doi:10.1006/jmbi.2000.3805 available online at http://www.idealibrary.com on IDEAL® J. Mol. Biol. (2000) 299, 1019–1034

JMB

The 1.7 Å Crystal Structure of BPI: A Study of How Two Dissimilar Amino Acid Sequences can Adopt the Same Fold

Gary Kleiger[1], Lesa J. Beamer[2], Robert Grothe[1], Parag Mallick[1] and David Eisenberg[1]*

[1]*UCLA-DOE Laboratory of Structural Biology and Molecular Medicine, Molecular Biology Institute, UCLA BOX 951570, Los Angeles CA 90095-1570, USA*

[2]*Biochemistry Department University of Missouri-Columbia, Columbia MO 65211, USA*

**Corresponding author*

We have extended the resolution of the crystal structure of human bactericidal/permeability-increasing protein (BPI) to 1.7 Å. BPI has two domains with the same fold, but with little sequence similarity. To understand the similarity in structure of the two domains, we compare the corresponding residue positions in the two domains by the method of 3D-1D profiles. A 3D-1D profile is a string formed by assigning each position in the 3D structure to one of 18 environment classes. The environment classes are defined by the local secondary structure, the area of the residue which is buried from solvent, and the fraction of the area buried by polar atoms. A structural alignment between the two BPI domains was used to compare the 3D-1D environments of structurally equivalent positions. Greater than 31 % of the aligned positions have conserved 3D-1D environments, but only 13 % have conserved residue identities. Analysis of the 3D-1D environmentally conserved positions helps to identify pairs of residues likely to be important in conserving the fold, regardless of the residue similarity. We find examples of 3D-1D environmentally conserved positions with dissimilar residues which nevertheless play similar structural roles. To generalize our findings, we analyzed four other proteins with similar structures yet dissimilar sequences. Together, these examples show that aligned pairs of dissimilar residues often share similar structural roles, stabilizing dissimilar sequences in the same fold.

Keywords: BPI; X-ray crystallography; 3D-1D environment; domain; fold

Introduction

Proteins with sequence similarity also display structural similarity. However, many proteins with no apparent sequence similarity display the same folds. For example, the mitochondrial enzyme rhodanese contains two domains of similar structure, but little sequence similarity (Hol *et al.*, 1983). This phenomenon is not rare. In fact, the database of distant aligned protein structures (DAPS) has over 1000 examples of structurally similar proteins with less than 25 % sequence identity (Rice & Eisenberg, 1997). Examples of both intra and inter-molecular fold similarity in the absence of amino acid similarity is given in Table 1.

To study how dissimilar protein sequences adopt similar folds, we analyze the structure of the bactericidal/permeability-increasing protein (BPI). BPI has two domains with the same fold but with dissimilar sequences. Both BPI domains are twisted, anti-parallel β-sheet barrels capped by two α-helices. The domain main-chain atoms can be superimposed without significant deformation (3.0 Å rmsd over 173 residues). The BPI domain is to date a unique fold. We take advantage of the apparent domain duplication in BPI to find structurally conserved positions for the BPI domain.

BPI is a mammalian protein located in polymorphonuclear neutrophils, a cell of the innate immune response that protects the host during microbial infection (Elsbach & Weiss, 1995). BPI specifically binds lipopolysaccharides in the outer-membrane of Gram-negative bacteria. Although

Abbreviations used: BPI, bactericidal/permeability-increasing protein; rmsd, root mean square deviation; FWLO, fractional weighted log-odds.

E-mail address of the corresponding author: david@mbi.ucla.edu

0022-2836/00/041019–16 $35.00/0

Table 1. Examples of domains with similar folds but dissimilar amino acid sequences

Protein 1	Protein 2	rmsd (Å)	% Seq. ID	Reference
Hexokinase domain I	Hexokinase domain II	2.8	11	Steitz *et al.* (1976)
Rhodanese domain I	Rhodanese domain II	1.8	13	Ploegman *et al.* (1978)
Rhizopuspepsin domain I	Rhizopuspepsin domain II	3.0	13	Subramanian *et al.* (1977)
Phosphoglycerate kinase dom. I	PGK domain II	4.5	8	Banks *et al.* (1979)
Arabinose-binding protein dom. I	Arabinose domain II	3.2	7	Quiocho *et al.* (1977)
Bovine F1-ATPase	Rec A protein	3.2	9	Abrahams *et al.* (1994); Story *et al.* (1992)
Bromoperoxidase A2	Triacylglycerol hydrolase	2.7	14	Hecht *et al.* (1994); Uppenberg *et al.* (1994)
β-amylase	Concanavalin B	4.1	10	Hennig *et al.* (1995); Mikami *et al.* (1994)
Cellobiohydrolase I	Serum amyloid component	3.4	10	Divne *et al.* (1994); Emsley *et al.* (1994)
Neuroglian	T cell antigen receptor	3.3	11	Bentley *et al.* (1995); Huber *et al.* (1994)

The first five examples are similar folds where each domain is from the same protein (rms deviations are for superimposed main-chain atoms); the last five examples are similar folds where each domain is from a different protein (rmsd are for superimposed C^{α} atoms). This table has been adapted and expanded from Richardson (1981).

BPI has two domains, a single BPI domain can adopt a stable fold: an N-terminal fragment has been expressed and retains the *in vitro* activity of wild-type BPI (Horwitz *et al.*, 1996). The BPI X-ray structure was determined in our laboratory to 2.4 Å resolution at room temperature (Beamer *et al.*, 1997), and here is extended to 1.7 Å using cryo-crystallography diffraction data.

Here, we use 3D-1D environment classes to identify the structurally equivalent positions in the two domains that have similar atomic environments. 3D-1D environment classes have been used successfully for tasks such as protein sequence fold assignment, and assessing the quality of protein structures derived from NMR or X-ray crystallography experiments (Bowie *et al.*, 1991; Luthy *et al.*, 1992). The 3D-1D environment classes describe the protein structure, where each environment is defined by the local secondary structure, the solvent accessibility, and the fraction of local atoms that are polar for each residue. Our hypothesis is that structurally equivalent positions with identical environments are important for conserving the fold. While one might expect to find positions with conserved structural environments in the core of each BPI domain, we show here that structurally equivalent surface positions with dissimilar residue types can conserve structural roles. With the environment classes determined from our 1.7 Å model, we compare structurally equivalent positions between the two domains of BPI and examine the physical-chemical properties that determine the unusual BPI structure.

Results

Extension of the resolution and refinement of the 1.7 Å structure of BPI

Using cryo-crystallography as described in Materials and Methods, the resolution of the BPI structure was extended to 1.7 Å and refined to an *R*-factor of 0.198 with an R_{free} of 0.249. Comparison of the high-resolution model of BPI with the room-temperature model reveals little structural change, with the exception of residues 42 to 48. In the new structure, several side-chains on one side of the loop now pack against the protein, whereas these side-chains in the previous model were mostly exposed to solvent. The loop rearrangement may be due to conditioning the crystals with 45% PEG 6000 for cryo-protection, freezing, or a combination of the two. Equivalent main-chain atoms for the two models superimpose with an rmsd of 0.9 Å. The ribbon diagram is shown in Figure 1(a).

Comparison of the N-terminal and C-terminal domains of BPI using 3D-1D environments

The structural alignment of the N-terminal and C-terminal BPI domains was used to generate a sequence alignment between the two domains (Figure 2). Conservation between the two domains was first examined at the residue level. A total of 21 pairs of residues are identical out of 164 aligned positions between the two domains, corresponding to a sequence identity of 13%. This level of sequence identity is significantly higher than would be expected for two independently generated random sequences of this length (*Z*-score equals 3.6, see Materials and Methods). Even so, the sequence identity is too low to allow sequence alignment methods to predict an alignment, and in fact the structural identity of the two domains was unsuspected prior to the 2.4 Å resolution structure.

To understand the fold identity of the two BPI domains in the absence of strong sequence similarity, positional similarities were examined using 3D-1D environment classes. The environments for each position in the two domains are shown on the structure-derived sequence alignment of the two domains (Figure 2). We reasoned that structurally equivalent positions with similar environments might also have similar structural roles in each BPI domain. Of the 164 structurally aligned positions, 51 have identical 3D-1D environments, correspond-

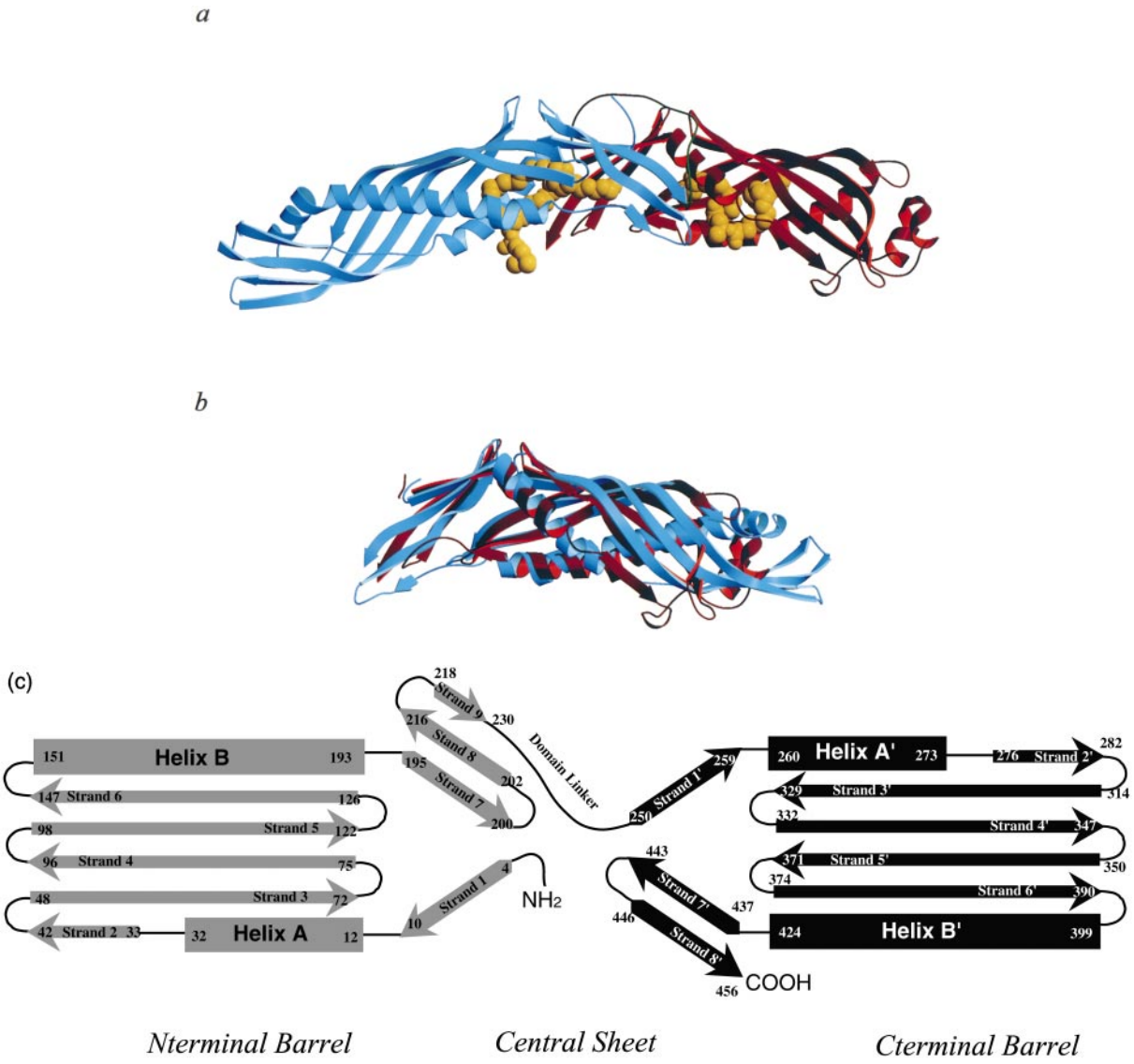

Figure 1. (a) Ribbon representation of the 1.7 Å crystal structure of human BPI. The N-terminal domain is blue. The C-terminal domain is red. Residues 10 to 193 fold into a structural element called the N-terminal barrel. The barrel is composed of five anti-parallel β-strands which twist about the barrel axis. Two α-helices complete the barrel by closing a gap in the β-sheet. Residues 260 to 430 fold into a similar structure called the C-terminal barrel. Amino acid residues 201 to 229, as well as 431 to 456, fold into the central β-sheet of six strands, located in the center of the molecule, which interacts with both the N-terminal and C-terminal barrels. A linker of residues 230 to 250 (olive) connects the N-terminal and C-terminal domains. (b) Superposition of the N-terminal domain (blue) on the C-terminal domain (red). Residues 1-229 were structurally aligned to residues 251-456 using the algorithm ALIGN_V2 (Cohen, 1986). The two domains align with 3.0 Å root mean square deviation over the main-chain atoms of the 173 structurally corresponding residues. (c) Schematic of BPI showing its elongated shape and two-domain structure. The two domains are related by a pseudodyad perpendicular to the page. Secondary structure units are represented by arrows (β-strands) and rectangles (helices). The N-terminal domain (residues 1-229) is gray; the C-terminal domain (residues 251 to 456) is black. Secondary structure units have been numbered, with the primes denoting the units in the C-terminal domain. Residue positions for the start and end of each secondary structure unit are shown. The three subdomains (N-terminal barrel, C-terminal barrel, and central sheet) are shown.

ing to 31% of the structurally equivalent positions in the BPI domain alignment. This level of conservation is highly significant relative to an alignment of two independently generated random profiles of this length (Z-score = 12.9. A fuller statistical analysis of the BPI domain alignment is given in Materials and Methods).

Analysis of environmentally conserved positions in the BPI alignment as a function of environment class

We then asked if certain environment classes are more conserved than others in the BPI alignment. We use a *p*-value which is the probability that at

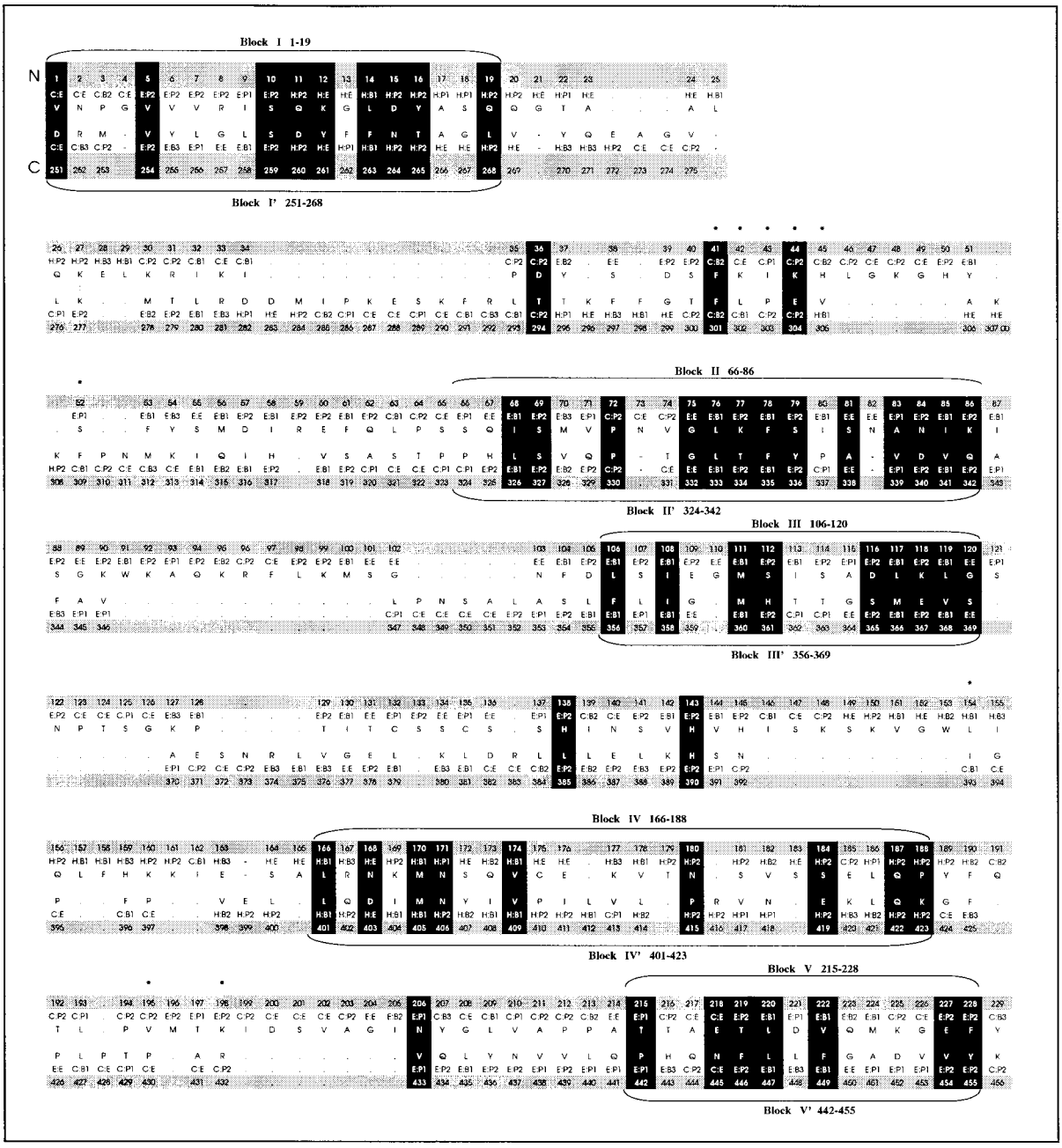

Figure 2. Structure-based alignment of the amino acid sequences of the N-terminal and C-terminal domains of BPI. Both the sequence and corresponding 3D-1D environments are shown. The N-terminal domain (residues 1 to 229) is the top sequence; the C-terminal domain (residues 251 to 456) is the bottom of each dual line. If the 3D-1D environments are identical for a given pair in the alignment, the two positions are considered structurally conserved and colored black. Consecutive stretches of 3D-1D-conserved residues are labeled blocks I-V. The nine positions that were removed from the initial alignment have stars located above the corresponding N-terminal position.

least M out of N pairs of environments from the BPI domain alignment would be the same if the environments were paired at random. Small p-values indicate correlation between the environments of structurally aligned positions; p-values for each environment class are shown in Table 2.

Notice that residues important for stabilizing the cores of proteins tend to be hydrophobic and buried in apolar environments, that is, in the B1 environment class (see Materials and Methods and Table 5 for the definition of each environment class). We would then expect to observe a low p-value for B1-B1 pairs because the structural role of residues that pack in the protein core tend to be conserved. The p-values for the H:B1 and E:B1 environment classes are 3×10^{-6} and 8×10^{-6}, respectively. Therefore, one would expect to observe at least as many H:B1 matches for the BPI

Table 2. Log-odds and fractional weighted log-odds (FWLO) values for 3D-1D environment classes (Env. Class) for the alignment of the N and C-terminal domains of BPI and for all alignments from the DAPS database

BPI				DAPS database	
Env. Class	*P*-value	Log-odds	FWLO	Log-odds	FWLO
H:E	5×10^{-2}	1.7	0.05	1.4	0.03
H:P2	2×10^{-8}	2.0	0.24	1.2	0.11
H:P1	NO	NO	NO	1.0	0.06
H:B3	NO	NO	NO	1.0	0.02
H:B2	NO	NO	NO	1.0	0.02
H:B1	3×10^{-6}	3.1	0.17	1.5	0.16
E:E	2×10^{-2}	1.4	0.06	2.1	0.01
E:P2	1×10^{-5}	1.0	0.20	1.9	0.05
E:P1	1×10^{-1}	0.9	0.04	1.3	0.04
E:B3	NO	NO	NO	1.5	0.03
E:B2	NO	NO	NO	1.3	0.03
E:B1	8×10^{-6}	1.3	0.19	1.7	0.23
C:E	1×10^{-1}	1.1	0.03	1.4	0.07
C:P2	2×10^{-1}	0.9	0.02	0.9	0.06
C:P1	NO	NO	NO	0.9	0.03
C:B3	NO	NO	NO	1.0	0.02
C:B2	NO	NO	NO	1.1	0.01
C:B1	NO	NO	NO	1.3	0.02

The log-odds gives the log of the ratio of the observed to expected probability for the pairing of two identical environmental classes. It is a measure of the likelihood that any instance of an environment class is conserved in the alignment. The FWLO is the log-odds weighted by the observed probability for the pairing of two identical environment classes in fractional form. A given pair may have a positive log-odds value, but if it occurs rarely in the structural alignment, the FWLO value will be low. Notice that the FWLO values for the BPI alignment and the DAPS alignments follow similar trends; the H:B1, E:B1, and H:P2 have high FWLO values. See Table 5 for definitions of the environment classes. NO, no observations.

domains in a random alignment only three out of one million times. Thus, positions that belong to the H:B1 or E:B1 classes are observed more often than expected at random, suggesting these positions have conserved structural roles (a more thorough discussion of Table 3 is given in Materials and Methods.)

The same argument can be applied to the surface of the protein. Positions which belong to the *P*2 environment class tend to be solvent exposed at the protein's surface. If the environment at the surface is conserved, we would expect to observe low *p*-values for the P2 class. The *p*-values for the H:P2 and E:P2 environment classes are 2×10^{-8} and 1×10^{-5}, respectively, which are as low as the values shown above for the H:B1 and E:B1 classes. While the structural conservation of residues in the protein core is well documented, the conservation of residues on the surface of proteins is not as well described. We next characterize the structural roles of residues in these positions, especially pairs from the BPI domain alignment which do not conserve residue type.

The structural roles of 3D-1D environmentally conserved positions with dissimilar residues in the BPI domain alignment

Thirty-one of the 51 3D-1D environmentally conserved pairs from the BPI domain alignment have similar or identical residues, defined by a positive substitution score from the GONNET matrix (Benner *et al.*, 1994). However, 20 structurally equivalent positions with conserved 3D-1D environments have dissimilar residues. We next wished to characterize the structural roles for residues at these 20 positions and determine if any could be important for stabilizing the BPI fold.

The structural roles for 3D-1D environmentally conserved positions with dissimilar residues were compared by analyzing the tertiary interactions they form with other residues. We analyzed favorable contacts, including hydrogen bonds, salt bridges, disulfide bonds, van der Waals interactions, or aromatic ring stacking. We then determined for each pair in the alignment whether the structural roles are similar.

We find that the structural roles for these pairs fall into three major categories: (i) conserved structural roles; (ii) auxiliary structural roles; and (iii) different structural roles.

Conserved structural roles

The structural roles for a pair of dissimilar residues are defined as conserved when a residue at position i in the N-terminal domain of BPI forms a tertiary interaction with residue j in the N-terminal domain, and residue i′ in the C-terminal domain forms a tertiary interaction with residue j′. Both residues i and i′ and j and j′ must be at structurally equivalent positions in the alignment for the structural roles to be defined as conserved. The type of tertiary interaction does not have to be the same (Figure 3(a)).

Auxiliary structural roles

Residue i forms a tertiary interaction with residue j in the N-terminal domain, and residue i′ in

Table 3. Classification of the structural roles for equivalent positions with conserved 3D-1D environment classes and dissimilar residue types. Comparison of structural roles for the N-terminal (n) and C-terminal (c) domains of BPI.

	N-terminal domain			C-terminal domain	
Residue (n)	Atom	Structural role	Residue (c)	Atom	Structural role
A. *Conserved structural roles*					
Ser79	OG	H bond to Ser69 OG through a water molecule	Tyr336	OH	H bond to Ser327 OG
Ser112	OG	H bond to bb carbonyl group of His138 through a water molecule	His361	CE1	VdW with Leu385 CB and CD1
Asp116	OD2	H bond to Ser134 OG	Ser365	OG	H bond to Lys380 NZ
Thr219	OG1	H bond to Thr216 OG1	Phe446	-	Ring stacking with His443
B. *Auxiliary structural roles*					
Tyr16	OH	H bond to Gln20 NH2	Thr265	OG1	H bond to bb carbonyl group of Phe262
Glu19	NE2	H bond to Gln20 OE1	Leu268	CG,CD1	VdW with Glu272 CD and CG
Asp36	OD1	H bond to bb amide of Ser55	Thr294	OG1	H bond to bb amide group of Met 312
Lys77	NZ	Salt bridge to Asp116 OD1	Thr334	OG1	H bond to Gln329 NE2
Ala83	CB	VdW with Leu63 CD2 and Ile 80 CG2	Val339	CG1	VdW with Met360 CB
Asn180	ND2	H bond to Ser184 OG	Pro415	CB,CD	VdW with Arg416 CZ and Leu414 CB
Ser184	OG	H bond to Asn 180 ND2	Glu419	OE2	Salt bridge to Arg416 NH2
Pro188	-	Solvent exposed;	Lys423	-	Solvent exposed
Asn206	CG	VdW with Lys225 CB and CG	Val433	CG1, CG2	VdW with Val453 CG1
Val222	CG1	VdW with Leu209 CG and CD1	Phe449	CD1,CE1,CE2	VdW with Leu440 CB and CD1 and Val368 CG1
Glu227	OE1	Salt bridge to Lys225 NZ	Val454	CB	VdW with Tyr436 CE1
C. *Different Structural Roles*					
Val1	-	Solvent exposed	Asp251	OD1	H bond with Arg252 NH1 and Tyr255 OH
Lys12	NZ	Salt bridge with Asp452 OD1	Tyr261	CG,CD1,CD2,CE2,CZ	VdW w/Pro241 CG,CD
Gly120	-	Solvent exposed;	Ser369	OG	H bond to bb carbonyl group of Ala370
His138	-	Solvent exposed	Leu385	CB,CD1	VdW with His361 CE1
Thr215	OG	H bond to ordered water	Pro442	CB,CG	VdW with Leu447 CD1

H bond, Hydrogen-bond interaction; VdW, a van der Waals interaction; backbone, bb.

the C-terminal domain forms a tertiary interaction with residue k′. Residues i and i′ must occupy structurally equivalent positions, but residues j and k′ do not have to be aligned (Figure 3(b)).

Different structural roles

Residues i and i′ belong to the same environment class but one or both of the residues have no corresponding structural role, or one or both residues are involved in stabilizing the interface between the domains in the BPI structure rather than the domain itself (Figure 3(c)). A summary of the structural roles for the 20 pairs of dissimilar residues is given in Table 3.

Four pairs of 3D-1D environmentally conserved positions with dissimilar residue types fall into the conserved structural roles category. For example, Ser79 and Tyr336 both belong to the E:P2 environment class, however the substitution of serine for tyrosine in the GONNET matrix (Benner *et al.*, 1994) is not favorable. The conserved structural roles of both Ser79 and Tyr336 are to stabilize the domain by connecting adjacent β-strands. These residues are located on β-strands 4 and 4′ respectively (see Figures 1(c) and 4(a)). Ser79 (N-terminal domain) forms a 2.7 Å hydrogen bond to a water molecule which is hydrogen bonded to Ser69, located on β-strand 3. The crystallographic temperature factor for this water molecule is 27 $Å^2$ suggesting that it is well ordered. In the C-terminal domain, Tyr336 forms a 2.8 Å hydrogen bond to Ser327, which lies on β-strand 3′. Ser69 and Ser327 are also paired in the BPI domain alignment. Therefore, Ser79 and Tyr336 interact with structurally equivalent residues in their respective domains, conserving the structural role for these dissimilar residues.

A similar example involves Asp116 and Ser365, located on β-strands 5 and 5′, respectively (Figure 4(b)). The conserved structural roles of Asp116 and Ser365 connect adjacent β-strands in their respective domains. Asp116 forms a 2.9 Å hydrogen bond to Ser134 on β-strand 6. Ser365 forms a 3.1 Å hydrogen bond to Lys380 on β-strand 6′. Ser134 and Lys380 are paired in the BPI domain alignment.

Conserved structural roles do not necessarily involve identical tertiary interactions. For example, Thr219 and Phe446 are structurally equivalent resi-

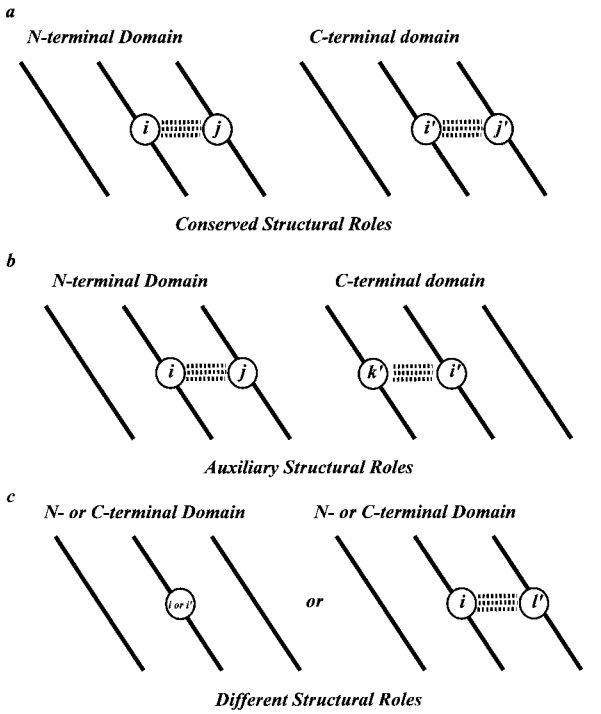

Figure 3. Pairs of 3D-1D environmentally conserved positions with dissimilar residues fall into three major categories. Each category describes the similarity of the structural role for each residue in the pair. (a) Conserved structural roles: residue i from the N-terminal domain interacts with residue j, and residue i′ in the C-terminal domain interacts with residue j′. Both residues i and i′ and j and j′ must be at structurally equivalent positions in the alignment. (b) Auxiliary structural roles: residue i from the N-terminal domain interacts with residue j, and residue i′ in the C-terminal domain interacts with residue k′. Residue i and i′ must occupy structurally equivalent positions. (c) Different structural roles: either residue i or i′ is solvent exposed, or residue i from the N-terminal domain interacts with residue l′ from the C-terminal domain.

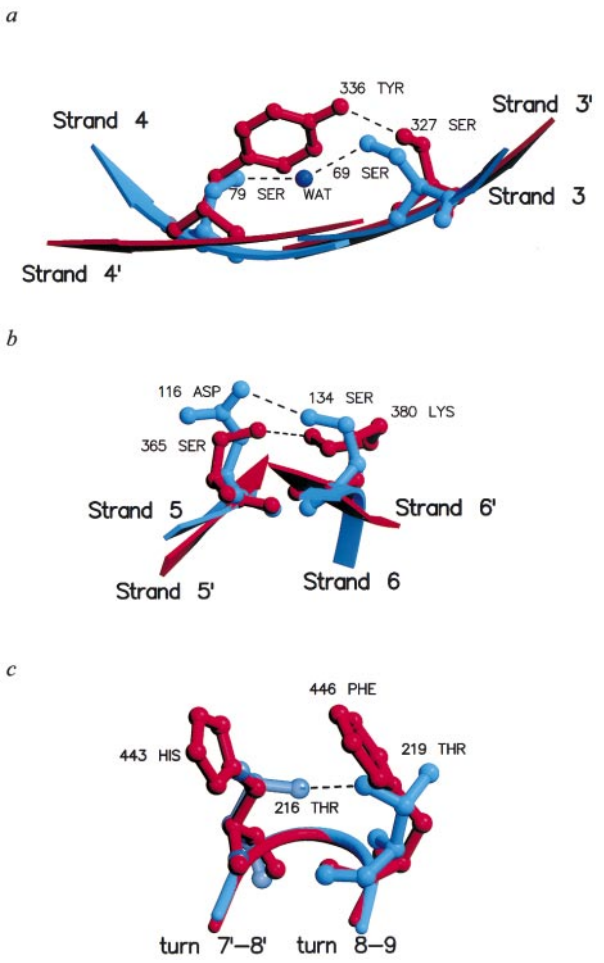

Figure 4. Examples of pairs of 3D-1D environmentally conserved residues with dissimilar residues yet conserved structural roles from the BPI domain alignment. Residues and secondary structure elements from the N-terminal domain are blue; residues and secondary structure elements from the C-terminal are red. The relationship of the secondary structure elements to the entire BPI molecule is shown in Figure 1(c). Hydrogen bonds are shown as broken lines between the hydrogen bond donor and acceptor atoms.

dues with identical environment classes (E:P2). Thr219 forms a 2.7 Å hydrogen bond with Thr216 which stabilizes a tight turn from β-strand 8 to β-strand 9 (Figure 4(c)). In the C-terminal domain, the aromatic ring of Phe446 is approximately 4 Å from the ring of His443. Ring stacking of phenylalanine and histidine residues has been shown to be an energetically favorable interaction (Mitchell *et al.*, 1994). Therefore, the interaction of Phe446 with His443 stabilizes a tight turn from β-strand 7′ to β-strand 8′, similar to the interaction found in the N-terminal domain.

Eleven 3D-1D environmentally conserved positions with dissimilar residue types have auxiliary structural roles in the BPI domains. The structural roles of these residue pairs are auxiliary because they help stabilize the fold, even though the structural roles are not strictly conserved. An example is Lys77 and Thr334, located in β-strand 4 and 4′ respectively (Figure 1(c)). Lys77 forms a salt bridge with Asp116 located on β-strand 5. Thr334 forms a hydrogen bond to Gln329 located on β-strand 3′. These interactions still support the fold of each BPI domain by connecting adjacent β-strands.

Five of the 3D-1D environmentally conserved positions with dissimilar residue types have different structural roles in the two domains. Some of these positions, such as His138, are completely exposed to bulk solvent and make no contacts with any ordered atoms in the BPI structure, whereas the structurally equivalent residue from the C-terminal domain, Leu385, makes van der Waals contacts with His361.

Another example of a pair of residues with different structural roles is Lys12 and Tyr261. Lys12 forms a salt bridge with Asp452, connecting

the N-terminal domain of BPI with the C-terminal domain. Tyr261 forms van der Waals contacts with Pro241, located in the domain linker. The structural roles for these two residues are different because Lys12 stabilizes the interface between the two domains and Tyr261 interacts with the domain linker.

Examples of conserved structural roles in proteins other than BPI with similar folds yet dissimilar sequences

The database of distant aligned protein structures (DAPS) has over 1000 pairs of structurally similar proteins with less than 25% sequence identity. To determine if other proteins of similar structure and dissimilar sequence also share residues with conserved structural roles, five examples were chosen from DAPS and compared in the same manner as the domains of BPI. These examples are shown as the last five entries in Table 1.

The first example aligns the α subunit of the bovine mitochondrial F1-ATPase with the *Escherichia coli* Rec A protein. These proteins share only 9% sequence identity over the aligned residues, yet 26% of the positions are 3D-1D environmentally conserved. Like the domains of BPI, we find pairs of 3D-1D environmentally conserved positions with dissimilar residues which share conserved structural roles. For example, Tyr300 of F1-ATPase forms a 3.5 Å hydrogen bond to Arg304 (Figure 5(a)). These residues are located on an α-helix. Tyr300 and Arg304 are separated by four residues corresponding to approximately one turn of the helix. This serves to align the two residues so that they can form a hydrogen bond stabilizing the helix. Likewise, the structurally equivalent residues in Rec A, Arg169 and Gln173, also form a 3.5 Å hydrogen bond. Tyr300 and Arg169 are paired in the structure-based sequence alignment and are not favored to substitute for one another according to the GONNET matrix (Benner *et al.*, 1994). Arg304 and Gln173 are also paired in the alignment, demonstrating that the structural roles in both proteins are conserved.

The second example aligns bromoperoxidase A2 with triacylglycerol hydrolase. The sequence identity is 14% over the aligned positions while 25% of the aligned positions are 3D-1D environmentally conserved. Gln42 in bromoperoxidase forms a 2.7 Å hydrogen bond with the carbonyl group oxygen atom of Leu260 (Figure 5(b)). The structurally equivalent residue in triacylglycerol hydrolase, Trp52, forms a 3.1 Å hydrogen bond with the carbonyl group oxygen atom of Leu228. Leu260 and Leu228 are paired in the alignment of these two proteins. The structural role of both Gln42 and Trp52 is a long-range tertiary interaction serving to connect two helices. It is noteworthy that this interaction is between two residues separated by approximately 200 residues.

A third example compares T cell antigen receptor with neuroglian. The sequence identity is 11% over the aligned positions. Of the aligned positions, 16% have 3D-1D environmentally conserved positions. Trp162 of the antigen receptor is structurally aligned to Gln650 of neuroglian. Trp162 forms a 3.4 Å hydrogen bond to Gln213, which also forms a 2.7 Å hydrogen bond to Arg211. Gln650 forms a 2.8 Å hydrogen bond to Arg687, which is paired with Arg211 in the alignment of neuroglian and T cell antigen receptor. This example demonstrates how hydrogen bonds through intermediate residues can promote conserved structural roles in proteins of similar structure.

In our fourth example, we align β-amylase with concanavalin B. These proteins share 10% sequence identity. Of the aligned positions, 22% are 3D-1D environmentally conserved. Leu78 of β-amylase is structurally aligned to Gln68 of concanavalin B. Unlike the above examples, these residues are buried in the core of each protein and belong to the H:B2 environment. Leu78 forms van der Waals contacts with other hydrophobic residues in the core, while Gln68 forms both van der Waals contacts with other residues as well as two hydrogen bonds to water molecules near the surface of the protein. The hydrogen bonding potential of the amide group hydrogen and oxygen atoms are thus satisfied, even though Gln68 is buried in the core. While this example does not satisfy the definition for conserved structural roles because each residue forms many non-specific contacts, it shows how dissimilar residues can stabilize the core of proteins.

Clustering of 3D-1D environmentally conserved positions in both the BPI fold and the lipid-binding pockets

Positions with conserved 3D-1D environments tend to cluster in the BPI fold. The location of the 51 3D-1D environmentally conserved pairs are highlighted in Figure 6(a). These positions tend to cluster around the core of each domain and the phospholipid binding pockets, while the two tips of the molecule contain very few 3D-1D environmentally conserved residues. Clustered positions predominantly belong to classes with mostly buried and apolar environments (H:B1 or E:B1) or mostly solvent accessible environments (H:P2, or E:P2).

The degree of clustering was assessed by calculating C^{α}-C^{α} distances for all 3D-1D environmentally conserved positions in each domain. For this analysis, a position is considered a tertiary neighbor of another position if its C^{α} atom is less than 7 Å away but is not within two residues on the peptide chain. We find 3D-1D environmentally conserved positions have at least one other conserved tertiary neighbor 41 and 44% of the time for the N-terminal and C-terminal domains, respectively. In contrast, positions that were not 3D-1D environmentally conserved has ''tertiary neighbors'' that were conserved only 24 and 28%,

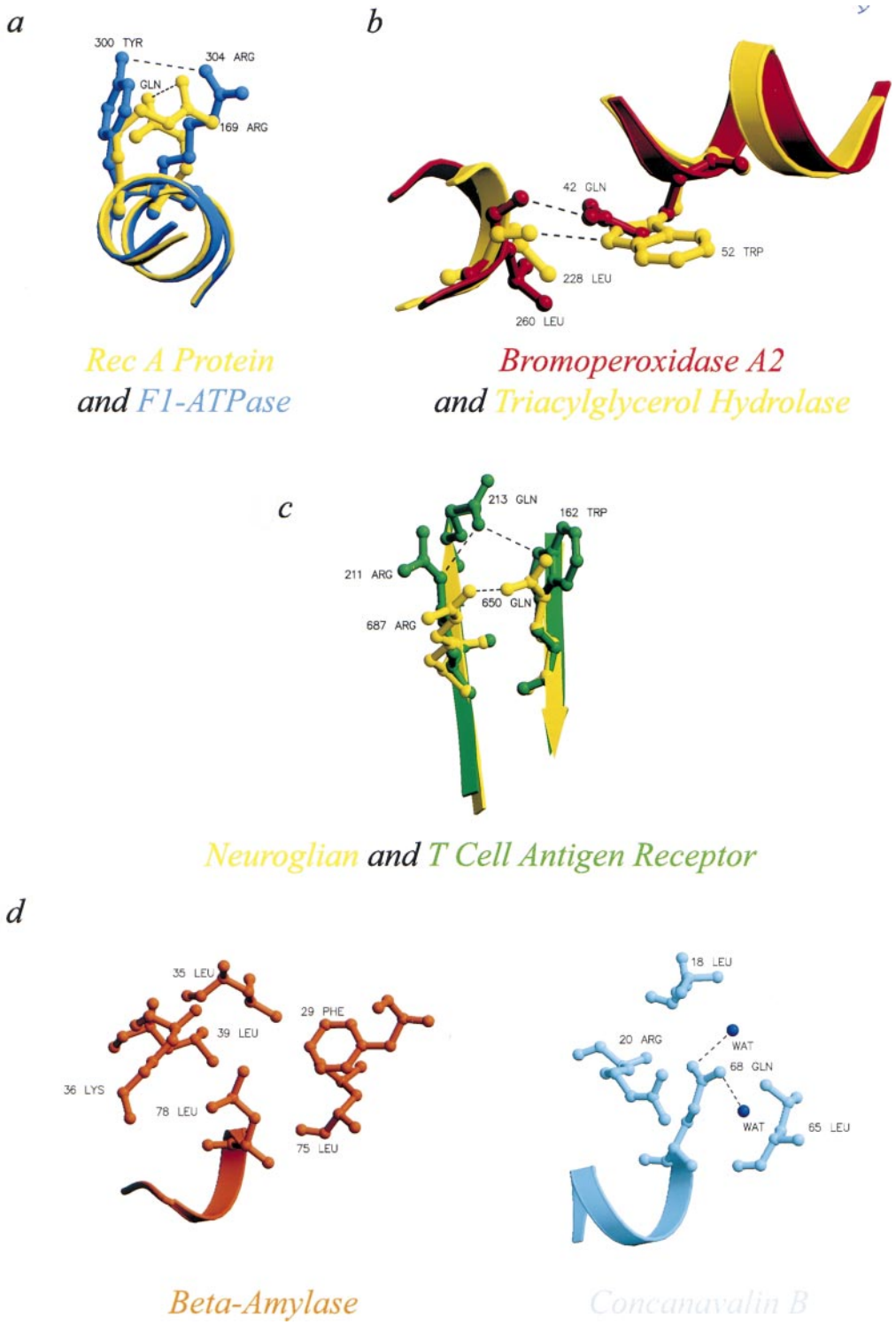

Figure 5. Four examples of pairs of 3D-1D environmentally conserved residues with dissimilar residues yet conserved structural roles. Each example is based on an alignment chosen at random from the DAPS database. (a) Rec A protein and F1-ATPase. (b) Bromoperoxidase A2 and triacylglycerol hydrolase. (c) Neruoglian and T-cell antigen receptor. (d) β-amylase and concanavalin B.

confirming that 3D-1D environmentally conserved positions tend to cluster.

3D-1D environmentally conserved positions in the lipid-binding pockets of each domain also cluster. The program CAST (Liang *et al.*, 1998) was used to identify all positions in contact with at least one lipid atom. Fifty-one positions contribute to the N-terminal lipid-binding pocket and 43 positions contribute to the C-terminal pocket. A total of 11 of these positions from the N-terminal pocket

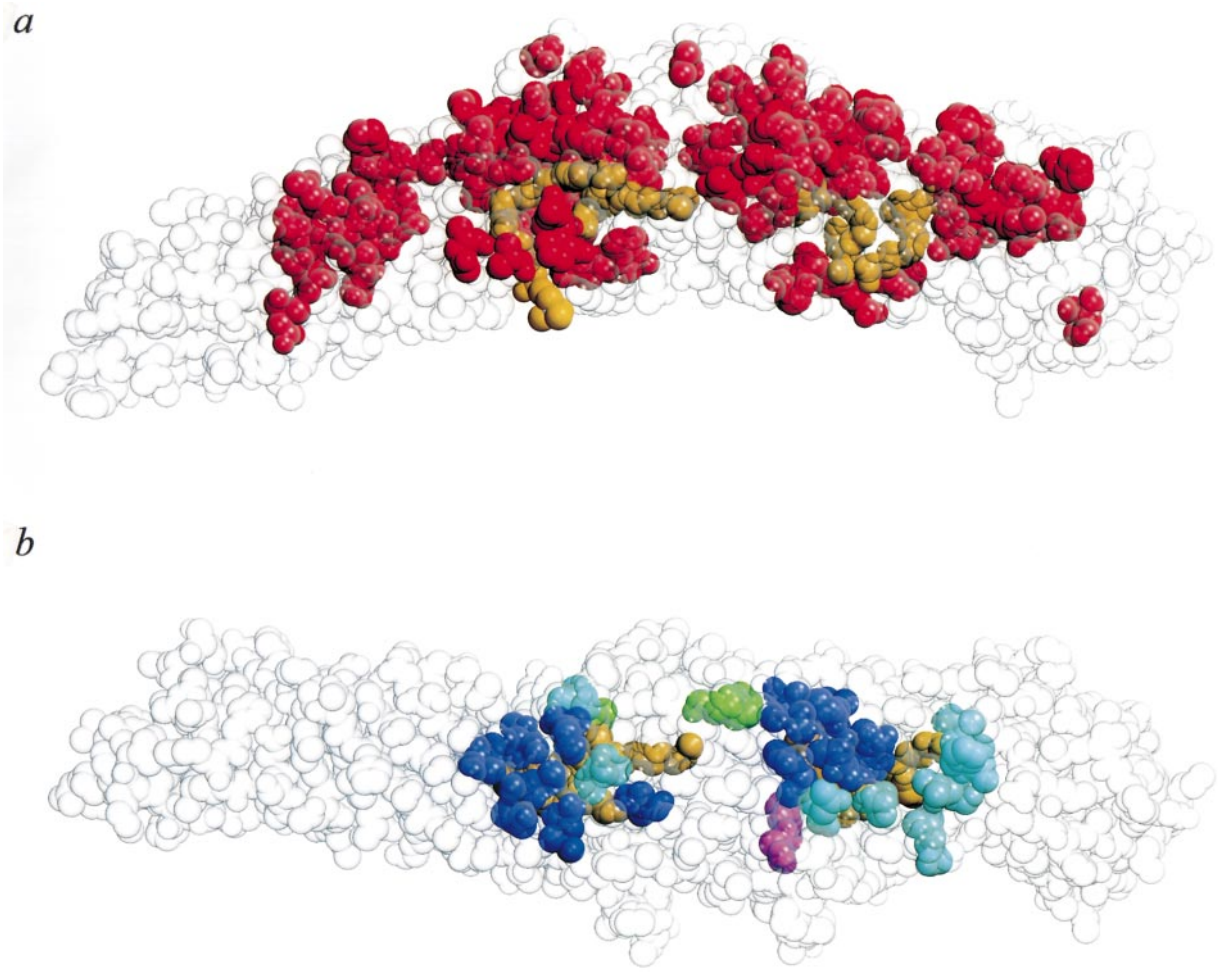

Figure 6 (*legend opposite*)

are 3D-1D environmentally conserved; 15 positions from the C-terminal pocket are 3D-1D environmentally conserved. While the number of 3D-1D environmentally conserved positions between the domains of BPI must be equal, the number of conserved positions in the lipid-binding pockets can differ. This is explained by the observation that some positions located in a lipid-binding pocket are aligned to positions that are not located in the other pocket. Therefore, only eight of the environmentally conserved positions in each pocket are paired in the BPI domain alignment.

The eight structurally equivalent positions from the lipid-binding pockets cluster in each domain (Figure 6(b)). Seven of the eight residues at these positions in the N-terminal lipid-binding pocket interact exclusively with only one acyl chain of the phospholipid (Figure 6(c)). Six of the eight residues at these positions in the C-terminal pocket interact exclusively with only one acyl chain of the lipid (Figure 6(d)). We return to the implications of this in the Discussion.

Nearly all of the 3D-1D environmentally conserved positions in the lipid-binding pockets belong to either the H:B1 (helical secondary structure and buried) or E:B1 (β-strand secondary structure and buried) environment classes. This is as one would expect because buried, hydrophobic residues contribute to pockets which bind apolar molecules. An exception occurs at the Phe228-Tyr455 pair, where the conserved environment class is E:P2. These residues are near the mouth of the lipid-binding pockets and are exposed to solvent.

Discussion

3D-1D Environment classes and the comparison of structurally equivalent positions with dissimilar residues in BPI and other proteins

Our goal is to understand how proteins with distantly related sequences fold into similar structures. The N-terminal and C-terminal domains of BPI are structurally similar, yet they display only 13 % sequence identity. Over 1000 pairs of proteins with similar structures and dissimilar sequences are available in the DAPS database.

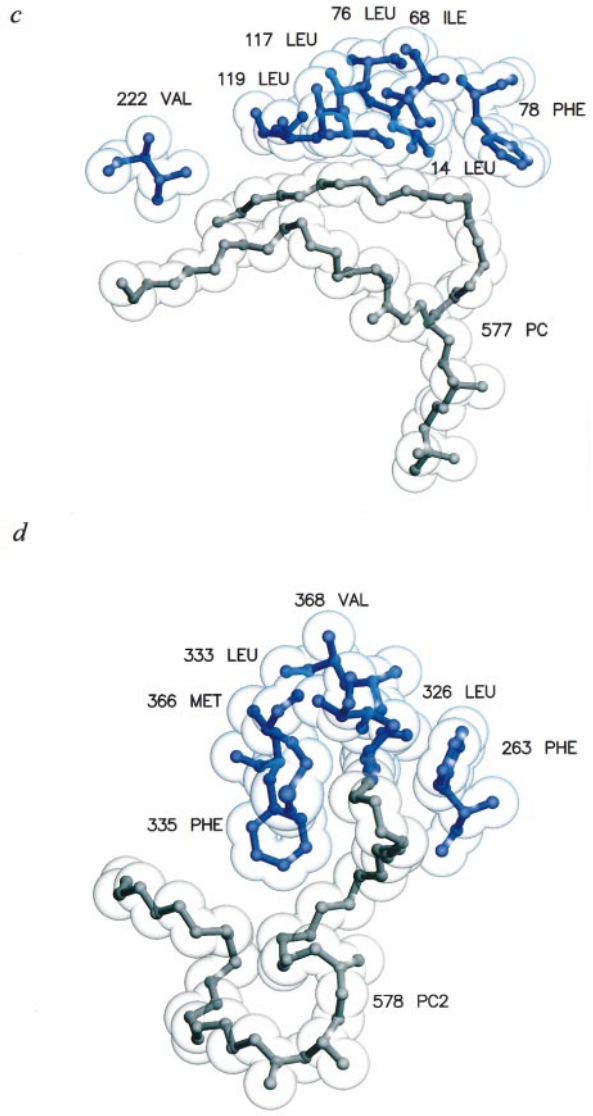

Figure 6. (a) Space-filling representation of BPI. 3D-1D environmentally conserved positions are colored red. The two bound phospholipids are gold. The N-terminal domain is shown on the left. (b) Space-filling representation of BPI rotated 90° about the *x*-axis relative to (a). Positions colored dark blue: (1) have conserved 3D-1D environments and are paired in the structural alignment; (2) contain at least one atom in contact with at least one lipid atom; (3) both positions are in contact with at least one lipid atom. Positions that are colored cyan: (1) and (2) hold but not (3). Positions that are colored green: 3D-1D environmentally conserved positions from the C-terminal domain that are in contact with the lipid-binding pocket of the N-terminal domain. Positions colored purple: 3D-1D-conserved positions from the N-terminal domain that are in contact with the lipid-binding pocket of the C-terminal domain. (c) The interaction of 3D-1D environmentally conserved positions in the N-terminal lipid-binding pocket. These seven positions satisfy the following constraints: (1) have conserved 3D-1D environments and are paired in the structural alignment; (2) contain at least one atom in contact with at least one lipid atom; (3) structurally equivalent positions in the C-terminal domain are also in contact with at least one lipid atom. The residues occupying these positions are also shown in (b) as blue spheres. The van der Waals radii are also shown for the atoms. These residues form a cluster which interacts exclusively with only one acyl chain of the phospholipid, shown as gray ball-and-sticks. (d) The interaction of 3D-1D environmentally conserved positions in the C-terminal lipid-binding pocket. These six positions also satisfy the same constraints as the equivalent N-terminal positions. The residues interact with only one acyl chain of the lipid.

To approach this question, we used the high-resolution crystal structure of human BPI to define positions that are structurally important to the BPI domain fold. Using 3D-1D environment classes, we compared the local environments of structurally equivalent positions in the two domains, assuming that positions with conserved environmental classes are structurally important.

Although only 21 of the residues in the BPI domain alignment are identical, 51 positions between the two domains have identical 3D-1D environment classes. The 13% sequence identity

between the two domains was shown to be statistically significant, with a Z-score of 3.6, and the 31% environment class identity was shown to have even greater significance, with a Z-score of 12.9. Focusing on identical environment classes reveals more clearly those positions in a fold of structural importance.

Of the 51 3D-1D environmentally conserved positions in the BPI domain alignment, 31 pairs are of identical or similar residues. We assumed the residues at these positions have similar structural roles, and this was verified by detailed examination. For example, several 3D-1D environmentally conserved positions were found in apolar environments. Residues at these positions are important for packing the BPI domain core; their structural roles are conserved (data not shown). In addition, many of the 3D-1D environmentally conserved positions located in the lipid-binding pockets of BPI also have conserved structural roles (next section). In addition to these pairs of similar residues, we assume that some of the 20 pairs of 3D-1D environmentally conserved positions with dissimilar residues also maintain similar structural roles and we chose to focus on them.

Conserved structural roles are found for four of the 20 3D-1D environmentally conserved positions with dissimilar residues. For example, we showed that the structurally equivalent residues Ser79 and Tyr336 formed hydrogen bonds to residues also located at structurally equivalent positions (Ser69 and Ser327 respectively). Intuitively, this result would be unexpected because the amino acid residue tyrosine has a side-chain of much greater length than serine. This example from BPI shows how this obstacle may be overcome; Ser79 forms a hydrogen bond to an ordered water molecule which bridges the distance to Ser69. Another example involves the 3D-1D environmentally conserved pair of Thr219 and Phe446. The amino acid residues threonine and phenylalanine are both chemically and sterically different. However, the structural role, stabilizing a β-turn, is still conserved. Thr219 forms a hydrogen bond to Thr216 and the aromatic ring of Phe446 stacks against His443. This example shows that several types of tertiary interactions, including a hydrogen bond and ring stacking, can conserve the structural roles of residues.

To explore whether these results are specific to BPI or more general, we analyzed four other protein pairs with similar structure and dissimilar sequence. As noted in Results, we found also that bromoperoxidase A2 and triacylglycerol hydrolase display 3D-1D environmentally conserved positions with dissimilar residues yet conserved structural roles. The structurally equivalent pair of dissimilar residues, Gln42 and Trp52, form long-range hydrogen bonds between residues separated over 200 residues in bromoperoxidase A2, and almost 180 residues in triacylglycerol hydrolase. We also found an example in the core of two proteins, β-amylase and concanavalin B, showing that structurally equivalent pairs of dissimilar residues in the cores of proteins can also have similar structural roles.

For the BPI domain alignment, 11 of the 20 3D-1D environmentally conserved positions with dissimilar residues have auxiliary structural roles. Six of these pairs contain at least one residue which make non-specific van der Waals contacts to other residues, while four pairs contain residues which make specific hydrogen bonds and/or salt bridges to other residues. One pair contains residues that are solvent exposed and do not contact any ordered atoms in the structure. These interactions would stabilize the fold although their role in conserving the fold is unclear.

Finally, five of the 20 3D-1D environmentally conserved positions with dissimilar residues apparently play different structural roles. Together, only 20% of the 20 environmentally conserved positions with dissimilar residues have conserved structural roles. Why do we not see more examples of pairs of dissimilar residues with conserved structural roles? The actual number is underestimated by our use of discrete environment classes. For example, it is possible that two structurally equivalent positions are in highly similar environments, yet they lie near each other on opposite sides of a boundary for two different environment classes. In this case, similar environments masquarade as being different. On the other hand, two structurally equivalent positions may be in different environments but lie on opposite extremes of the same environment class. Finally, while the BPI system is useful since its two covalently linked domains allow us to be nearly certain of their evolutionary relatedness, when these two domains are disconnected to perform the 3D-1D environment classification, some regions of the structure become solvent accessible, artificially changing their natural environments. Future research should include improved definitions of environment classes to overcome these obstacles. Despite present limitations, we detect four examples which help us understand how pairs of dissimilar residues can still have similar structural roles in BPI, as well as examples of proteins other than BPI with similar structure and dissimilar sequence.

The location of 3D-1D environmentally conserved positions in the lipid-binding pockets of BPI

Although the N-terminal and C-terminal domains of BPI both have apolar pockets which bind the phospholipid phosphatidylcholine, there are distinct differences between the two pockets: (i) the N-terminal domain lipid-binding pocket has a solvent-accessible surface area of 557 $Å^2$ and the C-terminal pocket has an area of 413 $Å^2$; (ii) the phosphocholine head group of the N-terminal phospholipid is visible in electron density maps and appears to hydrogen bond with a tyrosine residue, whereas the head group of the C-terminal

lipid is disordered; (iii) the mouth of the C-terminal lipid-binding pocket is slightly larger than the mouth of the N-terminal one; (iv) the expressed N-terminal domain retains the bactericidal and LPS-neutralization activities of the holoprotein, while a purified C-terminal fragment is not bactericidal and shows less LPS-neutralization activity than the N-terminal domain (Ooi *et al.*, 1991). The function of the C-terminal domain is to promote bacterial uptake by neutrophils. It is possible that the N-terminal domain is responsible for LPS-binding and the C-terminal domain may bind a different lipid important for the recognition of neutrophils (Elsbach & Weiss, 1998).

Despite these structural and functional differences, we observe that the BPI pockets display structural similarity. We found all 3D-1D environmentally conserved positions in the pockets and discovered that they cluster together (Figure 6(a), (b)). These clusters interact with only one acyl chain of the lipid ligand (Figure 6(c), (d)). We propose that this cluster of conserved positions represents a common structural origin between the N-terminal and C-terminal lipid-binding pockets of BPI. It also may represent a structural sub-motif that binds a single acyl chain. If the N-terminal and C-terminal lipid-binding pockets do have differing preferences for lipids, it may be true that these lipids have at least one acyl chain in common. Lipid-binding studies on recombinant versions of the N-terminal and C-terminal domains may illuminate the functions of the BPI lipid-binding pockets.

In conclusion, we have used 3D-1D environment classes to search for pairs of dissimilar residues with conserved structural roles in the N-terminal and C-terminal domains of BPI. 3D-1D environments are helpful because they do not rely on the residue identities of the structurally equivalent positions. Analysis of structurally equivalent pairs with dissimilar residue types provides additional insight into how two dissimilar protein sequences adopt similar folds. For BPI, clustering of environmentally conserved positions in the two BPI lipid-binding pockets identifies a structural sub-motif which may have evolved to bind single acyl chains of lipid molecules.

Materials and Methods

Refinement of the 1.7 Å crystal structure of BPI

Crystals of recombinant human BPI were grown as described (Beamer *et al.*, 1997). Crystals were equilibrated over-night to 25% (v/v) PEG 6000 by vapor diffusion, and then immediately into 45% PEG 6000 for approximately two minutes. The crystals were then mounted and frozen under a stream of liquid nitrogen. X-ray diffraction data were collected at the Brookhaven National Laboratory on beamline X-12B. Data were processed and scaled using the programs DENZO and SCALEPACK (Otwinowski, 1993).

The unit cell dimensions of frozen crystals differ from the room temperature crystal dimensions, especially the *b*-axis which contracted 16%. Rigid-body refinement in X-PLOR (Brünger, 1990) was used to reorient the room-temperature BPI model (protein atoms only) in the smaller unit cell, followed by simulated annealing and individual *B*-factor refinement. This model had an *R*-factor of 28.4% and R_{free} of 34.3%. Electron density for two molecules of phosphatidylcholine, each located in an apolar pocket in both domains of BPI, was apparent in both $2F_o - F_c$ and $F_o - F_c$ maps. Both lipids were also found during tracing of the original model, and likely were introduced to BPI during protein purification. Modeling of the N-terminal phospholipid reduced the *R*-factor to 27.3%, and R_{free} to 33.5%. Approximately 360 water molecules and the C-terminal phospholipid (excluding the disordered head) were gradually introduced into the model and refined to an *R*-factor of 21.3% and R_{free} of 29.6%.

Once we were not able to lower the *R*-factors using X-PLOR, we switched to the program CNS which uses a maximum-likelihood target function to minimize the gap between *R* and R_{free} (Brünger *et al.*, 1998). Upon using CNS we observed a noticeable improvement in the quality of the electron density maps which allowed us to further rebuild the model. Following a bulk-solvent and overall anisotropic *B*-factor correction, our final *R*-factor is 19.8% and final R_{free} is 24.9%. Refinement statistics are summarized in Table 4.

The following residues had weak density and therefore were not included in the refinement: K33, K44, K45, K86, K95, K118, K290, K307, K313. Multiple conformations were built for the following residues: Q26, D57, S79, S88, S107, H143, T179, S183, V243, H325, L326, T363, N373, R374, F425, N437, V453, V454.

Several structure validation methods were used to assess the quality of the high-resolution BPI model. No significant errors were found in the model using the programs VERIFY 3D, ERRAT, PROCHECK and WHAT IF (Colovos & Yeates, 1993; Laskowski *et al.*, 1993; Luthy *et al.*, 1992). Over 89% of the residues are in the most favorable position in a Ramachandran plot. Approximately 10% of the residues are in the additionally allowed regions. Composite, simulated-annealing omit maps were calculated for the entire molecule. Density from the omit maps agreed very well with the atomic positions of nearly all residues.

Structural alignment and calculation of 3D-1D profiles for the N-terminal and C-terminal domain

Coordinates for residues 1 to 229 corresponding to the N-terminal domain of BPI and residues 251 to 456 corresponding to the C-terminal domain were used for rigid-body alignment of the two domains. The program Align_v2 was used for the superposition, which outputs a structure-based sequence alignment of the two domains (Cohen, 1986). All main-chain atoms were used for both the superposition and calculation of rmsd. Initially 173 positions were aligned by the program. Nine pairs had no structurally equivalent main-chain atoms within 5 Å and were not considered aligned. They are shown in Figure 2 by stars. Therefore, the length of the entire BPI domain alignment is 164 positions.

3D-1D profiles were generated for both the N-terminal domain and the C-terminal domain (Bowie *et al.*, 1991). The boundaries used for calculating environment classes are shown in Table 5. The secondary structure at each residue was characterized using the program DSSP

Table 4. Statistics on crystallographic data and refinement for BPI

A. *Data Collection*	
Resolution range high (Å)	1.7
Resolution range low (Å)	50.0
Data cut-off (Sigma (F))	0.0
Overall completeness (final res. shell)	94.1 (77.2)
R_{merge} (%)	4.9
Redundancy	2.3
Number of reflections (test set)	4755
Number of reflections (overall)	47,197
B. *Crystal*	
Space group	*C*2
Unit cell parameters (Å)	$a = 184.3$, $b = 31.2$, $c = 80.6$, $\beta = 103.2^\circ$
C. *Refinement*	
R-value (%)	19.8
Free *R*-value (%)	24.9
Mean *B*-value (all atoms) (Å^2)	29.8
Estimated coordinate error (Luzzati) (Å)	0.20
rmsd from ideal values	
Bond lengths (Å)	0.017
Bond angles (deg.)	2.0
Dihedral angles (deg.)	25.9
Improper angels (deg.)	1.33

(Kabsch & Sander, 1983). 3D-1D profiles and secondary structures were calculated in a similar manner for examples from the DAPS database.

Calculation of *Z*-scores for assessing the statistical significance of % residue identity and % environment class identity in the BPI domain alignment

The % residue identity expected from two independently generated random sequences of 164 residues was calculated. The expected frequency of matches or % identity is given by:

$$\langle P \rangle = \sum_{i=1}^{m} p_i^2$$

where p_i is the probability of observing a residue of type i at a given position in a random sequence and m is the number of residue types, which in the case of a sequence alignment is 20 amino acid residues. The standard deviation, σ_p, in the frequency of matches is:

$$\sigma_p = \sqrt{P(1-P)/N}$$

where P is given above and N is the length of the alignment. Probabilities for each amino acid residue, as well as each environment class, were derived from the DAPS database.

The *Z*-score is given by:

$$Z = (P_{obs} - \langle P \rangle)/\sigma_p$$

Where P_{obs} is the observed % identity for either residues or environment classes between N-terminal and C-terminal domains of BPI. $\langle P \rangle$ and σ_p are both given above.

Log-odds values and fractional weighted log-odds values for each environment class based on the alignment of the N-terminal and C-terminal domains of BPI

To further compare and contrast structural properties between N-terminal and C-terminal domains, three statistical properties were computed for each environment class based upon the structural alignment of BPI. For each environment class, the statistical significance of observing pairs of positions (brought together by the alignment) both belonging to a given environment class was assessed by computing a *p*-value. The *p*-value for a given environment class is the probability that one would observe at least as many matches for that class in a random alignment where M is the number of observed matches of that type in the BPI domain structural alignment. The *p*-value is given by the following equation:

$$p = \sum_{X_{12}=M}^{Min(X_1,X_2)} \frac{(N-X_1)!(N-X_2)!X_1!X_2!}{(X_1-X_{12})!(X_2-X_{12})!(N-X_1-X_2+X_{12})!X_{12}!N!}$$

where X_1 is the number of observations for a given environment class from the N-terminal domain, X_2 is the number of observations for the same environment class for the C-terminal domain, and N is the total number of pairs in the alignment. Gaps are not considered. Several of the environment classes have low *p*-values, and therefore when they are paired in the domain alignment it is of statistical significance.

Second, the log-odds value for each environment class was calculated. The log-odds ratio for a given pair of environment classes is a measure of how likely it is that any given instance of an environment class is conserved in an alignment. Scores well above zero imply significance, whereas scores near zero imply that the observation is likely to occur when environments are paired randomly. The log-odds value, $LO(e)$ is given by:

$$LO(e) = \log\left[\frac{P_{AB}(e)}{P_A(e)P_B(e)}\right]$$

where $P_{AB}(e)$ is the joint probability of observing environment class e at aligned positions in both sequence A (N-terminal domain) and B (C-terminal domain). $P_A(e)$ is the probability of observing the environment class in sequence A, while $P_B(e)$ is the probability of observing the same class in sequence *B*.

Lastly, the fractional weighted log-odds value (weighted by the joint probability), FWLO(e), was calculated for each environment class. This value is expressed as a fraction of the total weighted log-odds scores for all environment classes. The FWLO(e) value is given by:

$$FWLO(e) = \left[\frac{LO(e) \times P_{AB}(e)}{\sum_{env.classes} LO(e) \times P_{AB}(e)}\right]$$

A given pair may have a positive $LO(e)$, but if it occurs only rarely in the structural alignment, the $FWLO(e)$ value will be low. Table 2 lists the $LO(e)$ and $FWLO(e)$ values for all identical environment pairs for both the N-terminal and C-terminal domains of BPI.

The H:P2, H:B1, E:P2 and E:B1 environment classes dominate the distribution of $LO(e)$ and $FWLO(e)$ values for each identical environment pair. Large $LO(e)$ values

Table 5. Boundaries for area buried (Ab) and fraction of the area buried by polar atoms (Fp) used for defining 3D-1D environment classes in the structures discussed in this paper. Each position defined by values of Ab and Fb can be occupied by a residue that is in helix (H), β-strand (E), or coil (C)

Environment class			Area buried (Å^2)	Fraction polar
H:B1	E:B1	C:B1	Ab $\geqslant$ 114	Fp $\leqslant$ 0.33
H:B2	E:B2	C:B2	Ab $\geqslant$ 114	0.33 $\leqslant$ Fp $\leqslant$ 0.46
H:B3	E:B3	C:B3	Ab $\geqslant$ 114	0.46 $\leqslant$ Fp $\leqslant$ 0.57
H:P1	E:P1	C:P1	40 $\leqslant$ Ab $\leqslant$ 114	0.46 $\leqslant$ Fp $\leqslant$ 0.57
H:P2	E:P2	C:P2	40 $\leqslant$ Ab $\leqslant$ 114	0.57 $\leqslant$ Fp $\leqslant$ 1.00
H:E	E:E	C:E	Ab $\leqslant$ 40	0.00 $\leqslant$ Fp $\leqslant$ 1.00

are observed for other environment classes, such as H:E or E:E, but these are relatively rare in BPI and thus unlikely to be as important. This is reflected in the low FWLO(e) values for the H:E and E:E pairs. High FWLO(e) values are observed for the H:P2, H:B1, E:P2 and E:B1 environment classes and are expected play a dominant structural role in the BPI domain.

Distribution of log-odds values and fractional weighted log-odds values for all structures in the DAPS database

To examine whether the H:P2, H:B1, E:P2 and E:B1 environment classes had structural significance to proteins other than BPI, we analyzed proteins in the database of aligned protein structures (DAPS) (Rice & Eisenberg, 1997). DAPS contains 1074 structurally-derived sequence alignments between structurally homologous proteins with less than 25 % sequence identity.

LO(*e*) and *FWLO*(*e*) values were calculated over all protein pairs in DAPS (Table 2). The H:P2, H:B1 and E:B1 residue environments had large *LO(e)* and *FWLO(e)* values. However, the *FWLO(e)* value for the E:P2 class is only slightly higher than values for several of the other environment classes. The importance of positions belonging the E:P2 environment class is not general and may be unique to the BPI domain.

Accession codes

The refined coordinates of BPI at 1.7 Å resolution have been deposited in the RCSB PDB with deposition code 1EWF.

Acknowledgments

We thank Edward Marcotte, Ioannis Xenarios, Michael Sawaya, Lukasz Salwinski, and Melinda Balbirnie for helpful discussion. We are grateful to XOMA and Steven Carroll for providing us with pure BPI protein. We also thank Duilio Cascio and Malcolm Capel for data collection and structure refinement. The work of G.K. was funded in part by USPHS Training Grant GM07185. This work was supported in part by the NIH and the Department of Energy. The work of P.M. was funded in part by Institutional National Research Service Award GM08375 from USHHS (NIH/NIGMS). L.J.B. was supported by a grant from the University of Missouri Research Board.

References

Abrahams, J. P., Leslie, A. G., Lutter, R. & Walker, J. E. (1994). Structure at 2.8 Å resolution of F1-ATPase from bovine heart mitochondria. *Nature,* **370**, 621-628.

Banks, R. D., Blake, C. C., Evans, P. R., Haser, R., Rice, D. W., Hardy, G. W., Merrett, M. & Phillips, A. W. (1979). Sequence, structure and activity of phosphoglycerate kinase: a possible hinge-bending enzyme. *Nature,* **279**, 773-777.

Beamer, L. J., Carroll, S. F. & Eisenberg, D. (1997). Crystal structure of human BPI and two bound phospholipids at 2.4 Ångstrom resolution. *Science,* **276**, 1861-1864.

Benner, S. A., Cohen, M. A. & Gonnet, G. H. (1994). Amino acid substitution during functionally constrained divergent evolution of protein sequences. *Protein Eng.* **7**, 1323-1332.

Bentley, G. A., Boulot, G., Karjalainen, K. & Mariuzza, R. A. (1995). Crystal structure of the beta chain of a T cell antigen receptor. *Science,* **267**, 1984-1987.

Bowie, J. U., Luthy, R. & Eisenberg, D. (1991). A method to identify protein sequences that fold into a known three-dimensional structure. *Science,* **253**, 164-170.

Brünger, A. T. (1990). *X-PLOR: A System for Crystallography and NMR*, Yale University Press, New Haven, CT.

Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallog. sect. D,* **54**, 905-921.

Cohen, G. H., Padlan, E. A. & Davies, D. R. (1986). Phosphocholine binding immunoglobulin Fab McPC603. An X-ray diffraction study at 2.7 Å. *J. Mol. Biol.* **190**, 593-604.

Colovos, C. & Yeates, T. O. (1993). Verification of protein structures: patterns of nonbonded atomic interactions. *Protein Sci.* **2**, 1511-1519.

Divne, C., Ståhlberg, J., Reinikainen, T., Ruohonen, L., Pettersson, G., Knowles, J. K., Teeri, T. T. & Jones, T. A. (1994). The three-dimensional crystal structure of the catalytic core of cellobiohydrolase I from *Trichoderma reesei*. *Science,* **265**, 524-528.

Elsbach, P. & Weiss, J. (1995). Prospects for use of recombinant BPI in the treatment of Gram-negative bacterial infections. *Infect. Agents Dis.* **4**, 102-109.

Elsbach, P. & Weiss, J. (1998). Role of the bactericidal/permeability-increasing protein in host defence. *Curr. Opin. Immunol.* **10**, 45-49.

Emsley, J., White, H. E., O'Hara, B. P., Oliva, G., Srinivasan, N., Tickle, I. J., Blundell, T. L., Pepys,

M. B. & Wood, S. P. (1994). Structure of pentameric human serum amyloid P component. *Nature,* **367**, 338-345.

Hecht, H. J., Sobek, H., Haag, T., Pfeifer, O. & van Pée, K. H. (1994). The metal-ion-free oxidoreductase from *Streptomyces aureofaciens* has an alpha/beta hydrolase fold. *Nature Struct. Biol.* **1**, 532-537.

Hennig, M., Jansonius, J. N., Terwisscha van Scheltinga, A. C., Dijkstra, B. W. & 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.

Hol, W. G., Lijk, L. J. & Kalk, K. H. (1983). The high resolution three-dimensional structure of bovine liver rhodanese. *Fund. Appl. Toxicol.* **3**, 370-376.

Horwitz, A. H., Leigh, S. D., Abrahamson, S., Gazzano-Santoro, H., Liu, P. S., Williams, R. E., Carroll, S. F. & Theofan, G. (1996). Expression and characterization of cysteine-modified variants of an amino-terminal fragment of bactericidal/permeability-increasing protein. *Protein Exp. Purif.* **8**, 28-40.

Huber, A. H., Wang, Y. M., Bieber, A. J. & Bjorkman, P. J. (1994). Crystal structure of tandem type III fibronectin domains from *Drosophila* neuroglian at 2.0 Å. *Neuron,* **12**, 717-731.

Kabsch, W. & Sander, C. (1983). Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers,* **22**, 2577-2637.

Laskowski, R. A., McArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallog.* **26**, 283-291.

Liang, J., Edelsbrunner, H. & Woodward, C. (1998). Anatomy of protein pockets and cavities: measurement of binding site geometry and implications for ligand design. *Protein Sci.* **7**, 1884-1897.

Luthy, R., Bowie, J. U. & Eisenberg, D. (1992). Assessment of protein models with three-dimensional profiles. *Nature,* **356**, 83-85.

Mikami, B., Degano, M., Hehre, E. J. & Sacchettini, J. C. (1994). Crystal structures of soybean beta-amylase reacted with beta-maltose and maltal: active site components and their apparent roles in catalysis. *Biochemistry,* **33**, 7779-7787.

Mitchell, J. B., Nandi, C. L., McDonald, I. K., Thornton, J. M. & Price, S. L. (1994). Amino/aromatic interactions in proteins: is the evidence stacked against hydrogen bonding? *J. Mol. Biol.* **239**, 315-331.

Ooi, C. E., Weiss, J., Doerfler, M. E. & Elsbach, P. (1991). Endotoxin-neutralizing properties of the 25 kD N-terminal fragment and a newly isolated 30 kD C-terminal fragment of the 55-60 kD bactericidal/permeability-increasing protein of human neutrophils. *J. Exp. Med.* **174**, 649-655.

Otwinowski, Z. (1993). *Proceedings of CCP4 Study Weekend: Data Collection and Processing* (Sawyer, L., Isaacs, N. & Baileys, S., eds), pp. 56-62, SERC Daresbury Laboratory, Warrington.

Ploegman, J. H., Drent, G., Kalk, K. H. & Hol, W. G. (1978). Structure of bovine liver rhodanese. I. Structure determination at 2.5 Å resolution and a comparison of the conformation and sequence of its two domains. *J. Mol. Biol.* **123**, 557-594.

Quiocho, F. A., Gilliland, G. L. & Phillips, G. N., Jr (1977). The 2.8 Å resolution structure of the L-arabinose-binding protein from *Escherichia coli*. Polypeptide chain folding, domain similarity, and probable location of sugar-binding site. *J. Biol. Chem.* **252**, 5142-5149.

Rice, D. W. & Eisenberg, D. (1997). A 3D-1D substitution matrix for protein fold recognition that includes predicted secondary structure of the sequence. *J. Mol. Biol.* **267**, 1026-1038.

Steitz, T. A., Fletterick, R. J., Anderson, W. F. & Anderson, C. M. (1976). High resolution X-ray structure of yeast hexokinase, an allosteric protein exhibiting a non-symmetric arrangement of subunits. *J. Mol. Biol.* **104**, 197-122.

Story, R. M., Weber, I. T. & Steitz, T. A. (1992). The structure of the *E. coli* recA protein monomer and polymer. *Nature,* **355**, 318-325.

Subramanian, E., Liu, M., Swan, I. D. & Davies, D. R. (1977). The crystal structure of an acid protease from *Rhizopus chinensis* at 2.5 Å resolution. *Advan. Exp. Med. Biol.* **95**, 33-41.

Uppenberg, J., Hansen, M. T., Patkar, S. & Jones, T. A. (1994). The sequence, crystal structure determination and refinement of two crystal forms of lipase B from *Candida antarctica*. *Structure,* **2**, 293-308.

Edited by D. Rees

(Received 21 December 1999; received in revised form 3 April 2000; accepted 19 April 2000)