol 2,3-dioxygenase gene of *Rhodococcus rhodochrous* CTM: Nucleotide sequence, comparison with isofunctional dioxygenases and evidence for an active-site histidine. *Microbiology* 140:321–330.
- Drocourt D, Calmels T, Reynes JP, Baron M, Tiraby G. 1990. Cassettes of the *Streptalloteichus hindustanus ble* gene for transformation of lower and higher eukaryotes to phleomycin resistance. *Nucleic Acids Res* 18:4009.
- Dumas P, Bergdoll M, Cagnon C, Masson JM. 1994. Crystal structure and site-directed mutagenesis of a bleomycin resistance protein and their significance for drug sequestering. *EMBO J* 13:2483–2492.
- Eltis LD, Bolin JT. 1996. Evolutionary relationships among extradiol dioxygenases. *J Bacteriol* 178:5930–5937.
- Felsenstein J. 1993. *PHYLIP (Phylogeny Inference Package), version 3.5c*. Seattle, Washington: University of Washington.
- Fukuyama K, Nagahara Y, Tsukihara T, Katsube Y, Hase T, Matsubara H. 1988. Tertiary structure of *Bacillus thermoproteolyticus* [4Fe-4S] ferredoxin. Evolutionary implications for bacterial ferredoxins. *J Mol Biol* 199:183–193.
- Han S, Eltis LD, Timmis KN, Muchmore SW, Bolin JT. 1995. Crystal structure of the biphenyl-cleaving extradiol dioxygenase from a PCB-degrading pseudomonad. *Science* 270:976–980.
- Happe B, Eltis LD, Poth H, Hedderich R, Timmis KN. 1993. Characterization of 2,2',3-trihydroxybiphenyl dioxygenase, an extradiol dioxygenase from the dibenzofuran- and dibenzo-*p*-dioxin-degrading bacterium *Sphingomonas* sp. strain RW1. *J Bacteriol* 175:7313–7320.
- Harayama S, Rekik M. 1989. Bacterial aromatic ring-cleavage enzymes are classified into two different gene families. *J Biol Chem* 264:15328–15333.
- Harayama S, Rekik M. 1993. Comparison of the nucleotide sequences of the meta-cleavage pathway genes of TOL plasmid pWW0 from *Pseudomonas putida* with other meta-cleavage genes suggests that both single and multiple nucleotide substitutions contribute to enzyme evolution. *Mol Gen Genet* 239:81–89.
- Heiss G, Stolz A, Kuhm AE, Muller C, Klein J, Altenbuchner J, Knackmuss HJ. 1995. Characterization of a 2,3-dihydroxybiphenyl dioxygenase from the naphthalenesulfonate-degrading bacterium strain BN6. *J Bacteriol* 177:5865–5871.
- Hofer B, Eltis LD, Dowling DN, Timmis KN. 1993. Genetic analysis of a *Pseudomonas* locus encoding a pathway for biphenyl/polychlorinated biphenyl degradation. *Gene* 130:47–55.
- Holm L, Sander C. 1994. The FSSP database of structurally aligned protein fold families. *Nucleic Acids Res* 22:3600–3609.
- Jones TA, Zou J-Y, Cowan SW, Kjeldgaard M. 1991. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr Sect A* 47:110–119.
- Kraulis PJ. 1991. MOLSCRIPT: A program to produce both detailed and schematic plots of protein structures. *J Appl Crystallogr* 24:946–950.
- Levitt M, Chothia C. 1976. Structural patterns in globular proteins. *Nature* 261:552–558.
- Martin JF, Liras P. 1989. Organization and expression of genes involved in the biosynthesis of antibiotics and other secondary metabolites. *Annu Rev Microbiol* 43:173–206.
- Mazodier P, Cossart P, Giraud E, Gasser F. 1985. Completion of the nucleotide sequence of the central region of Tn5 confirms the presence of three resistance genes. *Nucleic Acids Res* 13:195–205.
- Nakai C, Kagamiyama H, Nozaki M, Nakazawa T, Inouye S, Ebina Y, Nakazawa A. 1983. Complete nucleotide sequence of the metapyrocatechase gene on the TOL plasmid of *Pseudomonas putida* mt-2. *J Biol Chem* 258:2923–2928.
- Navas J, Leon J, Arroyo M, Garcia Lobo JM. 1990. Nucleotide sequence and intracellular location of the product of the fosfomycin resistance gene from transposon Tn2921. *Antimicrob Agents Chemother* 34:2016–2018.
- Ohlendorf DH, Lipscomb JD, Weber PC. 1988. Structure and assembly of protocatechuate 3,4-dioxygenase. *Nature* 336:403–405.
- Ohlendorf DH, Orville AM, Lipscomb JD. 1994. Structure of protocatechuate 3,4-dioxygenase from *Pseudomonas aeruginosa* at 2.15 Å resolution. *J Mol Biol* 244:586–608.
- Ranganathan S, Walsh ES, Godwin AK, Tew KD. 1993. Cloning and characterization of human colon glyoxalase-I. *J Biol Chem* 268:5661–5667.
- Roderick SL, Matthews BW. 1993. Structure of the cobalt-dependent methionine aminopeptidase from *Escherichia coli*: A new type of proteolytic enzyme. *Biochemistry* 32:3907–3912.
- Sawaya MR, Pelletier H, Kumar A, Wilson SH, Kraut J. 1994. Crystal structure of rat DNA polymerase β : Evidence for a common polymerase mechanism. *Science* 264:1930–1935.
- Schmitt E, Mechulam Y, Fromant M, Plateau P, Blanquet S. 1997. Crystal structure at 1.2 Å resolution and active site mapping of *Escherichia coli* peptidyl-tRNA hydrolase. *EMBO J* 16:4760–4769.
- Senda T, Sugiyama K, Narita H, Yamamoto T, Kimbara K, Fukuda M, Sato M, Yano K, Mitsui Y. 1996. Three-dimensional structures of free form and two substrate complexes of an extradiol ring-cleavage type dioxygenase, the BphC enzyme from *Pseudomonas* sp strain KKS102. *J Mol Biol* 255:735–752.
- Shu L, Chiou YM, Orville AM, Miller MA, Lipscomb JD, Que L Jr. 1995. X-ray absorption spectroscopic studies of the Fe(II) active site of catechol 2,3-dioxygenase. Implications for the extradiol cleavage mechanism. *Biochemistry* 34:6649–6659.
- Spence EL, Langley GJ, Bugg TDH. 1996. *Cis-trans* isomerization of a cyclopropyl radical trap catalyzed by extradiol catechol dioxygenases: Evidence for a semiquinone intermediate. *J Am Chem Soc* 118:8336–8343.
- Swofford DL. 1993. *PAUP: Phylogenetic analysis using parsimony, version 3.1.1*. Champaign, Illinois: Illinois Natural History Survey.
- Taira K, Hayase N, Arimura N, Yamashita S, Miyazaki T, Furukawa K. 1988. Cloning and nucleotide sequence of the 2,3-dihydroxybiphenyl dioxygenase gene from the PCB-degrading strain of *Pseudomonas paucimobilis* Q1. *Biochemistry* 27:3990–3996.
- Thompson JD, Higgins DG, Gibson TJ. 1994. CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 22:4673–4680.
- Volbeda A, Hol WG. 1989. Pseudo 2-fold symmetry in the copper-binding domain of arthropod haemocyanins. Possible implications for the evolution of oxygen transport proteins. *J Mol Biol* 206:531–546.
- Yrjälä K, Paulin L, Kilpi S, Romantschuk M. 1994. Cloning of *cmpE*, a plasmid-borne catechol 2,3-dioxygenase-encoding gene from the aromatic- and chloroaromatic-degrading *Pseudomonas* sp. HV3. *Gene* 138:119–121.
- Zilhao R, Courvalin P. 1990. Nucleotide sequence of the *fosB* gene conferring fosfomycin resistance in *Staphylococcus epidermidis*. *FEMS Microbiol Lett* 56:267–272.
- Zylstra GJ, Gibson DT. 1989. Toluene degradation by *Pseudomonas putida* F1. Nucleotide sequence of the *todC1C2BADE* genes and their expression in *Escherichia coli*. *J Biol Chem* 264:14940–14946.