doi:10.1016/j.jmb.2004.03.063 J. Mol. Biol. (2004) 339, 459–469

JMB

Available online at www.sciencedirect.com

SCIENCE DIRECT®

ELSEVIER

Sensor Domain of the *Mycobacterium tuberculosis* Receptor Ser/Thr Protein Kinase, PknD, forms a Highly Symmetric β Propeller

Matthew C. Good, Andrew E. Greenstein, Tracy A. Young Ho-Leung Ng and Tom Alber*

Department of Molecular and Cell Biology, University of California, 339 Hildebrand Hall #3206, Berkeley, CA 94720-3206, USA

*Corresponding author

Diverse pathogenic bacteria produce transmembrane receptor Ser/Thr protein kinases (STPKs), but little is known about the signals mediated by these "eukaryotic-like" proteins. To explore the basis for signaling in the bacterial STPK receptor family, we determined the structure of the sensor domain of *Mycobacterium tuberculosis* PknD. In two crystal forms, the PknD sensor domain forms a rigid, six-bladed β-propeller with a flexible tether to the transmembrane domain. The PknD sensor domain is the most symmetric β-propeller structure described. All residues that vary most among the blade subdomains cluster in the large "cup" motif, analogous to the ligand-binding surface in many β-propeller proteins. These results suggest that PknD binds a multivalent ligand that signals by changing the quaternary structure of the intracellular kinase domain.

Keywords: bacterial Ser/Thr protein kinases; NHL repeat; YVTD β propeller; protein evolution; receptor recognition

Tuberculosis is the world's leading infectious disease, killing over two million people annually.[1,2] The causative bacterial agent, *Mycobacterium tuberculosis* (*Mtb*), has a distinctive life cycle spanning different environments and developmental stages. The developmental programs and the molecular signals that mediate environmental adaptations are unknown. Along with two-component systems, 11 putative Ser/Thr protein kinases (STPKs) encoded in the *Mtb* genome[3,4] provide candidates for the sensors that initiate and maintain distinct developmental and metabolic states. Nine of the STPKs encode putative transmembrane receptors with an intracellular, N-terminal kinase domain linked to an extracellular, C-terminal sensor domain. The sequences of bacterial and phage genomes obtained in recent years revealed that members of this "eukaryotic-like" STPK family occur in numerous pathogenic bacteria.

Six of the *Mtb* STPKs or kinase domains have been expressed, purified and shown to be active *in vitro*.[5–9] The most studied *Mtb* STPK is PknB. Like eukaryotic STPKs, PknB is activated by autophosphorylation of the activation loop.[10–12] In addition, phosphorylation of two sites in the linker joining the kinase domain to the predicted transmembrane helix may regulate the binding of unidentified regulatory factors. Two crystal structures of the PknB kinase domain[12,13] revealed close structural similarities to eukaryotic STPKs, suggesting that mechanisms of regulation and substrate recognition are universal in this protein family. We proposed that dimerization through an intermolecular interface conserved in PknB orthologs may mediate inhibition and signaling of these STPKs.[12]

Less is known about the structures and functions of the bacterial STPK sensor domains. Genetic evidence suggests that the sensor and transmembrane domains of the *Bacillus subtilis* PknB homolog are capable of dimerizing *in vivo*.[10] The sensor domains of *Mtb* PknB and PknD have been predicted to contain PASTA domains[12,14] and a

Present address: Department of Biochemistry and Biophysics, University of California, San Francisco, CA 94143-2240, USA.

Abbreviations used: MAD, multiwavelength anomalous diffraction; STPKs, Ser/Thr protein kinases; SLP, surface-layer-protein; BraT, brain tumor; LDLR, low density lipoprotein receptor; *Mtb*, *Mycobacterium tuberculosis*; SLP, surface layer protein; rmsd, root-mean-square deviation.

E-mail address of the corresponding author: tom@ucxray.berkeley.edu

Table 1. Data collection, phasing and refinement statistics for the PknD sensor domain

Data collection and phasing				
Crystal symmetry	$P2_1$	$P2_1$	$P2_1$	$P2_12_12_1$
Sample	Thr606Cys PCMBS	Thr606Cys PCMBS	Native	Native
Wavelength (Å)	1.0092	0.9698	1.0781	1.0719
Resolution (Å)	75–3.0	75–3.0	75–1.8	75–1.9
Completeness (%)	97.6 (98.4)	97.5 (98.3)	99.4 (99.4)	99.1 (98.9)
Multiplicity	3.8 (3.9)	3.8 (3.9)	4.1 (3.9)	3.5 (3.6)
R_{merge} (%)[a,b]	3.6 (10.6)	3.7 (11.3)	5.2 (30.8)	6.9 (51.4)
$\langle I/\sigma I\rangle$[b]	21.0 (13.7)	20.2 (13.1)	18.0 (3.9)	8.5 (2.8)
Phasing power[c]	0.29/0.47			
Mean figure of merit[d]	0.282; 0.791 after solvent flattening			
Refinement				
Crystal symmetry	$P2_1$	$P2_12_12_1$		
Resolution (Å)	15.4–1.8	30.–1.9		
Reflections	50,571	17,367		
R_{cryst}/R_{free} (%)[e]	0.190/0.220	0.214/0.262		
rms Δ bonds, Δ angles[f]	0.013 Å, 1.476°	0.017 Å, 1.634°		
Average *B*-factor (Å^2)	23.8	32.9		
Main-chain dihedral angles	Most-favored, 86.1%, allowed, 13.9%	Most-favored, 84.7%, allowed, 15.3%		

PCMBS; *p*-chloromercurybenzenesulfonic acid.
[a] $R_{merge} = \sum|I - \langle I\rangle|/\sum I$; *I*, intensity.
[b] Parentheses denote values for highest resolution shell.
[c] Phasing power (dis/ano) $= [\sum_n |F_H|^2/\sum_n |E|^2|^{1/2}$; F_H, calculated heavy atom scattering factor; *E*, lack of closure error.
[d] Mean figure of merit $= \langle|\sum_\alpha P(\alpha)e^{i\alpha}/\sum_\alpha P(\alpha)|\rangle$; α, phase; $P(\alpha)$, phase probability distribution.
[e] Root mean square deviations from ideal values.
[f] $R_{cryst} = \sum|F_o - F_{calc}|/\sum F_o$; F_o, observed structure-factor amplitude; F_{calc}, calculated structure-factor amplitude.

NHL-repeat β propeller, respectively.[15,16] Mutant phenotypes and the nature of genes neighboring the STPK genes have supported working hypotheses for the functions of the bacterial STPKs.[4] For example, PknA and PknB have been proposed to regulate cell division,[6,17] and PknD has been associated with phosphate transport.[4] However, no structures, ligands, or oligomerization states of the sensor domains have been reported.

Here we describe the X-ray crystal structures of two crystal forms of the extracellular sensor domain of *Mtb* PknD. This domain forms a six-bladed β propeller built on a remarkably rigid and symmetric framework. The sites that differ most among the sequences of the six blades cluster in the large concave "cup" of the β propeller. Many β-propeller proteins bind cognate ligands on the analogous surface. These results restrict models for signaling through PknD.

Results

To characterize the sensor domain, we expressed and purified *Mtb* PknD residues 403–664. This sequence starts at the first charged residue following the predicted transmembrane helix. The expressed protein included a C-terminal His_6 tag that was not removed. This construct crystallized in two forms (Table 1). Because no suitable heavy atom derivatives of the wild-type protein were identified, we created five mutants that introduced a single cysteine, as well as a mutant that contained two Leu-to-Met substitutions. The cysteine mutations were introduced at the positions of threonine residues predicted to be partially solvent exposed in a homology model produced by the program 3D-PSSM.[18,19] Methionine residues were substituted for two leucine resides predicted to be buried in the homology model.

The structure of the monoclinic ($P2_1$) crystal form was determined at 1.8 Å resolution by multi-wavelength anomalous diffraction (MAD) analysis of a Hg derivative of the Thr606Cys mutant (Table 1, Figure 1A). The partially refined model of a monomer was used to determine phases for the orthorhombic ($P2_12_12_1$) crystal form by molecular replacement. The His_6 tag was fully or partially ordered in intermolecular contacts in both crystal forms, and the tag bound either seven ($P2_1$) or four ($P2_12_12_1$) Cd ions per monomer. Two regions of the structure were disordered in all three independent monomers: a four amino acid loop (455–458) in blade 1 and the N-terminal 11 residues (403–413) that connect the sensor domain to the transmembrane domain.

In both crystal forms, the PknD sensor domain forms a β propeller containing six blades arranged cyclically around a central pore (Figure 1(B)). Each blade contains four antiparallel β strands, with the first strand in the center of the structure and the last strand at the outer edge. The loop connecting strand 4 to strand 1 of the next blade (the 4–1 loop) and the loop connecting strands 2 and 3 (the 2–3 loop) form the "cup" of the β propeller. The 1–2 and 3–4 loops are on the opposite face of the structure. The PknD C terminus forms β strand 3 of the sixth blade, and the N terminus of the protein forms strand 4 of the sixth blade. This topological overlap creates a "latch" that knits the structure together (Figure 1(B)).

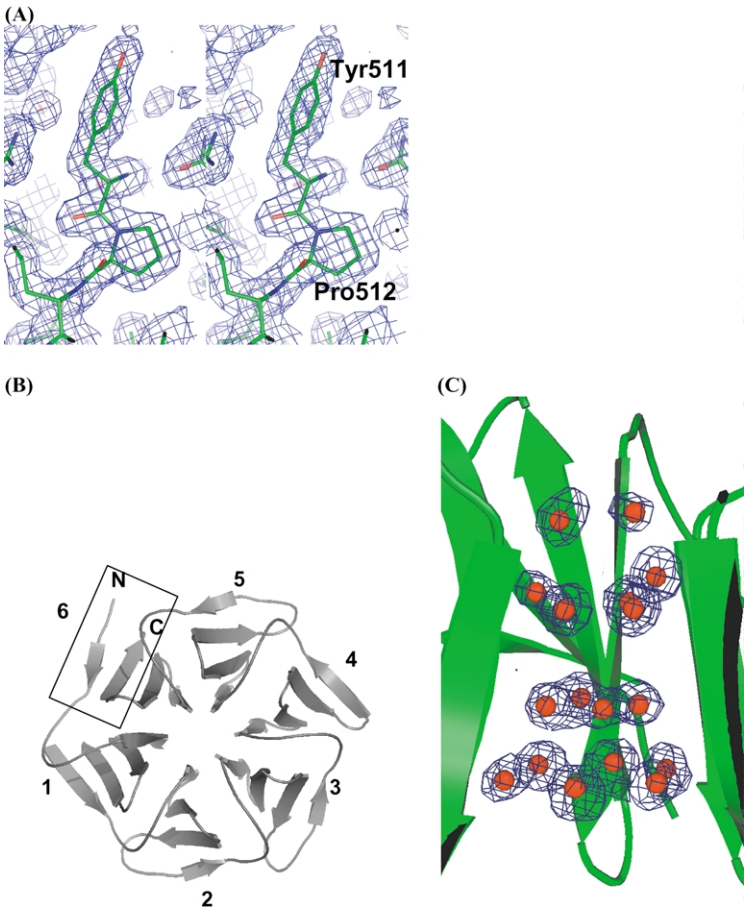

Figure 1. Crystal structure of the extracellular domain of PknD. (A) Stereo view of the electron density map ($2F_o - F_c$) of monoclinic crystal form. Electron density is displayed surrounding one of the invariant prolines in the "cup" and the exposed Tyr511. (B) The PknD sensor domain forms a six-bladed β propeller. Four strands constitute each of the six blades. The numbers around the periphery indicate the blade number, with the full blade nearest the N terminus labeled 1. The "latch" formed by the N and C termini is boxed. The 3–4 loop in blade 1 is disordered. The β propeller forms an approximate cylinder with a radius of 40 Å and a height of 35 Å. (C) Electron density ($2F_o - F_c$) showing rings of ordered water molecules stacked in the central channel of the PknD β propeller. The central pseudo-symmetry axis runs vertically through the center of the rings.

The central pore in the PknD sensor domain is ~8 Å in diameter and is bordered by the first strand of each blade. The generally small side chains in β strand 1 radiate tangentially around the pore. The pore contains stacked rings of water molecules (Figure 1(C)). On the face of the β propeller corresponding to the N-terminal side of strand 1 (the 4–1 loop or the "cup"), Trp597 closes the pore in the orthorhombic form. In the monoclinic form, no density is apparent for the side-chain of Trp597 in either monomer, which leaves the pore open to the solvent. On the face of the β propeller corresponding to the C-terminal end of the first β strands, Asp residues from all six blades point radially around the pore entrance, but no electron density for counterions was apparent.

We calculated the surface electrostatic potential to explore the possible relationship of the sensor domain to the bacterial membrane. The surfaces surrounding both ends of the central pore show a negative potential, and the remainder of the surface is generally neutral (Figure 3). The surface lacks foci of positive potential that might mediate direct interactions with the lipid head groups in the membrane. In the absence of membrane interactions mediated by counterions, the electrostatic

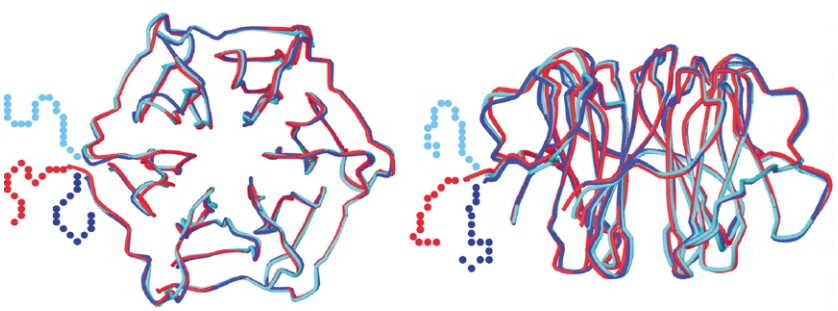

Figure 2. Similarity the independent PknD sensor domain structures. The two monomers from the monoclinic crystal form ($P2_1$, light blue and dark blue) and the model from the orthorhombic form ($P2_12_12_1$, red) were superimposed. The C^α rmsd between the three monomers ranged from 0.35–0.52 Å. The high similarity between the three models and the disorder of the N-terminal 11 residues (dots) implies that the PknD sensor domain forms a rigid framework that is flexibly tethered to the predicted transmembrane helix.

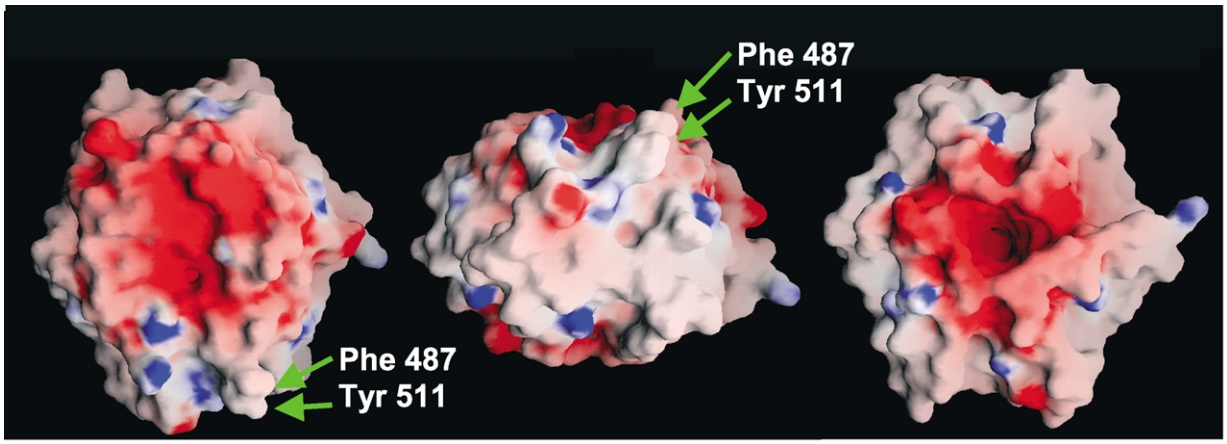

Figure 3. Surface electrostatic potential of the PknD sensor domain. Three views are shown looking along the central axis into the cup (left), rotated ~90° into the plane of the page with the cup on the top (center), and rotated 90° further looking into the central channel (right). Negative surface potential (−4 to −8 *kT*) is shown in red, neutral in white, and positive (+4.7 to +9.5 *kT*) in blue.[63] The distribution of negative surface potential implies the domain is oriented with the central pseudo-symmetry axis parallel to the plane of the membrane, leaving the ends of the β propeller accessible to a ligand. The exposed hydrophobic residues Phe487 and Tyr511 are seen on the rim of the cup.

potential is consistent with a receptor structure in which the sensor domain extends on edge, with the central axis parallel with the lipid bilayer, and exposes the loops connecting the β strands.

Most unanticipated was the high internal symmetry of the PknD sensor domain structure. The backbone rmsd between the six blades ranged from 0.18 Å to 0.67 Å (Figure 4(A)). The core of each blade is structurally conserved. The core sequence repeat, YVTD in β strand 2, forms crucial inter-blade contacts. The tyrosine in this repeat forms a hydrogen bond with the invariant Asp at the C-terminal end of strand 1 of the same blade. The Val forms an inter-blade hydrophobic pocket with four residues in the preceding blade, the Thr of the YVTD repeat, an invariant Val in strand 3, an invariant Phe in strand 4, and a Leu or Ile in strand 4. The Thr (which is replaced by Ala in two blades) is positioned to form a hydrogen bond to the backbone carbonyl of a variable amino acid in the 4–1 loop. Finally, the Asp in four blades contacts the ϵ and η-nitrogen atoms of the Arg four or five residues away. These interactions contribute to the marked internal structural similarity. The largest structural variation among the blades occurs in the exposed 4–1 and 2–3 loops in the cup (Figure 4(A)).

A structure-based alignment of the blade sequences (Figure 4(B)) revealed a level of conservation that is unprecedented in the β-propeller family. At 30 of 42 repeat positions, the same amino acid occurs in at least four of six blades. The pairwise sequence identity among the blades is a remarkable 56.4(±8.1)%, and all six blades are identical at 14 positions. Twelve of these invariant residues define the core of each blade, forming the β strand framework. Also invariant is the Asp that caps the C terminus of strand 1, lining the pore entrance. On the other side of each blade an invariant Pro forms a conserved bulge in β strand 4 (Figures 1(A) and 4(A)). An invariant Pro in the 4–1 loop positions the adjacent variable residues around the edge of the cup.

In contrast to this extensive internal sequence conservation, only three positions differ in at least 5/6 blades. Remarkably, these variable residues cluster in the structure in the 4–1 and 2–3 loops (Figure 4(B) and (C)). These exposed loops show the largest structural variation among the blades. Variable residues form the entire concave surface that encompasses the cup of the β propeller. This surface includes both hydrophilic and hydrophobic residues, with a hydrophobic ridge containing three exposed Phe and Tyr residues extending up to the rim of the cup (Figure 3). Trp597 is located in the bottom of the cup and seals off the central water channel in the structure in the monoclinic crystal form. In both crystal forms, the His_6 tag and associated Cd ions bind deeply in the cup of the neighboring molecule.

The three independent monomers seen in the two crystal forms are highly similar. The backbone rmsd among the three independent monomers varies from 0.35 Å to 0.52 Å (Figure 2). In contrast, the 11 N-terminal residues were disordered in both crystal forms. These results suggest that the PknD sensor domain forms a rigid framework that is loosely tethered to the putative transmembrane helix.

Discussion

As predicted,[15] the sensor domain of PknD forms a β propeller, a widespread motif found in numerous proteins (mostly in eukaryotes) with diverse functions.[20] First reported in the structure of the influenza virus neuraminidase,[21] β propellers have been described containing 4–8 cyclically arranged blades, each with four β strands. In

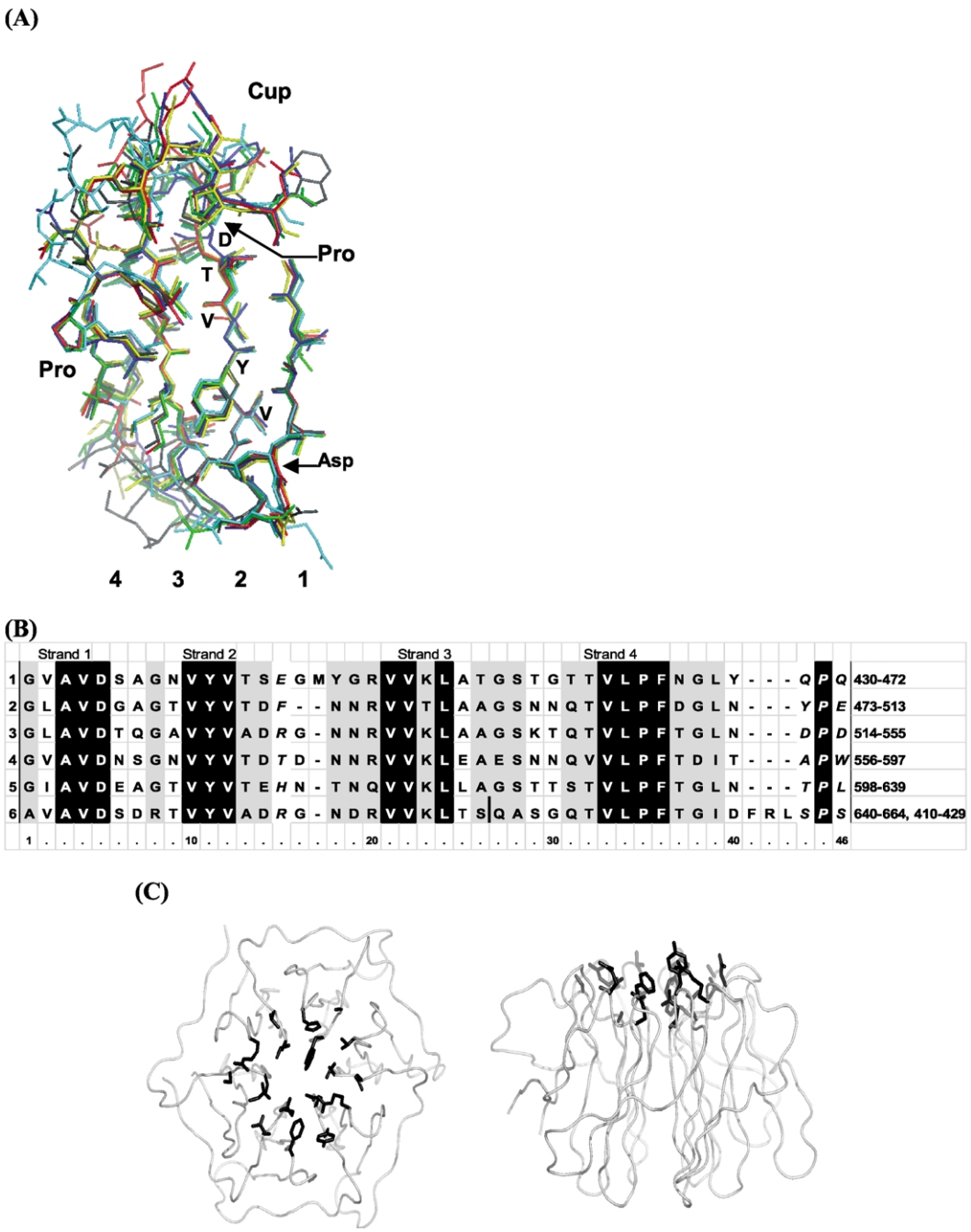

Figure 4. High internal symmetry of the PknD β propeller. (A) Structural superposition of the six blades. The central pseudo-symmetry axis and β strand 1 are vertical on the right side. The cup residues occur on the top right. The invariant proline residues in β strand 4 and the 4–1 loop and the invariant Asp near the C-terminal end of β strand 1 are indicated in the three-letter code. The backbone rmsd between the blades ranged from 0.18 to 0.67 Å. The core β sheet regions show exquisite similarity in all six blades. (B) Structure-based sequence alignment of the six blades of the PknD sensor domain. Blade numbers are listed on the left, and residue numbers are listed on the right. The 14 positions in black are invariant in all six blades, and the 16 positions boxed in grey are invariant in at least four of the six blades. Residues shown in italics are in the "cup" of the structure. With 30 of 42 positions conserved in at least four of six blades, the PknD sensor domain contains the most highly conserved sequence repeat observed in any β-propeller structure solved to date. (C) Sequence variation between the blades clusters in the structure. The side-chains for the three positions per blade (18 total positions) at which the amino acid residues differ in five of the six blades (black sticks) all map to the cup region of the structure. This variation would complement an asymmetric ligand. Residues in the cup are hydrophobic and hydrophilic in character, and include exposed aromatic side-chains. These features suggest that the cup region may play a role in ligand binding or protein–protein interactions.

some β-propeller structures, additional β strands occur on the outside of individual blades, or unrelated globular domains replace one or more blades.[22] With the notable exception of the viral neuraminidases, the vast majority of β propellers are built on an approximate sequence repeat, such as the Trp-Asp repeat embedded in a weak, ~40 amino acid repeat that defines the WD40 motif.[23] The sequence repeat does not correspond to an individual blade. Rather, each β propeller has a topological "latch" in which β strands from both termini of the protein chain are juxtaposed within one or more blades. The latch in PknD, in which the N terminus forms strand 4 of the last blade, corresponds to the latch structures observed in WD40 proteins[24] and the NHL-repeat protein, brain tumor (BraT).[25] A variation of this arrangement occurs in the P2 protein of the bacteriophage PRD-1, in which the N terminus forms the outer strands of the three C-terminal blades.[22] In contrast, the YWTD and the archaeal surface-layer-protein (SLP) β propellers start with strand 2 of the last blade.[15]

β-Propeller repeat motifs are classified on the basis of sequence and structural topology. The Pfam database of protein families[26] and the Conserved Domain Database (CDD)[27] define three classes of β-propeller repeats. Each consensus sequence can be represented by a single family member in the PDB. The G protein β subunit[24] represents the classic WD40 class, the low-density lipoprotein receptor[28] defines the YWTD class, and the NHL domain of BraT protein[25] defines the NHL class. Sequence and structural databases classify β-propeller repeats in a similar fashion. The Structural Classification of Proteins (SCOP) database[29] defines three superfamilies of β-propeller repeats: the WD40-repeat, the low-density lipoprotein receptor-YWTD, and the YVTN-repeat. The CATH Protein Structure Classification database[30] classifies the WD40-repeat and low-density lipoprotein receptor-YWTD repeats as homologous superfamilies.

On the basis of sequence profiles, PknD is a NHL-repeat β propeller with a core sequence motif, YVTD, which differs from any previously determined β-propeller structure. The repeat with highest sequence homology to the NHL consensus matches with 65% identity (E-value $= 5 \times 10^{-11}$). The NHL domain, named after NCL-1, HT2A and Lin-41, is found largely in tandem arrays in ring-finger beta-box coiled-coil (RBCC) regulatory proteins.[16] The NHL protein family has been proposed to mediate diverse protein interactions largely in eukaryotes.[16] The domain was defined structurally by the six-bladed β propeller in the translational repressor BraT.[25] The NHL repeats in BraT show at most 50% identity with the consensus NHL sequence. However, BraT contains an IVAD core sequence repeat that does not match the YVTD repeat in PknD. The overall sequence identity between BraT and the PknD sensor is 23% over 208 residues (E-value $= 9 \times 10^{-5}$). The C^{α} rmsd between the PknD and BraT β propellers is 2.6 Å, a value which is not better than matches to β propellers with other sequence repeats (see below). As the second structure of a NHL-repeat protein (and the first from a bacterial source), the PknD sensor domain reveals substantial structural variation within the NHL class.

Interestingly, the PknD sensor domain shows significant sequence homology to nidogen,[31] a YWTD β propeller (28% identity over 159 residues, E-value $= 5 \times 10^{-6}$), and to the β propeller of archaeal SLP[15] (30% identity over 189 residues, E-value $= 9 \times 10^{-5}$)). Nidogen and SLP, however, lack the NHL-repeat signature and show no significant homology to BraT. Nidogen and SLP also contain characteristic core contacts and a latch topology that differ from PknD. Thus, the sequence of PknD can be considered intermediate between the NHL proteins on the one hand and the YWTD proteins and YVTN repeat of SLP[15] on the other. Overall, the PknD sensor domain structure demonstrates a new core sequence repeat capable of forming a NHL-domain β propeller.

The network of contacts in the YVTD blade is unique to this repeat. In the YWTD repeat, for example, the Asp forms hydrogen bonds with the peptide backbone[32] instead of contacting a conserved Arg. Unlike the YVTD repeat of PknD, the SLP YVTN Tyr has no apparent hydrogen bond partner, and the Asn contacts a Ser side-chain and a backbone amide.[15] The distinctive latch structures and the absence of β propellers containing mixtures of different sequence repeats argues that the β-propeller family arose by serial duplications of the basic blade sequence followed by divergence and/or deletions of the repeats.

The pairwise sequence identity among the PknD blades (56.4(±8.1)%) is higher than in any other β-propeller structure. Compared to SLP, in which five of 42 positions are conserved in all seven blades,[15] 14 of 42 residues are identical in all six blades of PknD. In tachylectin-2, a five-bladed β propeller from the Japanese horseshoe crab *Tachypleus tridentatus*, the pairwise sequence identity between blades is 53.5(±5.5)%.[33] This symmetry in tachylectin-2 defines the core and the peripheral binding sites for the ligands *N*-acetylglucosamine and *N*-acetylgalactosamine.[33] In contrast, the pairwise sequence identity between the seven blades of the Gβ β propeller is only 20.7(±5.8)%.[24] Thus, the high sequence identity among the six PknD blades raises the possibility that PknD may bind six homologous ligands (analogous to tachylectin-2). Alternatively, PknD may bind a single asymmetric ligand through residues that differ in each blade. In this case, the conserved blade framework may indicate the protein arose recently in evolution. The idea that PknD is a "new" protein is consistent with the observation that orthologs are narrowly distributed, occurring only in pathogenic mycobacteria.

Remarkably, the 18 PknD residues that differ in at least five of the six PknD blades cluster in the

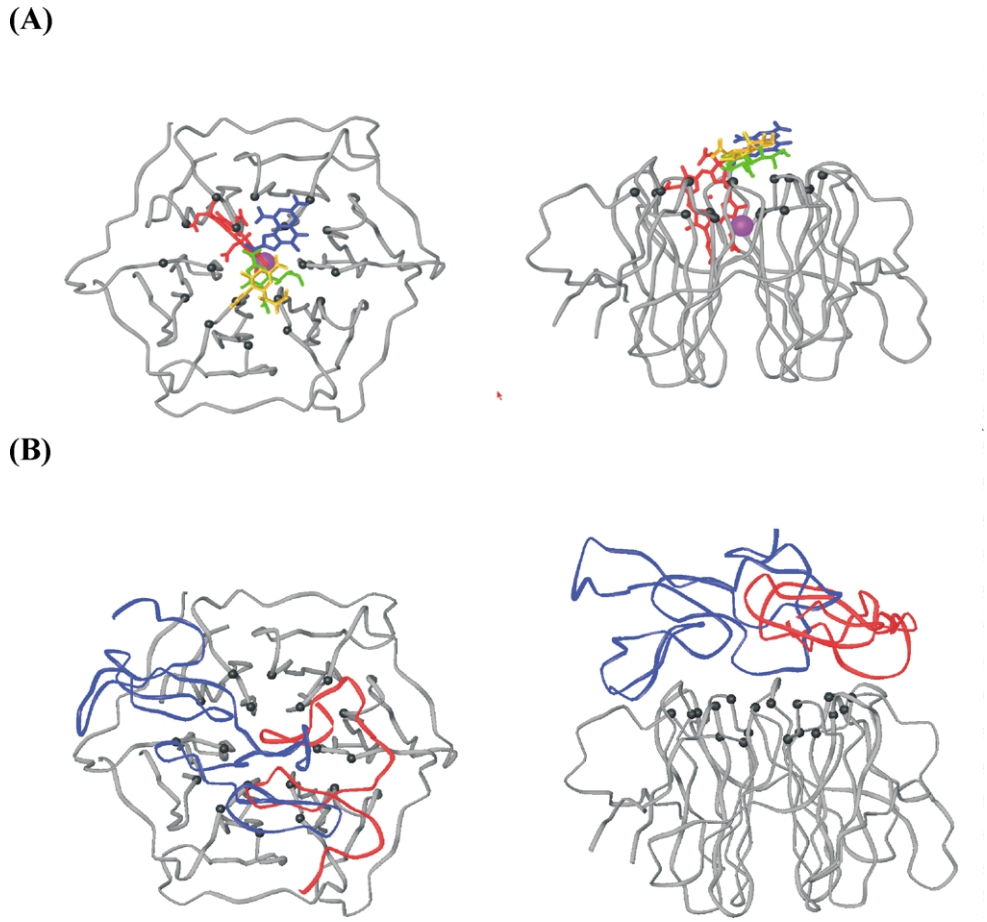

Figure 5. β-Propeller proteins bind ligands in the cup. (A) The backbones and small-molecule ligands of five β-propeller structures similar to PknD were superimposed with DALI[36] on the PknD structure. Ligands that bind in cup of the cognate structures include heme from the cytochrome Cd1 nitrite reductase[64] (red), Ca^{2+} in the active site of diisopropylfluorophosphatase[37] (purple), *N*-acetyl-D-glucosamine from neuraminidase[21] (green), 2-deoxy-2,3-dehydro-*N*-acetyl-neuramic acid from *trans*-sialidase[44] (orange), and the pyrroloquinoline quinone from soluble quinoprotein glucose dehydrogenase[39] (blue). Tachylectin-2 which binds sugars at conserved sites on the periphery of each blade,[33] provides an exception to this paradigm. Variable residues in the PknD sensor domain cup are shown as black dots. (B) Protein ligands of β propellers bind to surface around the rim of the cup. The PknD sensor domain was aligned with the structure of the low density lipoprotein receptor[28] and the nidogen–laminin complex.[31] The portion of the EGF domain of LDLR that interacts with the β-propeller domain is shown in red, and a portion of integrin that binds the nidogen β propeller is shown in blue. Both β-propeller proteins make protein–protein interactions in the cup. In contrast, clathrin terminal domain binds target peptides on the edge of the β propeller (not shown).[65]

cup of the β propeller (Figure 4). In β-propeller enzymes and protein binding modules, the cup often encompasses the active site or protein recognition surface (Figure 5). In the PknD crystals, the Cd^{2+}-ligated, His_6 tag bound in the cup of a neighboring molecule in both crystal forms. Like many protein binding sites,[34] the PknD cup exposes a large surface (931 $Å^2$) containing both polar and non-polar patches. The hydrophobic residues exposed in the cup include Phe487 and Tyr511. Phe is the most commonly buried residue in proteins,[35] suggesting that the exposed Phe487 may function in ligand binding. These characteristics support the idea that the cup is a functional binding surface of the PknD sensor domain.

None of the signaling ligands for bacterial receptor STPKs have been identified. The *PknD* gene in *Mtb* occurs in a cluster of genes involved in phosphate sensing and transport. In addition, *PknD* is co-expressed with *PstS2*, which encodes a periplasmic phosphate binding protein. Based on this genomic position and expression pattern, Av-Gay hypothesized that PknD may signal the levels of extracellular phosphate.[4] To test this idea, we searched the available structures using DALI[36] to identify proteins with backbone structures resembling the PknD sensor. No extremely similar structures were found. Diisopropylfluorophosphatase[37] and phytase[38] gave the closest global matches, with rms deviations of corresponding backbone atoms of 2.6 Å. Both of these enzymes metabolize organophosphates. Other proteins with similar structures included protein interaction motifs, such as nidogen[31] (2.1 Å C^{α} rmsd) and the pH sensor of the low density lipoprotein receptor[32] (2.4 Å C^{α} rmsd). Comparable similarity was found to several

enzymes with diverse functions, such as soluble quinoprotein glucose dehydrogenase[39] (2.7 Å C^{α} rmsd), sialidase[40] (2.9 Å C^{α} rmsd), and actin-related protein 3[41] (3.0 Å C^{α} rmsd). Because the β-propeller proteins mediate so many different activities (including numerous enzyme reactions, carbohydrate recognition, metal binding, and protein–protein interactions) the fold of the PknD sensor domain does not define a particular functional role.

How does the PknD sensor domain switch the activity of the intracellular kinase domain? Although β-propeller domains have been predicted in the *Drosophila* receptor tyrosine kinase *sevenless* and its vertebrate homolog *c-ros*,[23,42] no other structures of β-propeller domains that are linked to protein kinases are available for comparison with PknD. Generally, large conformational changes have not been observed in β-propeller proteins. The β subunits of heterotrimeric G-proteins, for example, show very limited conformational shifts upon binding of the α-subunits.[24] Similarly, the pH-sensor domain of the LDL receptor[28,32] and the β-propeller domain in the integrin αVβ3[43] change little upon associating with protein ligands. We found that backbone and side-chain positions are extraordinarily consistent between the three independent PknD sensor domain monomers (Figure 2). These results provide important support for the idea that the PknD sensor, like other β propellers, comprises a rigid structural framework that does not undergo a conformational change upon ligand binding. Moreover, although the PknD sensor domain construct begins at the first residue following the predicted transmembrane helix, the first 11 residues of the sensor domain are disordered in all three independent monomers. This disorder suggests that the sensor domain is tethered flexibly to the transmembrane domain. A disordered segment also links the transmembrane domain to the PknB kinase domain.[12] If this flexible coupling is a general property of these receptor STPKs, binding of the signaling ligands is unlikely to mediate conformational changes that propagate across the membrane. In the absence of propagating changes in tertiary structure, the PknD sensor domain likely mediates signaling through a direct or indirect change in quaternary structure or localization that alters the activity of the attached, intracellular kinase domain.

Generally, β propellers mediate heterotypic binding. The few homodimers include the hemagglutinin-neuraminidases of the paramixoviruses.[44,45] Although the β-propeller protein SEMA4D forms homodimers,[46] the β propeller does not mediate this pairing. Accordingly, the PknD sensor domain is monomeric under the solution conditions we tested (data not shown). Consequently, the fold and the high symmetry of the PknD β propeller imply that plausible candidate ligands include an oligomeric protein that recognizes the cup or receptors that recognize the repeated peripheral sites on the blades.

Materials and Methods

Cloning

An expression vector for PknD residues 403–664 followed by the C-terminal His_6 tag, LEHHHHHH, was created with pET24b (Novagen) using sticky-end PCR[47] as described.[12] Mutagenesis was carried out with the QuikChange kit (Stratagene). Constructs and mutants were confirmed by DNA sequencing.

Expression and purification

The wild-type and mutant PknD sensor domains were expressed at 25 °C in BL21(DE3) Codon Plus cells (Stratagene) as described.[12] The cells were harvested by centrifugation and resuspended in 50 mM Tris–HCl (pH 7.5), 0.3 M NaCl, 20% (v/v) glycerol, and 0.5 mM Tri(2-carboxyethyl)phosphine hydrochloride (TCEP). Cells were lysed by sonication on ice, and the lysate was clarified by centrifugation. The supernatant was loaded onto a 5 ml, immobilized metal affinity chromatography (IMAC) column (Sigma) equilibrated with 0.1 M $NiSO_4$ and eluted with a 10 mM–250 mM imidazole gradient in 100 ml. After dialysis into 50 mM Tris–HCl (pH 7.5), 150 mM NaCl, 5% glycerol, and 0.5 mM TCEP, the protein was further purified by size-exclusion chromatography using a HiLoad 26/60 Superdex 75 column (Amersham Pharmacia) equilibrated with 50 mM Tris–HCl (pH 7.5), 150 mM NaCl, 5% glycerol, and 0.5 mM TCEP. The single peak, corresponding to the PknD sensor domain monomer, was dialyzed into 20 mM Tris–HCl (pH 7.5), 75 mM NaCl, and 0.5 mM TCEP and loaded onto two tandem 5 ml HiTrap Q Sepharose ion-exchange columns (Amersham Pharmacia) in 20 mM Tris–HCl (pH 7.5), 75 mM NaCl, and 0.5 mM TCEP and eluted with a 0.075 M–1 M NaCl gradient in 100 ml. The purified protein was dialyzed into 25 mM Tris (pH 7.5) 75 mM NaCl, 0.5 mM reduced dithiothreitol (DTT).

Crystallization and heavy atom derivatives

The PknD sensor domain (20 mg ml^{-1}) was crystallized at 22 °C by vapor diffusion in hanging drops against 0.8 M sodium acetate (NaOAc), 0.1 M Hepes (pH 8.0) 0.05 M $CdSO_4$. These conditions yielded both crystal forms, although the monoclinic form was much more common.

The protein contains only a single internal Met residue, so we searched for heavy atom derivatives. Numerous metal compounds failed to provide adequate phasing signals. We created a suitable derivative by introducing sites for metal labeling using two parallel strategies. First, we introduced single Cys residues. Guided by the effects of Hg binding to Cys residues introduced in T4 lysozyme,[19] we sought to avoid buried sites where metal binding would disrupt the crystals and exposed sites that would be too mobile to restrict the position of the metal atom. Second, with the aim of obtaining phases from a protein labeled with SeMet, we introduced two Leu → Met substitutions at predicted buried sites.[48,49] The sites of substitutions were chosen using a homology model of the PknD sensor domain produced by the program 3D-PSSM.[18] The model was based on the β-propeller domain of the pH sensor of the low density lipoprotein receptor (LDLR). Despite only 17% sequence identity, the PknD sequence produced a

structural match with an *E*-value of 0.036. Thr455, Thr549, Ser570, Thr606 and Ser637 that were partially exposed in the homology model were changed individually to Cys. Leu578 and Leu620 buried in individual blades in the model were changed in tandem to Met.

The protein variants were purified as described for the wild-type sensor domain. The double leucine mutant was labeled by biosynthetic incorporation of SeMet.[50] Cysteine mutants were co-crystallized with 1 mM *p*-chloromercuribenzene sulfonic acid (Sigma). Back soaking was carried out in cryoprotectant.

Structure determination

X-ray data were collected at the Lawrence Berkeley National Laboratory Advanced Light Source Beamline 8.3.1. For data collection, crystals were soaked in 1 M NaOAc, 0.1 M Hepes (pH 8.5) 0.05 M $CdSO_4$ and 25% glycerol, mounted on a loop, and frozen in liquid N_2. Native data from the monoclinic crystal form were processed to 1.8 Å resolution using the Elves automation program.[51] Elves also was used to analyze Hg-derivative data collected at two wavelengths to 3 Å resolution. Although the derivative crystals diffracted to higher resolution, the need to collect short exposures to limit radiation damage reduced the accessible resolution of the data sets.

Running SOLVE[52] and MLPHARE[53] Elves located and refined two Hg sites, automatically explored the hand of the derivative, systematically ruled out alternative space groups and optimized the amount of solvent to use in solvent flattening.[51] The best electron density map at 3 Å resolution was interpreted manually using O,[54] using features in the homology model as a guide. Elements of the partial model common to both monomers in the asymmetric unit were used to locate and refine the non-crystallographic symmetry (NCS) operator using DM.[55] The partial model was refined against the 1.8 Å resolution data from the wild-type protein using CNS.[56] The refined model and the high-resolution native data were used to initiate automated model building of the wild-type structure with ARP/wARP.[57] This procedure produced a model containing ~90% of the residues with R/R_{free} values of 17.3/29.7%. The high-resolution electron density map was used to adjust the register of several of the β strands, and the model was completed by alternating cycles of building using O and refinement using Refmac.[58] Tight NCS restraints were used in the early stages of refinement. R_{free} was calculated with a random 5% of the data. The final model contains residues 413–454 and 459–664 in both monomers (plus the entire His_6 tag), 14 Cd^{2+} and 209 water molecules. No density was apparent in either monomer for the first 11 residues of the protein construct. The stereochemistry of the model analyzed using PROCHECK[59] scored better overall (*G*-factor = 0.0) than most structures refined at 1.8 Å resolution. All residues have allowed main-chain dihedral angles.

X-ray data from the orthorhombic crystal form were processed to 1.9 Å resolution using Elves.[51] Automated molecular replacement was performed using EPMR.[60] The search model was a partially refined structure of one monomer (lacking the His_6 tag) from the monoclinic crystals. The model of the orthorhombic form was refined at 1.9 Å resolution with CNS[56] and rebuilding was carried out with O. The final model contains one PknD sensor domain (residues 413–454, 459–664 and the first four residues of the His_6 tag), 71 water molecules and four Cd ions in the asymmetric unit. The last two residues of the His_6 tag were disordered. The stereochemistry of the model evaluated using PROCHECK scored better overall (*G*-factor = 0.0) than most structures refined at 1.9 Å resolution. No residues were in the disallowed region of the Ramachandran plot.

Structure analysis

DALI[36] was used to identify structures related to the PknD sensor domain. Sequence alignments were carried out with CLUSTAL X.[61] Overlays were drawn with Ribbons,[62] surface features were depicted with GRASP,[63] and other Figures were drawn with PyMOL.†

Coordinates

The coordinates and the structure factors were deposited in the Protein Data Bank under accession numbers 1RWI (monoclinic form) and 1RWL (orthorhombic form).

Acknowledgements

We thank David S. King for protein mass spectrometry measurements, and James Holton, Emmanuel Skordalakes, and Benedicte Delagoutte for help with X-ray structure determinations. We thank Kristi Pullen, Laurie Gay, and Danielle Tomkiel for stimulating discussions. We are grateful to Tom Terwilliger and the TB Structural Genomics Consortium for support. This work was supported by a grant from the NIH. The Advanced Light Source Beamline 8.3.1 was funded by the NSF, the University of California and Henry Wheeler.

References

1. Russell, D. G. (2001). *Mycobacterium tuberculosis*: here today, and here tomorrow. *Nature Rev. Mol. Cell. Biol.* **2**, 569–577.
2. Reed, S. G., Alderson, M. R., Dalemans, W., Lobet, Y. & Skeiky, Y. A. (2003). Prospects for a better vaccine against tuberculosis. *Tuberculosis (Edinb.)*, **83**, 213–219.
3. Cole, S. T. *et al.* (1998). Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature*, **393**, 537–544.
4. Av-Gay, Y. & Everett, M. (2000). The eukaryotic-like Ser/Thr protein kinases of *Mycobacterium tuberculosis*. *Trends Microbiol.* **8**, 238–244.
5. Av-Gay, Y., Jamil, S. & Drews, S. J. (1999). Expression and characterization of the *Mycobacterium tuberculosis* serine/threonine protein kinase PknB. *Infect. Immun.* **67**, 5676–5682.
6. Chaba, R., Raje, M. & Chakraborti, P. K. (2002). Evidence that a eukaryotic-type serine/threonine protein kinase from *Mycobacterium tuberculosis*

† DeLano, W. L. (2002). DeLano Scientific, San Fransisco, CA.

regulates morphological changes associated with cell division. *Eur. J. Biochem.* **269**, 1078–1085.
7. Koul, A. *et al.* (2001). Serine/threonine protein kinases PknF and PknG of *Mycobacterium tuberculosis*: characterization and localization. *Microbiology,* **147**, 2307–2314.
8. Molle, V. *et al.* (2003). Protein PknE, a novel transmembrane eukaryotic-like serine/threonine kinase from *Mycobacterium tuberculosis*. *Biochem. Biophys. Res. Commun.* **308**, 820–825.
9. Peirs, P., De Wit, L., Braibant, M., Huygen, K. & Content, J. (1997). A serine/threonine protein kinase from *Mycobacterium tuberculosis*. *Eur. J. Biochem.* **244**, 604–612.
10. Madec, E., Laszkiewicz, A., Iwanicki, A., Obuchowski, M. & Seror, S. (2002). Characterization of a membrane-linked Ser/Thr protein kinase in *Bacillus subtilis*, implicated in developmental processes. *Mol. Microbiol.* **46**, 571–586.
11. Boitel, B. *et al.* (2003). PknB kinase activity is regulated by phosphorylation in two Thr residues and dephosphorylation by PstP, the cognate phospho-Ser/Thr phosphatase, in *Mycobacterium tuberculosis*. *Mol. Microbiol.* **49**, 1493–1508.
12. Young, T. A., Delagoutte, B., Endrizzi, J. A., Falick, A. M. & Alber, T. (2003). Structure of *Mycobacterium tuberculosis* PknB supports a universal activation mechanism for Ser/Thr protein kinases. *Nature Struct. Biol.* **10**, 168–174.
13. Ortiz-Lombardia, M., Pompeo, F., Boitel, B. & Alzari, P. M. (2003). Crystal structure of the catalytic domain of the PknB serine/threonine kinase from *Mycobacterium tuberculosis*. *J. Biol. Chem.* **278**, 13094–13100.
14. Yeats, C., Finn, R. D. & Bateman, A. (2002). The PASTA domain: a β-lactam-binding domain. *Trends Biochem. Sci.* **27**, 438.
15. Jing, H. *et al.* (2002). Archaeal surface layer proteins contain β propeller, PKD, and β helix domains and are related to metazoan cell surface proteins. *Structure*, 1453–1464.
16. Slack, F. J. & Ruvkun, G. (1998). A novel repeat domain that is often associated with RING finger and B-box motifs. *Trends Biochem. Sci.* **23**, 474–475.
17. Rajagopal, L., Clancy, A. & Rubens, C. E. (2003). A eukaryotic type serine/threonine kinase and phosphatase in *Streptococcus agalactiae* reversibly phosphorylate an inorganic pyrophosphatase and affect growth, cell segregation, and virulence. *J. Biol. Chem.* **278**, 14429–14441.
18. Kelley, L. A., MacCallum, R. M. & Sternberg, M. J. (2000). Enhanced genome annotation using structural profiles in the program 3D-PSSM. *J. Mol. Biol.* **299**, 499–520.
19. Sun, D. P., Alber, T., Bell, J. A., Weaver, L. H. & Matthews, B. W. (1987). Use of site-directed mutagenesis to obtain isomorphous heavy-atom derivatives for protein crystallography: cysteine-containing mutants of phage T4 lysozyme. *Protein Eng.* **1**, 115–123.
20. Jawad, Z. & Paoli, M. (2002). Novel sequences propel familiar folds. *Structure*, **10**, 447–454.
21. Varghese, J. N., Laver, W. G. & Colman, P. M. (1983). Structure of the influenza virus glycoprotein antigen neuraminidase at 2.9 Å resolution. *Nature*, **303**, 35–40.
22. Xu, L., Benson, S. D., Butcher, S. J., Bamford, D. H. & Burnett, R. M. (2003). The receptor binding protein P2 of PRD1, a virus targeting antibiotic-resistant bacteria, has a novel fold suggesting multiple functions. *Structure*, **11**, 309–322.
23. Springer, T. A. (1998). An extracellular β-propeller module predicted in lipoprotein and scavenger receptors, tyrosine kinases, epidermal growth factor precursor, and extracellular matrix components. *J. Mol. Biol.* **283**, 837–862.
24. Sondek, J., Bohm, A., Lambright, D. G., Hamm, H. E. & Sigler, P. B. (1996). Crystal structure of a G-protein βγ dimer at 2.1 Å resolution. *Nature*, **379**, 369–374.
25. Edwards, T. A., Wilkinson, B. D., Wharton, R. P. & Aggarwal, A. K. (2003). Model of the brain tumor-Pumilio translation repressor complex. *Genes Dev.* **17**, 2508–2513.
26. Bateman, A. *et al.* (2004). The Pfam protein families database. *Nucl. Acids Res.* **32**, D138–D141.
27. Marchler-Bauer, A. *et al.* (2003). CDD: a curated Entrez database of conserved domain alignments. *Nucl. Acids Res.* **31**, 383–387.
28. Rudenko, G. *et al.* (2002). Structure of the LDL receptor extracellular domain at endosomal pH. *Science*, **298**, 2353–2358.
29. Andreeva, A. *et al.* (2004). SCOP database in 2004: refinements integrate structure and sequence family data. *Nucl. Acids Res.* **32**, D226–D229.
30. Pearl, F. M. *et al.* (2000). Assigning genomic sequences to CATH. *Nucl. Acids Res.* **28**, 277–282.
31. Takagi, J., Yang, Y., Liu, J. H., Wang, J. H. & Springer, T. A. (2003). Complex between nidogen and laminin fragments reveals a paradigmatic β-propeller interface. *Nature*, **424**, 969–974.
32. Jeon, H. *et al.* (2001). Implications for familial hypercholesterolemia from the structure of the LDL receptor YWTD-EGF domain pair. *Nature Struct. Biol.* **8**, 499–504.
33. Beisel, H. G., Kawabata, S., Iwanaga, S., Huber, R. & Bode, W. (1999). Tachylectin-2: crystal structure of a specific GlcNAc/GalNAc-binding lectin involved in the innate immunity host defense of the Japanese horseshoe crab *Tachypleus tridentatus*. *EMBO J.* **18**, 2313–2322.
34. Lo Conte, L., Chothia, C. & Janin, J. (1999). The atomic structure of protein–protein recognition sites. *J. Mol. Biol.* **285**, 2177–2198.
35. Rose, G. D., Geselowitz, A. R., Lesser, G. J., Lee, R. H. & Zehfus, M. H. (1985). Hydrophobicity of amino acid residues in globular proteins. *Science*, **30**, 834–838.
36. Holm, L. & Sander, C. (1993). Protein structure comparison by alignment of distance matrices. *J. Mol. Biol.* **233**, 123–138.
37. Scharff, E. I., Koepke, J., Fritzsch, G., Lucke, C. & Ruterjans, H. (2001). Crystal structure of diisopropylfluorophosphatase from *Loligo vulgaris*. *Structure*, **9**, 493–502.
38. Ha, N. C. *et al.* (2000). Crystal structures of a novel, thermostable phytase in partially and fully calcium-loaded states. *Nature Struct. Biol.* **7**, 147–153.
39. Oubrie, A. *et al.* (1999). Structure and mechanism of soluble quinoprotein glucose dehydrogenase. *EMBO J.* **18**, 5187–5194.
40. Gaskell, A., Crennell, S. & Taylor, G. (1995). The three domains of a bacterial sialidase: a β-propeller, an immunoglobulin module and a galactose-binding jelly-roll. *Structure*, **3**, 1197–1205.
41. Robinson, R. C. *et al.* (2001). Crystal structure of Arp2/3 complex. *Science*, **294**, 1679–1684.
42. Norton, P. A., Hynes, R. O. & Rees, D. J. (1990). *sevenless*: Seven found? *Cell*, **61**, 15–16.

43. Xiong, J. P. *et al.* (2002). Crystal structure of the extracellular segment of integrin αVβ3 in complex with an Arg-Gly-Asp ligand. *Science*, **296**, 151–155.
44. Crennell, S. J. *et al.* (1996). The structures of *Salmonella typhimurium* LT2 neuraminidase and its complexes with three inhibitors at high resolution. *J. Mol. Biol.* **259**, 264–280.
45. Lawrence, M. C. *et al.* (2004). Structure of the haemagglutinin-neuraminidase from human parainfluenza virus type III. *J. Mol. Biol.* **335**, 1343–1357.
46. Love, C. A. *et al.* (2003). The ligand-binding face of the semaphorins revealed by the high-resolution crystal structure of SEMA4D. *Nature Struct. Biol.* **10**, 843–848.
47. Zeng, G. (1998). Sticky-end PCR: new method for subcloning. *Biotechniques*, **25**, 206–208.
48. Rose, R. B., Endrizzi, J. A., Cronk, J. D., Holton, J. & Alber, T. (2000). High-resolution structure of the HNF-1α dimerization domain. *Biochemistry*, **39**, 15062–15070.
49. Gassner, N. C. & Matthews, B. W. (1999). Use of differentially substituted selenomethionine proteins in X-ray structure determination. *Acta Crystallog. sect. D*, **55**, 1967–1970.
50. Van Duyne, G. D., Standaert, R. F., Karplus, P. A., Schreiber, S. L. & Clardy, J. (1993). Atomic structures of the human immunophilin FKBP-12 complexes with FK506 and rapamycin. *J. Mol. Biol.* **229**, 105–124.
51. Holton, J. M. & Alber, T. (2004). Automated protein crystal structure determination using Elves. *Proc. Natl Acad. Sci. USA*, **101**, 1537–1542.
52. Terwilliger, T. C. & Berendzen, J. (1999). Automated MAD and MIR structure solution. *Acta Crystallog. sect. D*, **55**, 849–861.
53. CCP4 (1994). The CCP4 suite: programs for protein crystallography. *Acta Crystallog. sect. D*, 760–763.
54. Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. (1991). Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallog.* **47**, 110–119.
55. Cowtan, K. (1994). 'dm': an automated procedure for phase improvement by density modification. *Joint CCP4 ESF-EACBM Newsletter Protein Crystallog.* **31**, 34–38.
56. Brunger, A. T. *et al.* (1998). Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallog. sect. D*, **54**, 905–921.
57. Morris, R. J., Perrakis, A. & Lamzin, V. S. (2002). ARP/wARP's model-building algorithms. I. The main chain. *Acta Crystallog. sect. D*, **58**, 968–975.
58. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallog. sect. D*, **53**, 240–255.
59. Laskowski, R. A., MacArthur, 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.
60. Kissinger, C. R., Gehlhaar, D. K. & Fogel, D. B. (1999). Rapid automated molecular replacement by evolutionary search. *Acta Crystallog. sect. D*, **55**, 484–491.
61. Jeanmougin, F., Thompson, J. D., Gouy, M., Higgins, D. G. & Gibson, T. J. (1998). Multiple sequence alignment with Clustal X. *Trends Biochem. Sci.* **23**, 403–405.
62. Carson, M. (1991). Ribbons. *J. Appl. Crystallog.* **24**, 958–961.
63. Nicholls, A., Sharp, K. A. & Honig, B. (1991). Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins: Struct. Funct. Genet.*, **11**, 281–296.
64. Fulop, V., Moir, J. W., Ferguson, S. J. & Hajdu, J. (1995). The anatomy of a bifunctional enzyme: structural basis for reduction of oxygen to water and synthesis of nitric oxide by cytochrome cd1. *Cell*, **81**, 369–377.
65. ter Haar, E., Harrison, S. C. & Kirchhausen, T. (2000). Peptide-in-groove interactions link target proteins to the β-propeller of clathrin. *Proc. Natl Acad. Sci. USA*, **97**, 1096–1100.

Edited by I. Wilson

(Received 30 December 2003; received in revised form 16 March 2004; accepted 22 March 2004)