The 1.7 Å crystal structure of BPI: a study of how two dissimilar amino acid sequences can adopt the same fold

Parag Mallick

Journal of Molecular Biology

Mylan v. Regeneron
IPR2021-00880
U.S. Pat. 9,669,069
Exhibit 2069

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*

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.

© 2000 Academic Press

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

*Corresponding author

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

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-
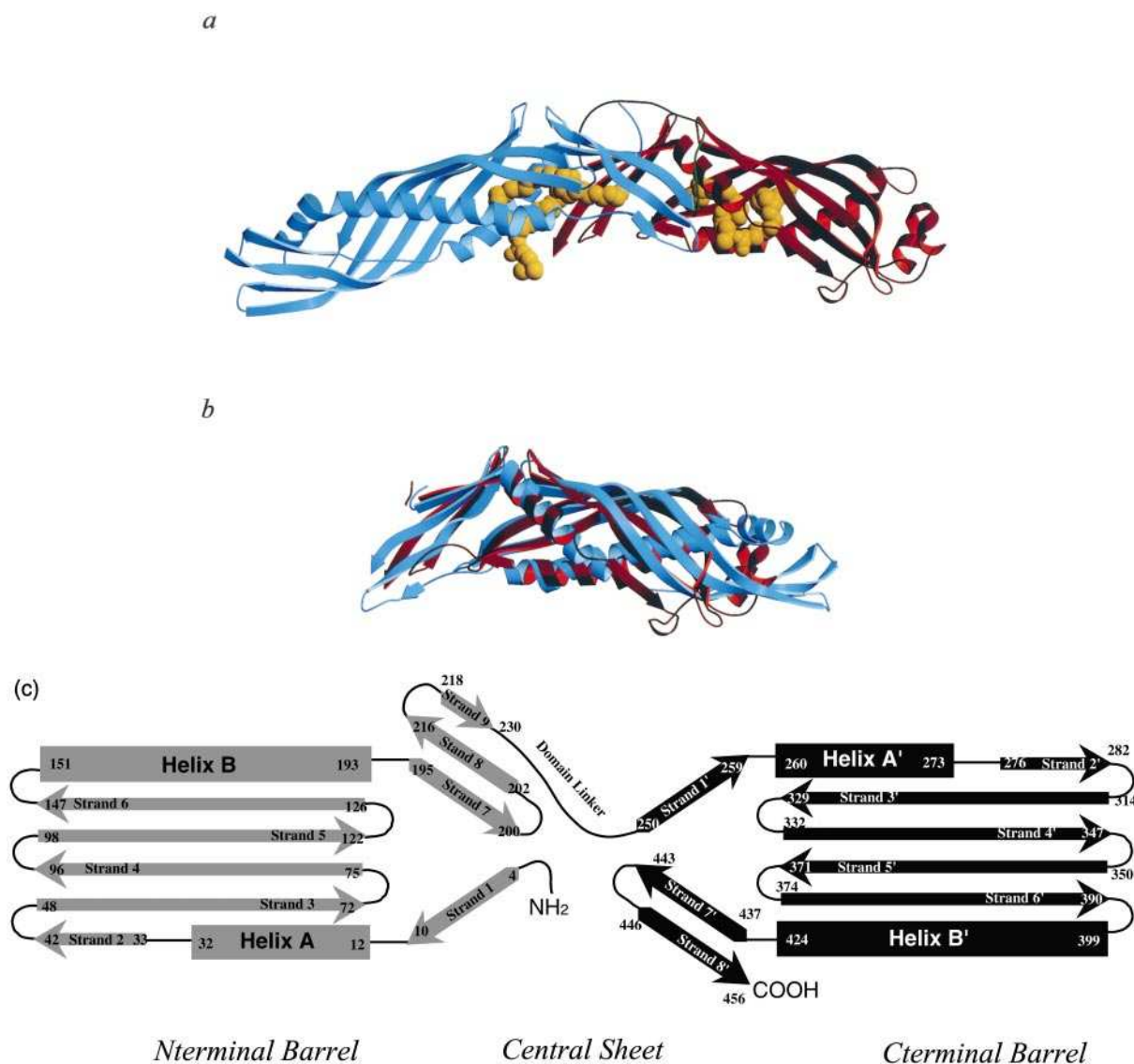
Figure 1. (a) Ribbon representation of the 1.7 Å crystal structure of human BPI. The N-terminal domain is blue. The C-terminal domain is red. Residues 10 to 193 fold into a structural element called the N-terminal barrel. The barrel is composed of five anti-parallel β-strands which twist about the barrel 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
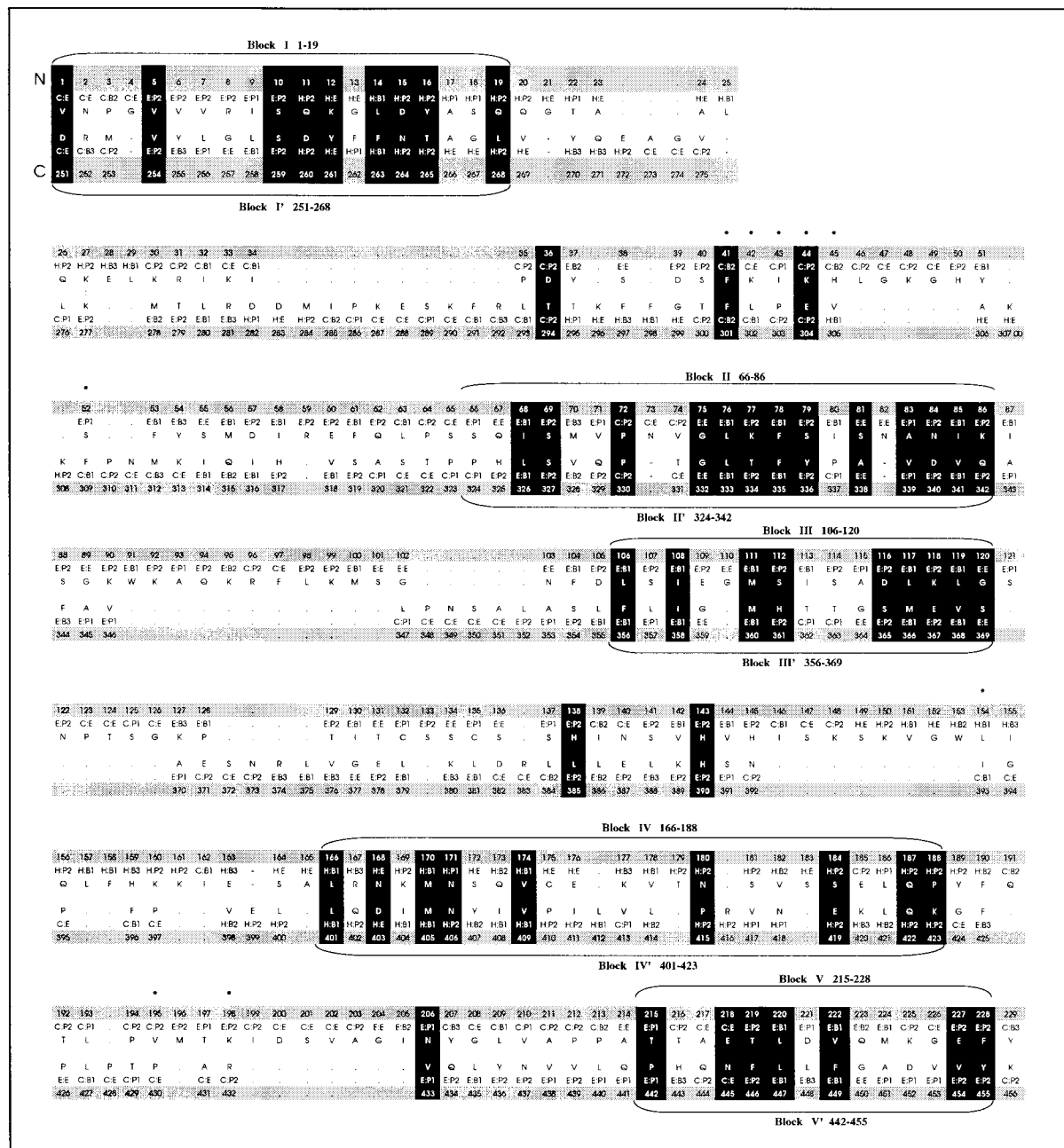
Figure 2. Structure-based alignment of the amino acid sequences of the N-terminal and C-terminal domains of BPI. Both the sequence and corresponding 3D-1D environments are shown. The N-terminal domain (residues 1 to 229) is the top sequence; the C-terminal domain (residues 251 to 456) is the bottom of 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

DOCKET ALARM

Explore Litigation Insights

Docket Alarm provides insights to develop a more informed litigation strategy and the peace of mind of knowing you're on top of things.

Real-Time Litigation Alerts

Keep your litigation team up-to-date with **real-time alerts** and advanced team management tools built for the enterprise, all while greatly reducing PACER spend.

Our comprehensive service means we can handle Federal, State, and Administrative courts across the country.

Advanced Docket Research

With over 230 million records, Docket Alarm's cloud-native docket research platform finds what other services can't. Coverage includes Federal, State, plus PTAB, TTAB, ITC and NLRB decisions, all in one place.

Identify arguments that have been successful in the past with full text, pinpoint searching. Link to case law cited within any court document via Fastcase.

Analytics At Your Fingertips

Learn what happened the last time a particular judge, opposing counsel or company faced cases similar to yours.

Advanced out-of-the-box PTAB and TTAB analytics are always at your fingertips.

API

Docket Alarm offers a powerful API (application programming interface) to developers that want to integrate case filings into their apps.

LAW FIRMS

Build custom dashboards for your attorneys and clients with live data direct from the court.

Automate many repetitive legal tasks like conflict checks, document management, and marketing.

FINANCIAL INSTITUTIONS

Litigation and bankruptcy checks for companies and debtors.

E-DISCOVERY AND LEGAL VENDORS

Sync your system to PACER to automate legal marketing.