Structure, Vol. 10, 1453–1464, October, 2002, ©2002 Elsevier Science Ltd. All rights reserved. PII S0969-2126(02)00840-7

Archaeal Surface Layer Proteins Contain β Propeller, PKD, and β Helix Domains and Are Related to Metazoan Cell Surface Proteins

Hua Jing,[1] Junichi Takagi,[1] Jin-huan Liu,[2]
Sara Lindgren,[2] Rong-guang Zhang,[4]
A. Joachimiak,[4] Jia-huai Wang,[3,5]
and Timothy A. Springer[1,5]
[1]Center for Blood Research and
Department of Pathology
[2]Dana-Farber Cancer Institute and
Department of Medicine
[3]Dana-Farber Cancer Institute and
Department of Pediatrics and
Department of Biological Chemistry
and Molecular Pharmacology
Harvard Medical School
Boston, Massachusetts 02115
[4]Biosciences Division
Structural Biology Center
Argonne National Laboratory
Chicago, Illinois

Summary

The surface layer of archaeobacteria protects cells from extreme environments and, in *Methanosarcina*, may regulate cell adhesion. We identify three domain types that account for the complete architecture of numerous *Methanosarcina* surface layer proteins (SLPs). We solve the crystal structure for two of these domains, which correspond to the two N-terminal domains of an *M. mazei* SLP. One domain displays a unique, highly symmetrical, seven-bladed β propeller fold, and the other belongs to the polycystic kidney disease (PKD) superfamily fold. The third domain is predicted to adopt a β helix fold. These domains have homologs in metazoan cell surface proteins, suggesting remarkable relationships between domains in archaeal SLPs and metazoan cell surface proteins.

Introduction

Almost all studied archaeobacteria and many eubacteria contain a surface layer (S layer) external to the cytoplasmic membrane. The S layer is composed exclusively of proteins or glycoproteins organized regularly into a porous and often crystalline surface [1–5]. In many *Archaea*, the S layer is the only cell envelope component and thus provides all the functions normally associated with a cell wall, such as a protective barrier and maintenance of cell shape. In the methanogen *Methanosarcina mazei*, the surface layer proteins (SLPs) may also help regulate cell-cell association [6]. The Latin root "sarcina," meaning bundle, refers to the habit of *Methanosarcina* of growing in well-defined multicellular structures. Growth as single cells, packets, or lamina is regulated by *Methanosarcina* in response to environmental conditions [7]. Expression of one of the SLPs, which is defined structurally in this paper, is increased in lamina, as shown by immunofluorescence [8], and is present at cell-cell interfaces, as shown by immunogold staining (A. Marcario, personal communication).

Despite the physiological importance of SLPs, little is known about their 3D structure. Only two small portions of SLPs have been characterized by structure prediction or determination. One is a peptidoglycan-anchoring S layer homology (SLH) repeat in some eubacterial SLPs, which is predicted to contain two α helices [9–11]. The other is a 52-residue peptide fragment from the *Staphylothermus marinus* tetrabrachion SLP, which tetramerizes into a four-stranded right-handed coiled coil, as shown by a crystal structure [12]. These short terminal fragments provide tantalizing glimpses of SLPs, but little insight into their overall architecture.

To gain insights into SLP structure, as well as into relationships between surface proteins in *Archaea* and *Metazoa*, we have examined SLPs in *M. mazei*. Two types of sequence repeats have previously been recognized in *M. mazei* SLPs, termed AB and C repeats [6, 13, 14]. We report here that the AB repeats, which contain a YVTN (Tyr-Val-Thr-Asn) motif, are remote sequence homologs of sequence repeats that contain a YWTD (Tyr-Trp-Thr-Asp) motif, are abundant in metazoan extracellular proteins [15], and fold into a six-bladed β propeller [16, 17]. The crystal structure of the *M. mazei* SLP fragment presented here demonstrates a seven-bladed β propeller domain. This SLP fragment also contains a type C repeat, which we demonstrate is closely related to the polycystic kidney disease (PKD) domain found in polycystin-1 and defined by an NMR structure [18]. Mutations in polycystin-1 cause most cases of autosomal dominant polycystic kidney disease, which affects 1 out of 800 people of all ethnic groups. We also find that PKD domains are present in many SLPs in combination with a third type of sequence repeat, which is present in SLPs from many *Archaea*. Detailed analysis of this repetitive segment suggests that it folds into a parallel right-handed β helix domain. In *M. mazei*, as well as in the completely sequenced *Methanosarcina acetivorans* [19], the YVTN β propeller, PKD, and parallel right-handed β helix domains account for the entirety of numerous SLPs. Our results for the first time define the complete architectures of SLPs and reveal evolutionary relationships between SLPs and metazoan cell surface proteins. Furthermore, we report the first crystal structures for a YVTN β propeller domain and a PKD domain.

Results and Discussion

Prediction

A PSI-BLAST search [20] with the consensus sequence from predicted metazoan YWTD β propellers [16] revealed homology to an *M. mazei* SLP (ORF492 of

---

[5]Correspondence: springeroffice@cbr.med.harvard.edu (T.A.S.), jwang@red.dfci.harvard.edu (J.-h.W.)

Key words: surface layer protein; β propeller; PKD; β helix; domain architecture; *Methanosarcina*

gi1213020) [14] with an expectation value of 0.00027. Six tandem YWTD repeats had previously been predicted to fold into a six-bladed β propeller domain [16], as later confirmed by the crystal structure of a representative YWTD domain in the low density lipoprotein receptor (LDLR) [17]. By contrast, we found seven tandem repeats containing a YVTN motif in place of the YWTD motif in ORF492. Using previously described methods [16], we predicted them to fold into a seven-bladed propeller. The initial sequence of clone Wp2A suggested three independent, closely spaced open reading frames: ORF492, ORF375, and ORF783, the last two with GTG initiation codons [14]. ORF375 appeared to contain a fragment of another β propeller domain. PairWise [21] searches with the YVTN β propeller protein sequence from ORF492 and C repeat protein sequence from ORF783 against the translated nucleotide sequences centering on the regions between ORF492 and ORF375 and between ORF375 and ORF783 predicted additional YVTN repeats and one more C repeat, in which frameshift errors were present. Resequencing of these two portions of clone Wp2A revealed that ORF492, ORF375, and ORF783 indeed are portions of a single open reading frame encoding 1673 amino acid residues deposited under GenBank accession number AF394229.

Thus the *M. mazei* SLP precursor of 1673 residues is comprised entirely of a 42-residue signal sequence, 14 YVTN repeats, which correspond to the previously described AB repeats [14], and 12 domains, which correspond to the previously described C repeats (Figure 1A). The YVTN repeats are present in two groups of seven repeats each, which fold into two predicted seven-bladed β propellers. A fragment containing the first predicted β propeller domain and the first type C repeat (Figure 1A) was expressed in the periplasm of *E. coli*, purified, and crystallized.

Overall Structure

The crystal structure was solved at 2.4 Å resolution using MAD (Figure 1B; Table 1). The R factor and R_{free} are 0.214 and 0.263, respectively. Of the 1564 total residues present in the four independent molecules in the asymmetric unit, 1562 are in the most-favored and allowed regions in the Ramachandran plot, and only two surface residues are in the additionally allowed region. The four molecules are related by two noncrystallographic translation vectors and show up to a 10° difference in the angle between the two domains (Figure 1C).

The crystal structure contains representative structures for each of the two types of domains present in the SLP (Figures 1A and 1B). As predicted, residues 1–298 fold into a β propeller domain that contains seven, four-stranded β sheets. These β sheets are also known as blades, because of their arrangement in the propeller, or "W"s, because each antiparallel β sheet traces out a "W" shape. The blades pack radially about a pseudo-7-fold symmetry axis, which passes through a narrow, water-filled central channel (Figure 1B). β strand 1 lines the central channel, and β strand 4 forms the circumference of the torus-shaped propeller. The 4-1 and 2-3 loops define the top, and the 1-2 and the 3-4 loops define the bottom, of the propeller. The β propeller domain is 28 Å high and has a wider diameter at the bottom (43 Å) than at the top (30 Å). Residues 303–383 form an immunoglobulin-like β sandwich domain, which belongs to the polycystic kidney disease (PKD) domain superfamily (see below). The PKD domain is 44 Å long and 15 Å in diameter. The β propeller domain connects to the PKD domain through a linker that extends from the end of β strand 1 of W7, across the bottom of the β propeller domain, and along the junction between W7 and W1 (Figures 1B and 1C).

The PKD domain extends away from the side of the β propeller domain (Figures 1B–1D). The total solvent-accessible surface area buried in the domain interface is 650 $Å^2$. The size of this domain interface is consistent with the limited differences in interdomain angles observed in the four molecules in the crystal structure (Figure 1C). The bottom surface of the β propeller (Figure 1D) stacks on the top face of the β propeller domain of another molecule in the crystal lattice, with a total buried interface area of $\sim$1300 $Å^2$.

The YVTN β Propeller Domain

The YVTN β propeller domain is highly pseudosymmetric, correlating with high sequence identity of 46.6 ± 7.6% between the seven blades. The entire β propeller domain superimposes upon itself with an rmsd of 1.08 ± 0.08 Å for $C\alpha$ atoms when compared at all six pseudosymmetric positions. Some β propellers show much higher symmetry around the pseudosymmetry axis than others. We examined other known highly symmetric β propellers: the eight-bladed cd1 nitrite reductase [22], seven-bladed β propellers in transducin G_β [23] and Tup1 [24], and six-bladed β propellers in the LDL receptor [17] and TolB [25]. The YVTN β propeller is more symmetric than any other six-, seven-, or eight-bladed β propeller described to date, on the basis of rmsd in rotational superposition, the uniformity of the angle between neighboring blades around the pseudosymmetry axis, and the uniformity of the tilt of the blade plane relative to the central axis (data not shown).

The unusually high symmetry of the YVTN β propeller structure is enforced by three clusters of interactions, centered in highly conserved YVTN, SPDGT, and SVIDT sequence motifs (Figure 2A), that stabilize each blade and the orientation between adjacent blades in a uniform manner. The YVTN sequence motif is situated in the middle of β strand 2 and interacts extensively with β strands 3 and 4 from the preceding blade and β strands 1 and 2 in the following blade. The Tyr, Val, and Thr/Ala residues participate in several hydrophobic clusters located in the middle of the blade interface (Figure 2B). The Asn residue of the YVTN motif is situated at the center of a striking hydrogen bond network consisting of three backbone amides (Figure 3A, blade positions 1, 2, and 3) in the loop connecting β strands 4 and 1 (4-1 loop) of adjacent blades and five side chains within the 2-3 loop (Figure 2B; Figure 3A, blade positions 20, 23, 24, 25, and 27). This network constrains the orientation between neighboring blades on the upper edge of the hydrophobic interface. The Tyr of the YVTN motif and the preceding Lys (Figure 3A, blade positions 15 and 17) form side chain hydrogen bonds, which often are

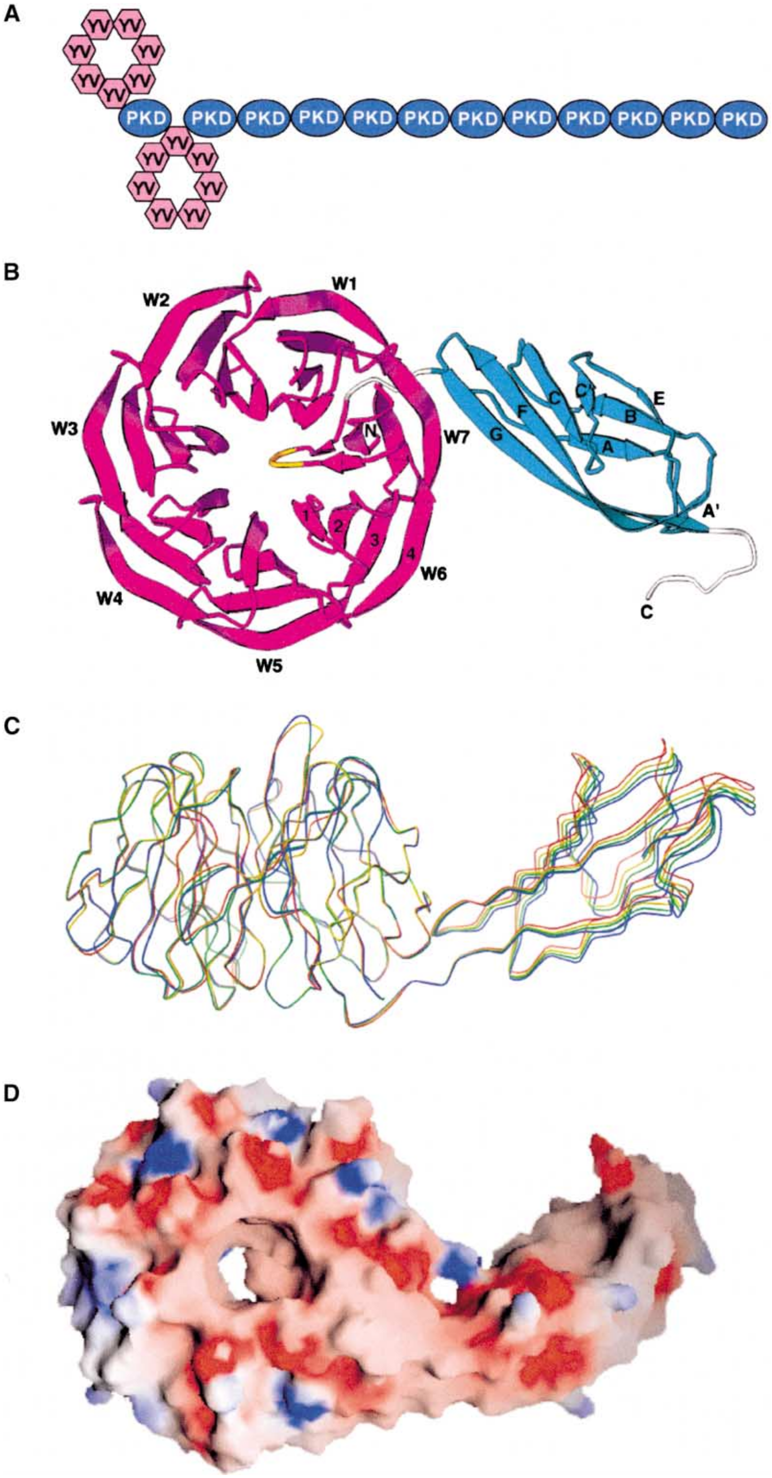


Figure 1. Overall Structure

(A) Schematic domain structure of the full-length *M. mazei* SLP.

(B) Ribbons diagram of YVTN β propeller (magenta) and PKD (blue) domains as viewed from the top of the β propeller domain. Domain linkers, white; bulge on strand 1 of W7, green.

(C) Side view of the Cα traces of four molecules in the asymmetric unit, superimposed on the β propeller domain.

(D) Charges on the bumpy bottom surface of the β propeller domain. The GRASP representation of surface charge is colored from red (−10 kEV) to blue (+10 kEV). The view is ~90 degrees from (C) and ~180 degrees from (B).

water mediated, to residues in the 1-2 and 3-4 loops of the following blade (Figure 2B). This stabilizes the orientation on the lower edge of the hydrophobic blade interface. The SPDGT sequence motif in the 1-2 loop and the SVIDT motif in strand 3 are also highly conserved (Figure 2A), possess substantial hydrogen bond networks, and each contributes to domain stability and the regularity of the 1-2 and 3-4 loops (Figure 3A). Notably, the Asp residue in the SVIDT motif corresponds to the D in the WD40 repeats of the G_β β propeller domain, and its side chain provides the same function of forming a hydrogen bond to the amide group of residue $n + 2$ to stabilize a type I turn.

In less symmetric seven- and eight-bladed β propellers, one of the sheet-sheet interfaces differs markedly from the others, and one of the sheets at this interface

Table 1. Statistics on Diffraction Data and Crystal Structure

Crystallographic Parameters				
Space group	C2			
Cell dimension	a = 188.2 Å, b = 65.9 Å, c = 140.0 Å, β = 124.3°			
Asymmetric unit contents	4 molecules of 41.5 kDa each			
Solvent content (%)	42.7			
Vm (Å^3/Da)	2.16			
Data and Phasing Statistics				
	Peak^a	Inflection	Remote	
Wavelength (Å)	0.979067	0.979221	0.939270	
Resolution limit (Å)	2.4	2.5	2.7	
Total observations	318,719	287,818	231,145	
Unique reflections	52,881	47,105	37,167	
I/σ (I)^b	14.6 (3.1)	14.2 (3.4)	13.3 (3.8)	
R_{merge} (%)^b	10.9 (43.4)	11.1 (40.6)	12.3 (43.2)	
Completeness (%)^b	93.7 (89.0)	94.1 (95.2)	93.3 (95.3)	
R_{ano} (%)^b	5.9 (25.6)	5.5 (20.8)	5.8 (22.1)	
Anomalous completeness (%)^b	88.3 (80.9)	88.5 (88.0)	87.6 (89.4)	
Figure of merit (30–2.8 Å)	0.36			
Structure Statistics				
Resolution (Å)	20–2.4			
R factor (%)	21.4			
R_{free} (%)^c	26.3			
Structure completeness	4 × 391 residues and 1364 waters			
Rmsd bond length (Å)	0.007			
Rmsd bond angle (°)	1.354			
Luzzati coordinate error (Å)	0.29			
Average B Factor (Å^2)	A	B	C	D
Propeller domain	18.0	16.9	17.6	15.8
PKD domain	30.9	35.9	46.5	29.4
Water	32.4			

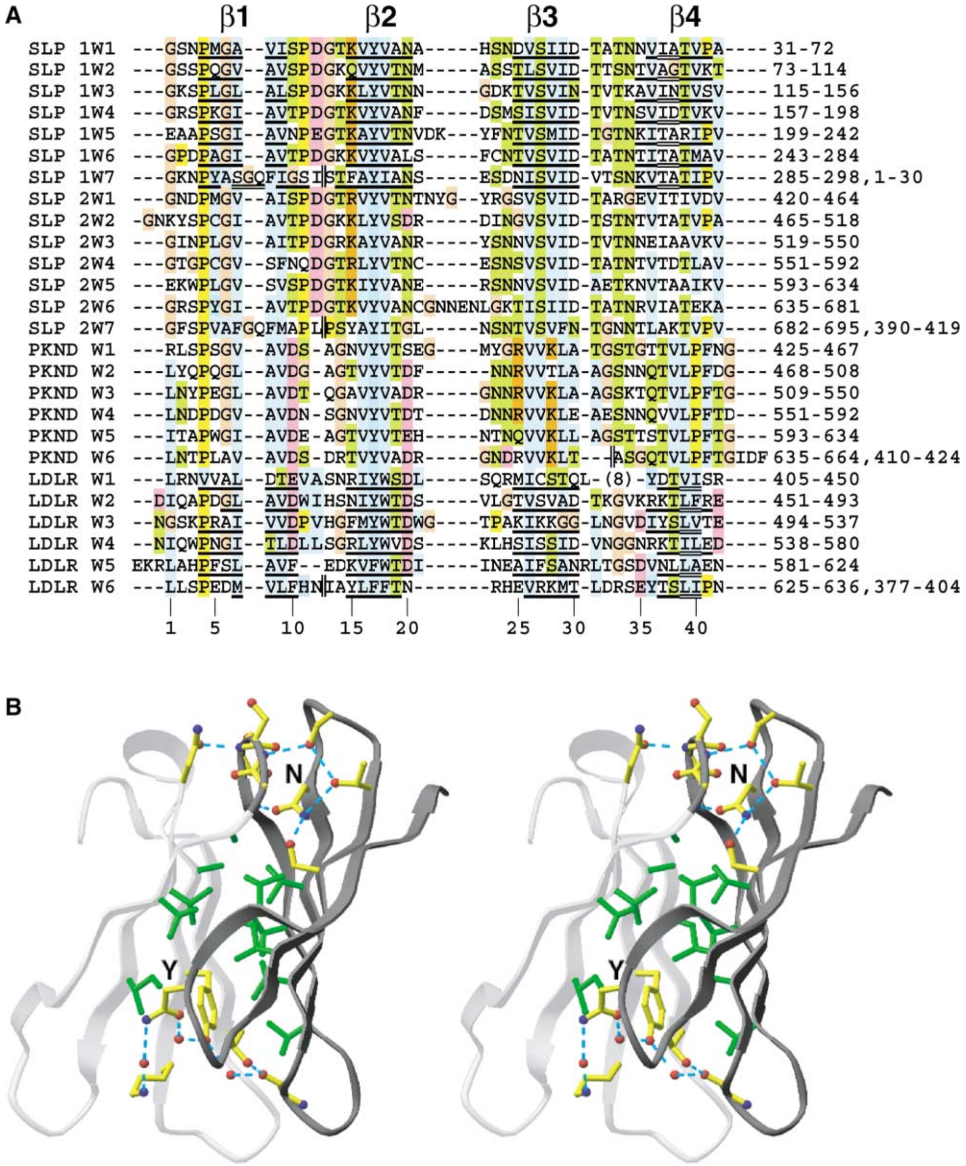
[a] Used in structure refinement after merging anomalous pairs.
[b] Last shell statistics (2.49–2.40 Å) are shown in parentheses.
[c] Five percent of the reflections were allocated to the test set.

slides in to fill part of the central cavity [26]. To prevent this in the YVTN β propeller, three residues from β strand 1 of W7 bulge into the central cavity (Figures 1B and 2A) and help to maintain the pseudosymmetry. The structure and sequence in this bulge is similar to the "Tyr corner" motif [27], but a Phe replaces the Tyr.

Interesting sequence homologs of YVTN β propellers are present in both *Archaea* and eubacteria that have either six or seven blades and, in some cases, differ in the offset between the blades and the sequence repeats. Usually the N- and C-terminal β strands of a β propeller are hydrogen bonded to one another in the last sheet and thus knit together the circular domain. In the seven-bladed YVTN β propeller domain described here, the N-terminal β strand is strand 2 (Figure 2A); thus, the sequence repeats begin with strand 2. By contrast, the N-terminal β strand is strand 4 in PknD, a sequence homolog in *Mycoplasma tuberculosis* ($E < 3 \times 10^{-15}$ with PSI-BLAST) with an extracellular domain composed solely of six repeats, with a YVTD motif in place of the YVTN motif (Figure 2A). This putative six-bladed YVTD β propeller is linked by a type II transmembrane segment to a Ser/Thr protein kinase cytoplasmic domain. This interesting homology suggests that there are evolutionary pathways for interconverting both the number of blades in a β propeller and the beginning β strand. These conversions could occur simultaneously. For example, a seven-bladed propeller beginning with strand 2 could convert to a six-bladed propeller beginning with strand 4 by loss of the first and last two β strands.

We initially predicted the YVTN β propellers because of their sequence homology with metazoan YWTD β propellers, which begin with β strand 2, like the YVTN β propellers, but contain six instead of seven blades (Figure 2A). The YWTD and YVTN motifs in strand 2 and the Pro and Gly residues at positions 1 and 3 of strand 1 have structurally equivalent positions and functions in the two types of propellers (Figures 2A, 2B, 3A, and 3B). Furthermore, the YWTD and YVTN blades are almost identical in number of residues and in backbone hydrogen bonds, except for the position of the β bulge in β strand 4 (Figure 3A). These structural similarities validate the sequence homology and support the evolutionary relationship between YWTD and YVTN β propellers.

Differences between YWTD and YVTN β propellers provide insights into the structural features that lead to six- or seven-bladed propeller folds. The distance between adjacent blades at β strand 1 is similar in six- and seven-bladed propellers, because, in six-bladed propellers, the blades move closer to the pseudosymmetry axis [26] (compare the distance to the axis in Figure 3B). However, the distance between adjacent blades must be greater at β strand 4 in six-bladed pro-



```
SLP 1W1   ---GSNPMGA--VISPDGTKVYVANA-----HSNDVSIID-TATNNVIATVPA---- 31-72
SLP 1W2   ---GSSPQGV--AVSPDGKQVYVTNM-----ASSTLSVID-TTSNTVAGTVKT---- 73-114
SLP 1W3   ---GKSPLGL--ALSPDGKKLYVTNN-----GDKTVSVIN-TVTKAVINTVSV---- 115-156
SLP 1W4   ---GRSPKGI--AVTPDGTKVYVANF-----DSMSISVID-TVTNSVIDTVKV---- 157-198
SLP 1W5   ---EAAPSGI--AVNPEGTKAYVTNVDK---YFNTVSMID-TGTNKITARIPV---- 199-242
SLP 1W6   ---GPDPAGI--AVTPDGKKVYVALS-----FCNTVSVID-TATNTITATMAV---- 243-284
SLP 1W7   ---GKNPYASGQFIGSI‖STFAYIANS-----ESDNISVID-VTSNKVTATIPV---- 285-298,1-30
SLP 2W1   ---GNDPMGV--AISPDGTRVYVTNTNYG--YRGSVSVID-TARGEVITIVDV---- 420-464
SLP 2W2   -GNKYSPCGI--AVTPDGKKLYVSDR-----DINGVSVID-TSTNTVTATVPA---- 465-518
SLP 2W3   ---GINPLGV--AITPDGRKAYVANR-----YSNNVSVID-TVTNNEIAAVKV---- 519-550
SLP 2W4   ---GTGPCGV--SFNQDGTRLYVTNC-----ESNSVSVID-TATNTVTDTLAV---- 551-592
SLP 2W5   ---EKWPLGV--SVSPDGTKIYVANE-----RSNNVSVID-AETKNVTAAIKV---- 593-634
SLP 2W6   ---GRSPYGI--AVTPDGTKVYVANCGNNENLGKTISIID-TATNRVIATEKA---- 635-681
SLP 2W7   ---GFSPVAFGQFMAPL‖PSYAYITGL-----NSNTVSVFN-TGNNTLAKTVPV---- 682-695,390-419
PKND W1   ---RLSPSGV--AVDS-AGNVYVTSEG----MYGRVVKLA-TGSTGTTVLPFNG--- 425-467
PKND W2   ---LYQPQGL--AVDG-AGTVYVTDF------NNRVVTLA-AGSNNQTVLPFDG--- 468-508
PKND W3   ---LNYPEGL--AVDT-QGAVYVADR-----GNNRVVKLA-AGSKTQTVLPFTG--- 509-550
PKND W4   ---LNDPDGV--AVDN-SGNVYVTDT-----DNNRVVKLE-AESNNQVVLPFTD--- 551-592
PKND W5   ---ITAPWGI--AVDE-AGTVYVTEH-----NTNQVVKLL-AGSTTSTVLPFTG--- 593-634
PKND W6   ---LNTPLAV--AVDS-DRTVYVADR-----GNDRVVKLT--‖ASGQTVLPFTGIDF 635-664,410-424
LDLR W1   ---LRNVVAL--DTEVASNRIYWSDL-----SQRMICSTQL-(8)-YDTVISR---- 405-450
LDLR W2   --DIQAPDGL--AVDWIHSNIYWTDS-----VLGTVSVAD-TKGVKRKTLFRE---- 451-493
LDLR W3   --NGSKPRAI--VVDPVHGFMYWTDWG----TPAKIKKGG-LNGVDIYSLVTE---- 494-537
LDLR W4   --NIQWPNGI--TLDLLSGRLYWVDS-----KLHSISSID-VNGGNRKTILED---- 538-580
LDLR W5   EKRLAHPFSL--AVF--EDKVFWTDI-----INEAIFSANRLTGSDVNLLAEN---- 581-624
LDLR W6   ---LLSPEDM--VLFHN‖IAYLFFTN-------RHEVRKMT-LDRSEYTSLIPN---- 625-636,377-404
             |   |     |    |    |       |    |     |    |
             1   5    10   15   20      25   30    35   40
```

Figure 2. The YVTN β Propeller Domain

(A) Blade sequence alignments for the two SLP YVTN β propeller domains, a YVTD six-bladed β propeller domain comprising the entire extracellular domain of the *M. tuberculosis* protein Ser/Thr kinase D (PknD, gi 15608071), and the YWTD β propeller domain of the LDL receptor. β strands are underlined; bulged residues are double underlined; the N and C termini, which are adjacent in the last "knitting" blade in each propeller, are separated by two vertical lines. Residues are colored according to the consensus patterns derived from the blade alignment of each β propeller type with ClustalX [51]. Eight residues disordered in W1 of the YWTD β propeller structure are omitted and indicated in parentheses. Numbers below the alignment show blade positions (see Figure 3A).

(B) Stereo view of the interface between W1 (light gray) and W2 (dark gray). The residues forming the hydrophobic core are in green; the residues forming the hydrogen bond networks (dashed blue lines) are in yellow, with red oxygen atoms and blue nitrogen atoms. The isolated red spheres are water molecules.

pellers because each blade must span one-sixth, instead of one-seventh, of the circumference of the propeller. Indeed, distances measured between equivalent residues on strand 4 of adjacent blades are greater in YWTD (19.37 ± 0.97 Å) than in YVTN (17.38 ± 0.43 Å) β propellers. This separation is regulated in part by the distinct hydrogen bond pattern of the Asn and Asp side chains of the YVTN and YWTD motifs, respectively. Compared with the Asn side chain, the Asp side chain rotates more perpendicularly to the blade, hydrogen bonds to different backbone amide groups in the 4-1 loop, and causes more separation between the 4-1 and 2-3 loops in YWTD β propellers (Figures 3A and 3B). This extra space is filled by the side chain at blade position 1, which is bulky and hydrophobic in YWTD blades, and a Gly in YVTN blades (Figures 2A and 3A). Another notable difference is the bulky Trp of the YWTD motif, which orients out radially toward β strands 3 and 4 and occupies more space in the blade interface than the compact Val of the YVAN motif. Finally, the β bulge

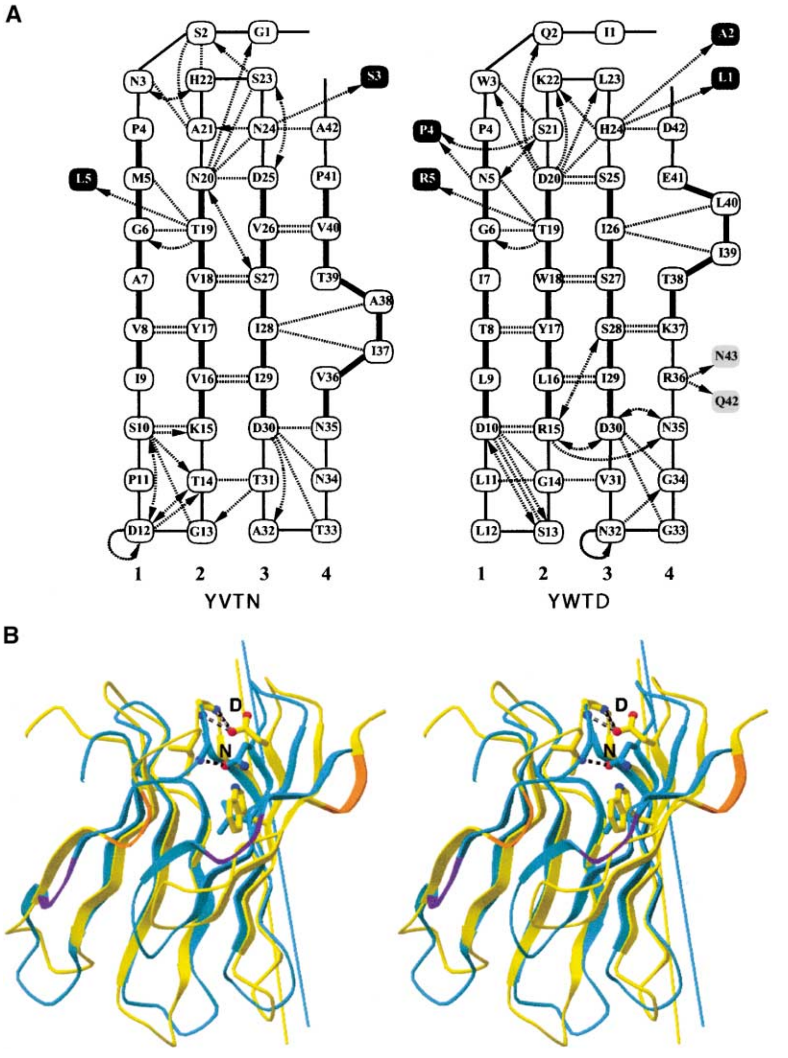


Figure 3. Comparisons of YVTN ad YWTD Blades

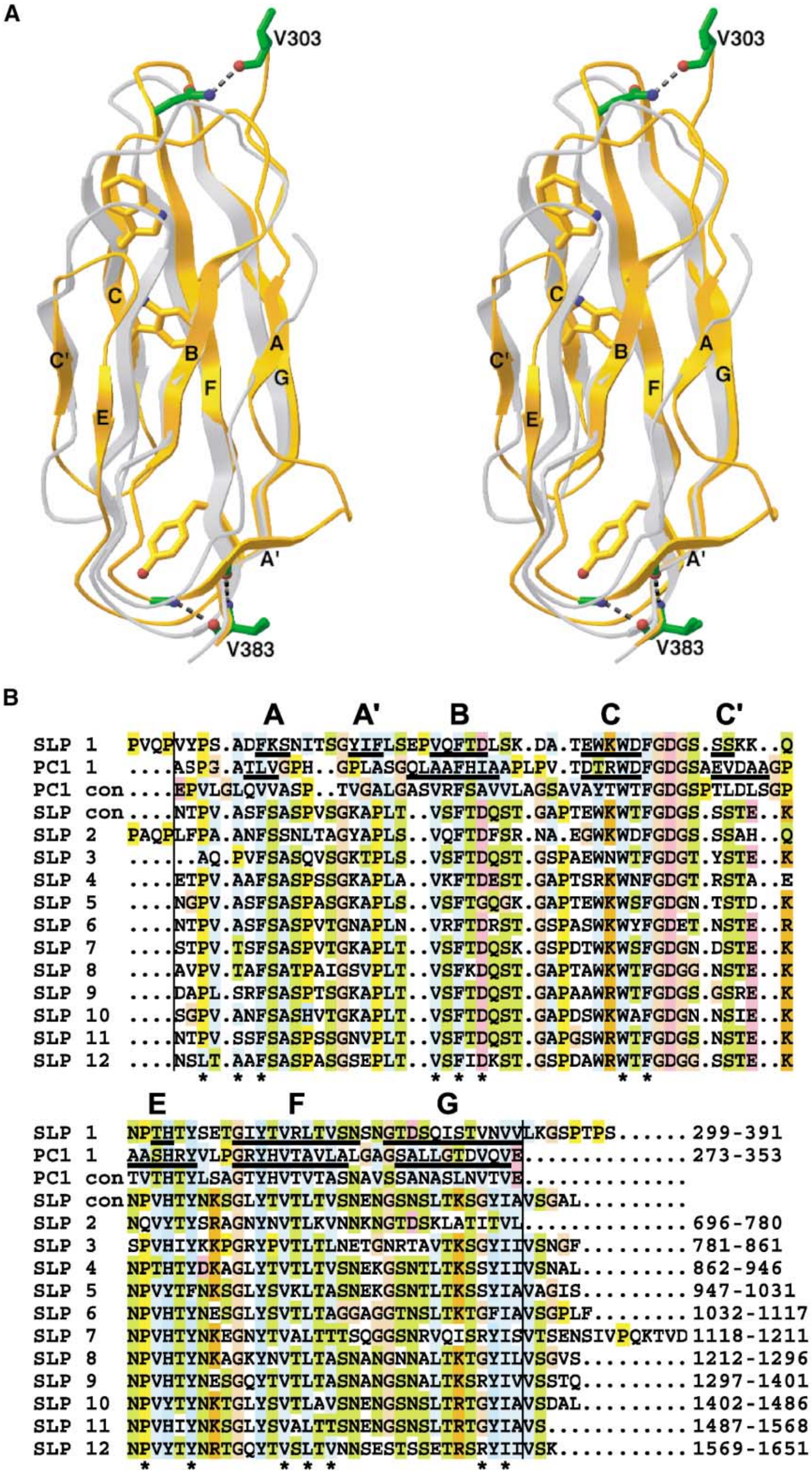
(A) Diagrams of hydrogen bond networks in YVTN and YWTD blades. The numbers refer to the 42 consensus blade positions. Residues and hydrogen bonds are shown for the most typical blade identified from blade superpositions, which is W1 in YVTN blades and W4 in YWTD blades, except for the hydrogen bonds of Thr19 that are adopted from W2 in both β propellers. Residues are connected by thick lines in β strands and thin lines in loops. Dashed lines indicate main chain to main chain hydrogen bonds, dashed lines with single arrows indicate side chain to main chain hydrogen bonds, with the point to main chain, and dashed lines with double arrows indicate side chain to side chain hydrogen bonds. Residues from the preceding blade are in light gray boxes, and those from the following blade are in dark gray boxes.

(B) Comparison of the most typical YVTN (blue) and YWTD (yellow) blade pairs superimposed on the first blade. Side chains in the blade interface are shown for the Val and Asn of the YVTN motif and for the Trp and Asp of the YWTD motif. The β bulges in strand 4 are shown in purple and orange, respectively. The pseudosymmetry axis of each β propeller is shown as a color-keyed straight line.

in strand 4 is offset by two residues in YWTD blades relative to the YVTN blades (Figure 3A). The two bulged side chains are larger, more hydrophobic, and located more equatorially in YWTD blades, and these side chains make more interblade contacts than the equivalent residues in YVTN blades. The position of the bulge in the six-bladed YWTD propeller is the same as it is in the six-bladed TolB [25] and five-bladed tachylectin propellers [28]. By contrast, the position of the bulge in the seven-bladed YVTN propeller is the same as it is in the seven-bladed G protein β subunit [23], Tup1 [24], nitrous oxide reductase [29], and methylamine dehydrogenase [30] propellers. Therefore, the position of the β bulge appears to be relevant to whether propellers are six bladed or seven bladed. The predicted six-bladed *M. tuberculosis* YVTD β propeller lacks the Trp of the YWTD motif; however, the Asp residue of the YVTD motif and the bulky hydrophobic side chain in blade position 1 suggest the larger separation between the 4-1 and 2-3 loops seen in six-bladed YWTD propellers. Furthermore, bulky hydrophobic residues in β strand 4 are in a position appropriate for the bulge of six-bladed propellers.

The PKD Domain

The PKD domain contains eight β strands arranged in two antiparallel β sheets (Figure 4). One sheet contains the ABE strands, and the other contains the A′GFCC′ strands. This topology is similar to that described for the I2 subset of the immunoglobulin superfamily (IgSF) [31], the fibronectin type III (FN3) superfamily, and the PKD superfamily [18], all of which belong to the immunoglobulin-like β sandwich fold [32]. However, a DALI search [33] with this domain revealed the highest similarity to the NMR structure of PKD domain 1 from human polycystin-1, the sole previous structural representative of the PKD superfamily [18]. The two structures superimpose with an rmsd of 2.2 Å for 76 aligned Cα positions (Figure 4A). In agreement, statistically significant sequence homology is detectable between them with sensitive methods [18] (Figure 4B). Several features confirm that these two structures define a superfamily distinct from other Ig-like folds [18]. The second Trp of strand C, although conserved in both PKD and IgSF domains, packs in a different position in the hydrophobic cores of the two domains [18]. A highly conserved tyrosine in



```
                   A       A'       B            C         C'
SLP 1  PVQPVYPS.ADFKSNITSGYIFLSEPVQFTDLSK.DA.TEWKWDFGDGS.SSKK..Q
PC1 1  ....ASPG.ATLVGPH..GPLASGQLAAFHIAAPLPV.TDTRWDFGDGSAEVDAAGP
PC1 con....EPVLGLQVVASP..TVGALGASVRFSAVVLAGSAVAYTWTFGDGSPTLDLSGP
SLP con....NTPV.ASFSASPVSGKAPLT..VSFTDQST.GAPTEWKWTFGDGS.SSTE..K
SLP 2  PAQPLFPA.ANFSSNLTAGYAPLS..VQFTDFSR.NA.EGWKWDFGDGS.SSAH..Q
SLP 3  ......AQ.PVFSASQVSGKTPLS..VSFTDQST.GSPAEWNWTFGDGT.YSTE..K
SLP 4  ....ETPV.AAFSASPSSGKAPLA..VKFTDEST.GAPTSRKWNFGDGT.RSTA..E
SLP 5  ....NGPV.ASFSASPASGKAPLS..VSFTGQGK.GAPTEWKWSFGDGN.TSTD..K
SLP 6  ....NTPV.ASFSASPVTGNAPLN..VRFTDRST.GSPASWKWYFGDET.NSTE..R
SLP 7  ....STPV.TSFSASPVTGKAPLT..VSFTDQSK.GSPDTWKWSFGDGN.DSTE..K
SLP 8  ....AVPV.TAFSATPAIGSVPLT..VSFKDQST.GAPTAWKWTFGDGG.NSTE..K
SLP 9  ....DAPL.SRFSASPTSGKAPLT..VSFTDQST.GAPAAWRWTFGDGS.GSRE..K
SLP 10 ....SGPV.ANFSASHVTGKAPLT..VSFTDQST.GAPDSWKWAFGDGN.NSIE..K
SLP 11 ....NTPV.SSFSASPSSGNVPLT..VSFTDQST.GAPGSWRWTFGDGS.NSTE..K
SLP 12 ....NSLT.AAFSASPASGSEPLT..VSFIDKST.GSPDAWRWTFGDGG.SSTE..K
             *  * *              * * *           * *

         E           F            G
SLP 1  NPTHTYSETGIYTVRLTVSNSNGTDSQISTVNVVLKGSPTPS......299-391
PC1 1  AASHRYVLPGRYHVTAVLALGAGSALLGTDVQVE..............273-353
PC1 conTVTHTYLSAGTYHVTVTASNAVSSANASLNVTVE.............
SLP conNPVHTYNKSGLYTVTLTVSNENGSNSLTKSGYIAVSGAL........
SLP 2  NQVYTYSRAGNYNVTLKVNNKNGTDSKLATITVL..............696-780
SLP 3  SPVHIYKKPGRYPVTLTLNETGNRTAVTKSGYIIVSNGF.........781-861
SLP 4  NPTHTYDKAGLYTVTLTVSNEKGSNTLTKSSYIIVSNAL.........862-946
SLP 5  NPVYTFNKSGLYSVKLTASNEKGSNTLTKSSYIAVAGIS.........947-1031
SLP 6  NPVHTYNESGLYSVTLTAGGAGGTNSLTKTGFIAVSGPLF........1032-1117
SLP 7  NPVHTYNKEGNYTVALTTTSQGGSNRVQISRYISVTSENSIVPQKTVD1118-1211
SLP 8  NPVHTYNKAGKYNVTLTASNANGNNALTKTGYILVSGVS.........1212-1296
SLP 9  NPVHTYNESGQYTVTLTASNANGSNALTKSRYIVVSSTQ.........1297-1401
SLP 10 NPVYTYNKTGLYSVTLAVSNENGSNSLTRTGYIAVSDAL.........1402-1486
SLP 11 NPVHIYNKSGLYSVALTTSNENGSNSLTRTGYIAVS............1487-1568
SLP 12 NPVYTYNRTGQYTVSLTVNNSESTSSETRSRYIIVSK...........1569-1651
        *   *       * * *            * *
```

Figure 4. The PKD Domain

(A) Stereo view of the SLP PKD domain (gold) with the superimposed NMR structure of PKD domain 1 of polycystin-1 (silver). Side chains are shown in the SLP PKD domain for the two Trp residues (gold) conserved on strand C and the Tyr residue forming the Tyr corner near the beginning of strand F (gold). The residues delimiting domain boundaries (V303 and V383) are shown with the side chains or backbone atoms of their hydrogen bonding partners (green).

(B) Structure-based sequence alignments of PKD domain 1 in the *M. mazei* SLP (SLP 1) and in human polycystin-1 (PC1 1), together with the aligned sequences of the other 11 PKD domains (SLP 2–SLP 12) in the *M. mazei* SLP, the consensus of the 12 SLP PKD domains (SLP con), and the consensus of 16 PKD domains in polycystin-1 (PC1 con). β strands are underlined; residues buried in the SLP PKD domain are indicated by asterisks; residues disordered near the N terminus of the polycystin-1 PKD domain structure are in italics; domain boundaries are indicated by two vertical dashed lines. Residues are colored according to the consensus patterns derived from the SLP PKD domain alignment and polycystin-1 PKD domain alignment with ClustalX [51].

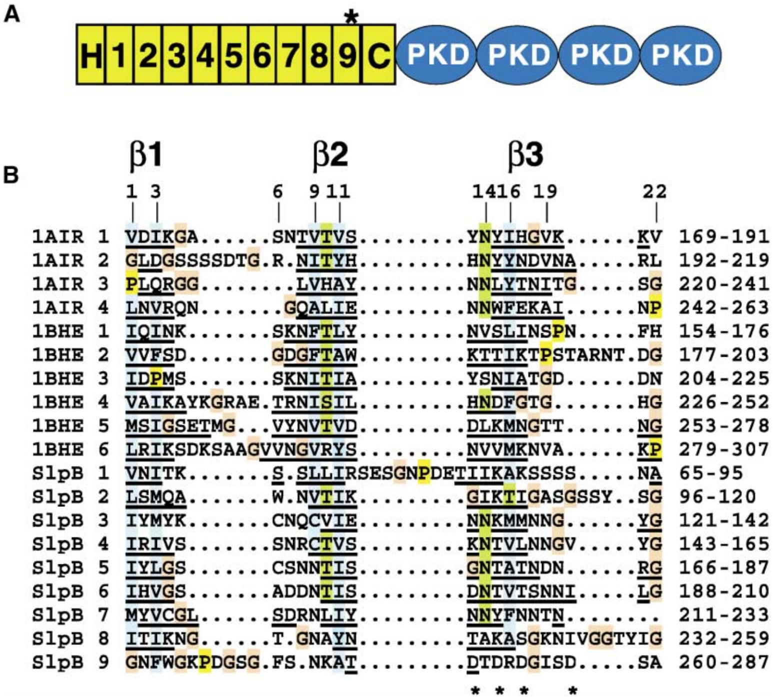


Figure 5. The β Helix Domain

(A) Domain architecture of SlpB. The 652-residue protein contains a signal peptide (residues 1–30; data not shown), an "H" segment (31–64) predicted to contain an α helix and two short β strands, nine β-helical repeats (65–287; one is asterisked because it contains a Ca^{2+} binding site; see panel [B]), a "C" segment (288–303) with predicted coiled conformation, and four PKD domains (308–387, 394–474, 477–561, and 569–648; interdomain linkers not included).

(B) Sequence alignment of nine β-helical repeats in SlpB with structurally defined repeats in pectate lyase C from *Erwinia Chrysanthemi* (Protein Data Bank ID 1AIR [52]) and polygalacturonase from *Erwinia Carotovora* (Protein Data Bank ID 1BHE [53]). Structurally observed or predicted β strands are underlined; the positions of conserved residues are marked on the top; asterisks indicate Asp residues in repeat 9 of SlpB, which may dictate a β hairpin Ca^{2+} binding motif [54]. Residues are colored according to the consensus patterns derived from this alignment with ClustalX [51].

strand F of both IgSF and PKD domains forms a "Tyr corner," characteristic of Ig-like folds [27], which makes one corner of the turn in the E-F loop (Figures 4A and 4B). However, in both PKD domain structures, the Tyr of this corner is shifted down by two ladder positions relative to IgSF domains and, thus, is not structurally equivalent [18]. This shift in position makes the hydrophobic core in PKD domains more elongated than in IgSF domains. Thus, the PKD domain indeed belongs to a distinct superfamily [18], albeit adopting the similar Ig-like fold.

However, there are some significant differences between the two PKD domains. The top of the sheet containing the C, F, and G strands is shifted in the SLP PKD domain to make way for the bulky Trp at the beginning of strand C, which is a Thr in PKD domain 1 of polycystin-1 (Figure 4B). SLP PKD domain 1 may be more representative of polycystin-1 PKD domains 2–16 in this respect because they have bulky Tyr or Phe residues in this position. Also, the position of strand E is markedly shifted. The position of this edge strand would have been hard to assign accurately in the NMR structure because it has hydrogen bonds only on one side. A conserved His in this strand was noted to be in an unusual hydrophobic environment in the NMR structure [18], whereas it is less buried and forms side chain hydrogen bonds in the SLP crystal structure.

The structure of the linked β propeller and PKD domains allows us to delimit the boundaries of all domains in this SLP, estimate the length of the interdomain linkers, and gain insight into the architecture of the whole molecule. Intradomain hydrogen bonds to the main chain N and O atoms of residues Val-303 and Val-383 delimit the first PKD domain (Figure 4A). These boundaries are likely to be conserved in the other SLP PKD domains because the hydrogen bonds are donated by the highly conserved side chain of Asn-369 in the FG turn at the N-terminal end (Figure 4B) and by the main chain at the C-terminal end. Hydrogen bonds within the first β propeller domain and strong homology (51% identity) to the second β propeller domain delimit these domains (Figure 2A). The linker between β propeller domain 1 and PKD domain 1, residues 299–302 (PVQP), and the linker between PKD domain 1 and β propeller domain 2, residues 384–389 (LKGSPT), are both included in the crystal structure. The linker between β propeller domain 2 and PKD domain 2 (PAQP) is almost identical to that between β propeller domain 1 and PKD domain 1, suggesting a similar architecture (Figure 4B). The linkers between PKD domains 3–7 and 8–12 are, on average, five residues in length and attach to the N- and C-terminal tips of the PKD domain (Figure 4A), suggesting considerable scope for interdomain flexibility. The linker between PKD domains 7 and 8 of 14 residues should permit even greater interdomain flexibility. However, the junction between PKD domains 2 and 3 is truncated, so that the first two residues of PKD domain 3 are deleted (Figure 4B), suggesting a tight interdomain linkage. The linker analysis suggests that the N-terminal portion of the SLP containing two β propeller and three PKD domains has a reasonably well-defined orientation between domains, whereas the C-terminal portion containing nine PKD domains exhibits considerable segmental flexibility.

A β Helix Domain in Other SLPs

We examined a second SLP that has been characterized in *M. mazei* and found to be a major surface antigen, SlpB [6, 13]. Four type C repeats, i.e., PKD domains, are present in the C-terminal portion of SlpB (Figure 5A). Searching the N-terminal portion for repeats using PAIRWISE [21] revealed previously unreported repeats

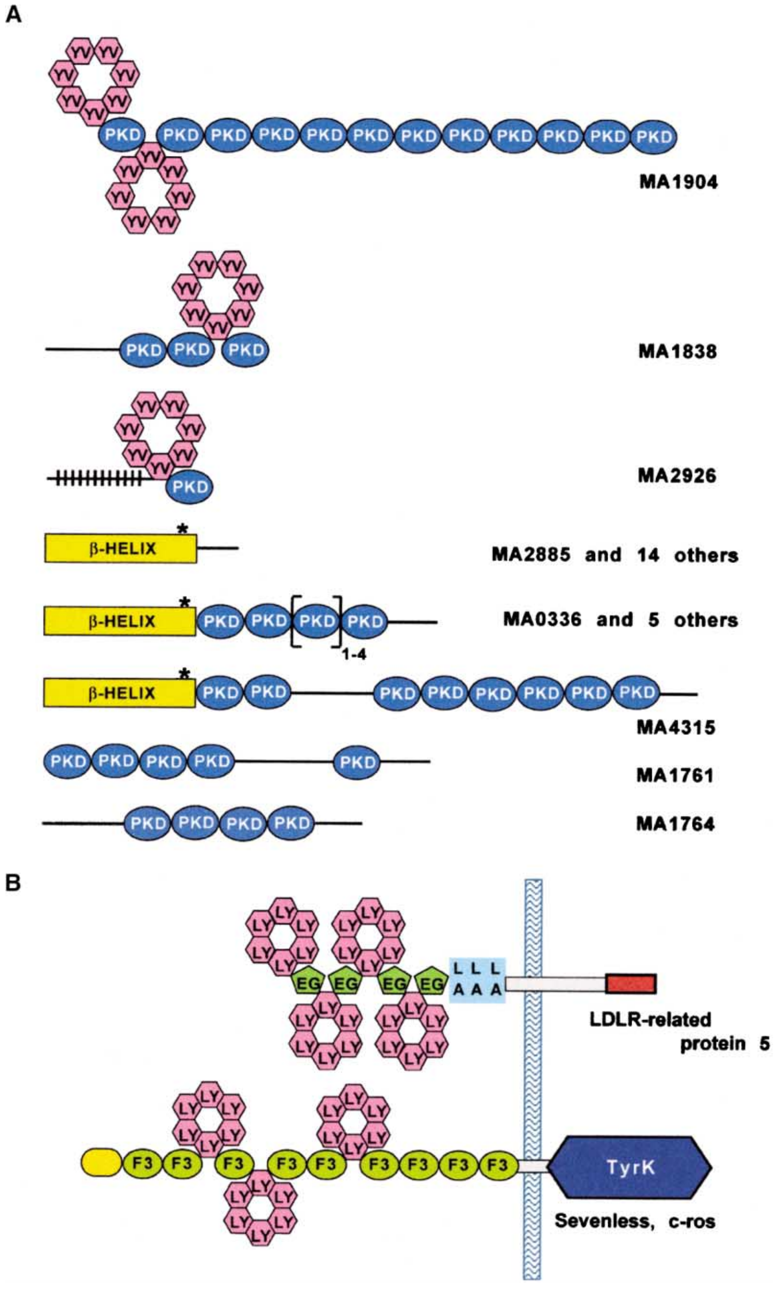


Figure 6. Domain Architectures of Related Surface Proteins in *Archaea* and *Metazoa*

(A) Examples of *M. acetivorans* surface proteins containing YVTN β propeller, PKD, and β helix domains. Protein locus numbers beginning with MA are shown to the right [19].
(B) Examples of metazoan cell surface receptors containing YWTD β propeller domains. The domain symbols and abbreviations are adopted from Bork and Bairoch [55] (or http://www.bork.embl-heidelberg.de/modules/), except that YV is used instead of LY (LDLR YWTD repeats) for the YVTN repeats. EG, epidermal growth factor; F3, fibronectin type III; LA, LDLR type A; TyrK, tyrosine kinase. The asterisk indicates a β hairpin Ca^{2+} binding motif. The straight line with vertical bars is a low-complexity region homologous to human zonadhesin [19].

of ~22 residues (Figures 5A and 5B). Using the 3D-PSSM server [34], we found the entire region between the signal peptide and the first PKD domain to be significantly homologous to the pectate lyase-like superfamily of the single-stranded, right-handed parallel β helix fold (expectation values of 0.0037–0.0086 to four structures). The pectate lyase-like β helix domains contain an N-terminal α helix and 9–16 tandem repeats with 22 consensus positions, which form three major β strands (β1, β2, and β3) and three turns or loops (T1, T2, T3) [35]. Of the 22 consensus positions, 9 positions are buried and are often conserved as hydrophobic residues, except for Ser at position 6 in T1, Asn at position 14 in T2, which dictates an Asn ladder across the repeats, and small residues at position 22 in T3 [35].

Similarly, in the N-terminal portion of SlpB, an N-terminal α helix and usually three β strands per repeat are predicted that align with the secondary structures of the pectate lyase superfamily (Figure 5B). In further similarity, hydrophobic residues are conserved at positions 1, 3, 11, and 16, and Ser, Asn, and Gly are found at positions 6, 14, and 22, respectively (Figure 5B). Position 9 in the last five SlpB β helix repeats is changed from a hydrophobic residue to Asn, suggesting the presence of an additional Asn ladder. These similarities support the assignment to the pectate lyase-like superfamily. Thus, a signal sequence, a β helix domain, and four PKD domains account for the entirety of *M. mazei* SlpB.

The region of SlpB containing the β helix domain is homologous [13] to the N-terminal portion of SlpA of the *Archaea Methanothermus sociabilis* and *Methanothermus fervidus* [36], which we confirmed also contain 22-residue repeats (data not shown). Furthermore, using the SlpB β helix domain, we detected strong homology ($E < 3 \times 10^{-36}$ with PSI-Blast) to SLPs in other sequenced *Archaea*, such as SlpB-1 and SlpB-2 of *Archaeoglobus fulgidus* and four SLPs in *Methanobacterium thermoautotrophicum* (gi7482051, gi7482052, gi7482054, and gi 2621804). Recently, sequence homologs of the pectate lyase superfamily have been found with other approaches, including SLPs [37, 38]. Domain homologs are also present in some eubacterial cell surface and pathogenic proteins and in some human and mouse proteins [37, 38].

YVTN β propeller, PKD, and β Helix Domains in the *Methanosarcina acetivorans* Genome

In the 5.7 Mb genome of *Methanosarcina acetivorans* [19], we found a large number of proteins with YVTN β propeller, PKD, and β helix domains. The methanogen *M. acetivorans* is closely related to *M. mazei* in the regulated ability to grow as single cells, lamina, or packets of firmly adherent cells and in the presence of a noncrystalline S layer that is closely associated with the cytoplasmic membrane. *M. acetivorans* conserves not only the overall domain organization of the two SLPs described above, but also the exact number of tandem PKD domains (Figure 6A). Furthermore, many other extracellular proteins contain YVTN β propeller, PKD, and β helix domains, representatives of which are shown in Figure 6A. These proteins contain signal peptides and lack transmembrane domains, suggesting they may be SLPs. The PKD domain has been particularly expanded in *M. acetivorans*, relative to other prokaryotes, and appears in 60 different proteins [19]. The PKD domain often associates with either YVTN β propeller or β helix domains, but, intriguingly, the YVTN β propeller and β helix domains do not coexist in the same proteins in this genome (Figure 6A). The multiplicity of SLPs with diverse domain architectures (Figure 6A) suggests that the assembly of SLPs in *Methanosarcina* is more complex than the assembly of the monomolecular and crystalline S layers in other *Archaea* and eubacteria; this correlates with the absence of crystallinity in *M. mazei* and *M. acetivorans* S layers and the regulated ability of these species to form multicellular structures.

Biological Implications

It is striking that *Archaea* that demonstrate regulated cell-cell interactions contain domains in their SLPs that are closely related to those in metazoan cell surface and extracellular matrix proteins. The evolutionary relationship between the YWTD and YVTN sequence repeats in *Metazoa* and *Archaea*, respectively, and between the β propeller domains into which they fold, is described here for the first time. We speculate that the expansion in the number of SLPs in *Methanosarcina* relative to other sequenced *Archaea* and in the number of proteins that contain PKD and YVTN domains relates to their function in cell-cell interactions. As described in the Introduction, growth of *Methanosarcina* as single cells, packets, or lamina is regulated by environmental conditions, and there are correlated changes in the expression of the surface layer protein studied here [8]. The related YWTD β propeller domains and PKD domains are found in metazoans on the cell surface and in the extracellular matrix, but not in the unicellular eukaryote *S. cerevesia*, other fungi, or plants [39]. YWTD β propeller domains are found in a large number of endocytic receptors, in receptors that regulate development, and in the laminin binding domain of nidogen, an extracellular matrix component present in all sequenced metazoans [16]. The arrangement of YWTD β propeller domains in multidomain proteins in metazoans is similar to that of the YVTN β propeller domains in *Methanosarcina*, with epidermal growth factor domains or fibronectin type III domains replacing the PKD domains (Figure 6B). PKD domains are found in multiple metazoan cell surface proteins [18, 39], including several on melanocytes and polycystin-1 on kidney epithelial cells. Polycystin-1 associates with polycystin-2 to form a nonselective cation channel [40], and its 16 tandem PKD domains mediate intercellular contacts between epithelial cells through high-affinity homophilic interactions [41]. *Archaea* arose in evolution long before *Metazoa*. Therefore, our finding of YVTN β propeller and PKD domains in archaeal surface layer proteins suggests that proteins that already were adapted for cell surface functions and intercellular interactions in *Archaea* later evolved to regulate cell-cell interactions in *Metazoa*.

Previous structural studies on SLPs have been limited to the prediction or determination of the structure of short terminal segments [9, 12]. The three types of SLP domains described here are substantially larger and constitute the entirety or the majority of numerous archaeal SLPs. Thus, our work provides the first insight into the domains that are responsible for the overall architecture of SLPs.

Experimental Procedures

Homolog Detection and Fold Prediction

Sequence repeats were identified and consensus sequences were derived with PairWise [21]. Secondary structures were predicted using PHD [42] and PSI-Pred [43]. Sequence and structure homologs were identified through iterative PSI-Blast searches [20] and position-specific scoring matrices with 3D-PSSM [34]. Sequences were aligned by structure or with PRRP [44] and manual adjustment.

Protein Expression, Purification, and Crystallization

The N-terminal segment of the *M. mazei* S-6 SLP encompassing the signal peptide, the first YVTN β propeller domain, and the first PKD domain was PCR amplified with the genomic DNA clone Wp2A [14] (gift from Dr. E.C. de Macario, Wadsworth Center, New York State Department of Health) as template. A 6 × His tag was attached to the C-terminal end of this construct, and it was cloned into the pQE-60 vector (QIAGEN) with EcoRI and BglII sites; a C-terminal linker (Ala-Gly-Ser) is present before the hexahistidine tag. These additional nine residues are disordered in the crystal structure.

The authentic signal peptide (42 aa residues) enabled periplasmic expression in *E. coli* strain M15. Transformed bacteria were cultured in LB media containing ampicillin and kanamycin at 37°C. Cultures were induced with 1 mM IPTG at an $A_{600} = 0.4$–0.5, and bacteria were harvested 6 hr later. Cells were washed and suspended in lysis buffer (50 mM phosphate, 0.3 M NaCl, 10 mM imidazole [pH 8.0], 1 μg/ml leupeptin, 1 μg/ml pepstatin A, and 0.5 mM PMSF), lysed by freezing and thawing four times, and sonicated. The supernatant from the lysate was applied to Ni-NTA agarose (QIAGEN) and eluted in lysis buffer containing 0.25 M imidazole. This resulted in a ∼90% pure protein, which showed a 42 kDa band on reducing SDS-PAGE, but nonreducing SDS-PAGE revealed both the 42 kDa band and a disulfide-linked homodimer of 85 kDa. Purified protein was in equilibrium between monomer and dimer; the monomeric 42 kDa protein purified on a gel filtration column developed the 85 kDa species during storage, and, conversely, the isolated dimeric 85 kDa protein developed the 42 kDa species during storage. Apparently, the presence of unpaired Cys-265 caused this behavior. Therefore, Ni-NTA-purified material was reduced with 25 mM β-mercaptoethanol and alkylated with 25 mM iodoacetamide at pH 8.0. Alkylated protein was further purified on a Mono-S cation exchange column (Pharmacia) equilibrated with 50 mM MES and 100 mM NaCl (pH 6.0), with a linear gradient of NaCl up to 500 mM. The monomeric protein eluted at about 250 mM NaCl. The SeMet-substituted protein was produced as described [45] in M9 minimal medium and purified using a similar protocol.

The purified sample was concentrated to 13 mg/ml and crystallized using the hanging drop method at room temperate by adding an equal volume of the well solution of 20% PEG 8000, 0.3 M KCl, and 0.1 M NaAc (pH 3.8). The crystals could be frozen after soaking in a solution additionally containing 10% glycerol.

Data Collection, Phase Determination, and Structure Refinement

Multiple-wavelength anomalous dispersion (MAD) data were collected on beamline SBC-CAT, 19ID, at the Advanced Photon Source. The crystallographic parameters, qualities of diffraction data at three wavelengths, and phasing statistics are shown in Table 1. The native Patterson map showed two strong peaks located approximately at (0.25, 0.25, 0.0) and (0.5, 0.0, 0.0), with peak heights of approximately one-half and one-fourth, respectively, of the origin peak. These peaks indicate that two noncrystallographic translation vectors relate four molecules in the asymmetric unit, and the difference in peak heights shows that the (0.25, 0.25, 0.0) translation is stricter than the other. As a result, half of the reflections were weak, a quarter of the reflections were strong, and the remaining quarter of the reflections were close to extinction.

Five SeMet substitutions were present in each β propeller domain. Their positions were found by visual inspection on the basis of anomalous Patterson and difference Fourier maps calculated with CCP4 for a reduced, small pseudocell, with a quarter of the cell volume, which, like the unit cell, was in the C2 space group. Then, the 20 Se sites in the asymmetric unit of the real cell were derived using the translation vectors and refined. The experimental MAD map was much improved by the solvent-flipping density modification procedure in CNS [46], which clearly delineated most of the protein boundaries. Combining the maps calculated at different resolution ranges allowed building of most residues in the β propeller domain and the positioning of five major strands in the PKD domain. However, the R factors were still high, mostly because the predominance of weak reflections deteriorated the scaling between F_o and F_c. Therefore, we split the reflections into strong, weak, and very weak sets and used only the weak set for rigid-body refinement in the earlier stages to better determine the differences in interdomain orientations. After cycles of manual rebuilding in O [47] and reciprocal space refinement in CNS, the model gradually became more complete and accurate, and the R factors dropped. After the first two refinement cycles employing strict noncrystallographic restraints, different weights for the restraints were tested and optimized in each following cycle. The restraints were gradually released as the R factor fell.

The final structure has been validated by stereochemical criteria implemented in PROCHECK, CNS, O, and ADIT. The structures were aligned with O, LSQMAN, or STAMP [48], and then visual analysis was performed in O. The structure figures were prepared with Ribbons [49] and GRASP [50].

Acknowledgments

We would like to thank Dr. Stephen Harrison at Harvard University for helpful suggestions, Dr. Kemin Tan at the Dana-Farber Cancer Institute for help in data collection, and Drs. James Galagan, Bruce Birren, and Eric Lander at the Whitehead Institute for access to the *M. acetivorans* genome sequence before publication. The project was supported by NIH grants HL48675 to T.A.S. and J.H.W. and DOE contract W-31-109-Eng-38 to A.J.

Received: May 1, 2002
Revised: July 2, 2002
Accepted: July 3, 2002

References

1. Konig, H. (1988). Archaeobacterial cell envelopes. Can. J. Microbiol. *34*, 395–406.
2. Messner, P., and Sleytr, U.B. (1992). Crystalline bacterial cell-surface layers. Adv. Microb. Physiol. *33*, 213–275.
3. Baumeister, W., and Lembcke, G. (1992). Structural features of archaebacterial cell envelopes. J. Bioenerg. Biomembr. *24*, 567–575.
4. Beveridge, T.J. (1994). Bacterial S-layers. Curr. Opin. Struct. Biol. *4*, 204–212.
5. Sara, M., and Sleytr, U.B. (2000). S-Layer proteins. J. Bacteriol. *182*, 859–868.
6. Mayerhofer, L.E., Conway de Macario, E., Yao, R., and Macario, A.J.L. (1998). Structure, organization, and expression of genes coding for envelope components in the archaeon *Methanosarcina mazei* S-6. Arch. Microbiol. *169*, 339–345.
7. Mayerhofer, L.E., Macario, A.J., and de Macario, E.C. (1992). Lamina, a novel multicellular form of *Methanosarcina mazei* S-6. J. Bacteriol. *174*, 309–314.
8. Macario, A.J., and Conway De Macario, E. (2001). The molecular chaperone system and other anti-stress mechanisms in archaea. Front. Biosci. *6*, D262–283.
9. Lupas, A., Engelhardt, H., Peters, J., Santarius, U., Volker, S., and Baumeister, W. (1994). Domain structure of the *Acetogenium kivui* surface layer revealed by electron crystallography and sequence analysis. J. Bacteriol. *176*, 1224–1233.
10. Olabarria, G., Carrascosa, J.L., de Pedro, M.A., and Berenguer, J. (1996). A conserved motif in S-layer proteins is involved in peptidoglycan binding in *Thermus thermophilus*. J. Bacteriol. *178*, 4765–4772.
11. Mesnage, S., Fontaine, T., Mignot, T., Delepierre, M., Mock, M., and Fouet, A. (2000). Bacterial SLH domain proteins are non-covalently anchored to the cell surface via a conserved mechanism involving wall polysaccharide pyruvylation. EMBO J. *19*, 4473–4484.
12. Stetefeld, J., Jenny, M., Schulthess, T., Landwehr, R., Engel, J., and Kammerer, R.A. (2000). Crystal structure of a naturally occurring parallel right-handed coiled coil tetramer. Nat. Struct. Biol. *7*, 772–776.
13. Yao, R., Macario, A.J.L., and Conway de Macario, E. (1994). An archaeal S-layer gene homolog with repetitive subunits. Biochim. Biophys. Acta *1219*, 697–700.
14. Mayerhofer, L.E., Conway de Macario, E., and Macario, A.J.L. (1995). Conservation and variability in *Archaea*: protein antigens with tandem repeats encoded by a cluster of genes with common motifs in *Methanosarcina mazei* S-6. Gene *165*, 87–91.
15. Bork, P., Downing, A.K., Kieffer, B., and Campbell, I.D. (1996). Structure and distribution of modules in extracellular proteins. Q. Rev. Biophys. *29*, 119–167.
16. 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.
17. Jeon, H., Meng, W., Takagi, J., Eck, M.J., Springer, T.A., and Blacklow, S.C. (2001). Implications for familial hypercholesterolemia from structure of the LDL receptor YWTD-EGF domain pair. Nat. Struct. Biol. *8*, 499–504.
18. Bycroft, M., Bateman, A., Clarke, J., Hamill, S.J., Sandford, R., Thomas, R.L., and Chothia, C. (1999). The structure of a PKD domain from polycystin-1: implications for polycystic kidney disease. EMBO J. *18*, 297–305.
19. Galagan, J.E., Nusbaum, C., Roy, A., Endrizzi, M.G., Macdonald, P., FitzHugh, W., Calvo, S., Engels, R., Smirnov, S., Atnoor, D.,

et al. (2002). The genome of *Methanosarcina acetivorans* reveals extensive metabolic and physiological diversity. Genet. Res. *12*, 532–542.

20. Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. *25*, 3389–3402.
21. Birney, E., Thompson, J.D., and Gibson, T.J. (1996). PairWise and SearchWise: finding the optimal alignment in a simultaneous comparison of a protein profile against all DNA translation frames. Nucleic Acids Res. *24*, 2730–2739.
22. Williams, P.A., Fulop, V., Garman, E.F., Saunders, N.F., Ferguson, S.J., and Hajdu, J. (1997). Haem-ligand switching during catalysis in crystals of a nitrogen-cycle enzyme. Nature *389*, 406–412.
23. Lambright, D.G., Sondek, J., Bohm, A., Skiba, N.P., Hamm, H.E., and Sigler, P.B. (1996). The 2.0 Å crystal structure of a heterotrimeric G protein. Nature *379*, 311–319.
24. Sprague, E.R., Redd, M.J., Johnson, A.D., and Wolberger, C. (2000). Structure of the C-terminal domain of Tup1, a corepressor of transcription in yeast. EMBO J. *19*, 3016–3027.
25. Carr, S., Penfold, C.N., Bamford, V., James, R., and Hemmings, A.M. (2000). The structure of TolB, an essential component of the tol-dependent translocation system, and its protein-protein interaction with the translocation domain of colicin E9. Struct. *8*, 57–66.
26. Murzin, A.G. (1992). Structural principles for the propeller assembly of β-sheets: the preference for seven-fold symmetry. Proteins *14*, 191–201.
27. Hemmingsen, J.M., Gernert, K.M., Richardson, J.S., and Richardson, D.C. (1994). The tyrosine corner: a feature of most Greek key beta-barrel proteins. Protein Sci. *3*, 1927–1937.
28. Beisel, H.G., Kawabata, S., Iwanaga, S., Huber, R., and 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.
29. Brown, K., Tegoni, M., Prudencio, M., Pereira, A.S., Besson, S., Moura, J.J., Moura, I., and Cambillau, C. (2000). A novel type of catalytic copper cluster in nitrous oxide reductase. Nat. Struct. Biol. *7*, 191–195.
30. Chen, L., Doi, M., Durley, R.C., Chistoserdov, A.Y., Lidstrom, M.E., Davidson, V.L., and Mathews, F.S. (1998). Refined crystal structure of methylamine dehydrogenase from Paracoccus denitrificans at 1.75 A resolution. J. Mol. Biol. *276*, 131–149.
31. Wang, J.-h., and Springer, T.A. (1998). Structural specializations of immunoglobulin superfamily members for adhesion to integrins and viruses. Immunol. Rev. *163*, 197–215.
32. Murzin, A.G., Brenner, S.E., Hubbard, T., and Chothia, C. (1995). SCOP: a structural classification of proteins database for the investigation of sequences and structures. J. Mol. Biol. *247*, 536–540.
33. Holm, L., and Sander, C. (1993). Protein structure comparison by alignment of distance matrices. J. Mol. Biol. *233*, 123–138.
34. Kelley, L.A., MacCallum, R.M., and Sternberg, M.J. (2000). Enhanced genome annotation using structural profiles in the program 3D-PSSM. J. Mol. Biol. *299*, 499–520.
35. Heffron, S., Moe, G.R., Sieber, V., Mengaud, J., Cossart, P., Vitali, J., and Jurnak, F. (1998). Sequence profile of the parallel beta helix in the pectate lyase superfamily. J. Struct. Biol. *122*, 223–235.
36. Brockl, G., Behr, M., Fabry, S., Hensel, R., Kaudewitz, H., Biendl, E., and Konig, H. (1991). Analysis and nucleotide sequence of the genes encoding the surface-layer glycoproteins of the hyperthermophilic methanogens *Methanothermus fervidus* and *Methothermus sociabilis*. Eur. J. Biochem. *199*, 147–152.
37. Bradley, P., Cowen, L., Menke, M., King, J., and Berger, B. (2001). BETAWRAP: successful prediction of parallel β-helices from primary sequence reveals an association with many microbial pathogens. Proc. Natl. Acad. Sci. USA *98*, 14819–14824.
38. Ciccarelli, F.D., Copley, R.R., Doerks, T., Russell, R.B., and Bork, P. (2002). CASH—a beta-helix domain widespread among carbohydrate-binding proteins. Trends Biochem. Sci. *27*, 59–62.
39. Ponting, C.P., Schultz, J., Copley, R.R., Andrade, M.A., and Bork, P. (2000). Evolution of domain families. Adv. Protein Chem. *54*, 185–244.
40. Hanaoka, K., Qian, F., Boletta, A., Bhunia, A.K., Piontek, K., Tsiokas, L., Sukhatme, V.P., Guggino, W.B., and Germino, G.G. (2000). Co-assembly of polycystin-1 and -2 produces unique cation-permeable currents. Nature *408*, 990–994.
41. Ibraghimov-Beskrovnaya, O., Bukanov, N.O., Donohue, L.C., Dackowski, W.R., Klinger, K.W., and Landes, G.M. (2000). Strong homophilic interactions of the Ig-like domains of polycystin-1, the protein product of an autosomal dominant polycystic kidney disease gene, PKD1. Hum. Mol. Genet. *9*, 1641–1649.
42. Rost, B. (1996). PHD: predicting one-dimensional protein structure by profile based neural networks. Methods Enzymol. *266*, 525–539.
43. Jones, D.T. (1999). Protein secondary structure prediction based on position-specific scoring matrices. J. Mol. Biol. *292*, 195–202.
44. Gotoh, O. (1996). Significant improvement in accuracy of multiple protein sequence alignments by iterative refinement as assessed by reference to structural alignments. J. Mol. Biol. *264*, 823–838.
45. Doublie, S. (1997). Preparation of selenomethionyl proteins for phase determination. Methods Enzymol. *276*, 523–530.
46. Brunger, 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., et al. (1998). Crystallography and NMR system: a new software suite for macromolecular structure determination. Acta Crystallogr. D Biol. Crystallogr. *54*, 905–921.
47. Jones, T.A., Zou, J.Y., Cowan, S.W., and Kjeldgaard, M. (1991). Improved methods for binding protein models in electron density maps and the location of errors in these models. Acta Crystallogr. A *47*, 110–119.
48. Russell, R.B., and Barton, G.J. (1992). Multiple protein sequence alignment from tertiary structure comparison: assignment of global and residue confidence levels. Proteins *14*, 309–323.
49. Carson, M. (1997). Ribbons. Methods Enzymol. *277*, 493–505.
50. Nicholls, A., Sharp, K.A., and Honig, B. (1991). Protein folding and association; insights from the interfacial and thermodynamic properties of hydrocarbons. Proteins *11*, 281–296.
51. Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., and Higgins, D.G. (1997). The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. *25*, 4876–4882.
52. Lietzke, S.E., Scavetta, R.D., Yoder, M.D., and Jurnak, F. (1996). The refined three-dimensional structure of pectate lyase E from *Erwinia chrysanthemi* at 2.2 angstrom resolution. Plant Physiol. *111*, 73–92.
53. Pickersgill, R., Smith, D., Worboys, K., and Jenkins, J. (1998). Crystal structure of polygalacturonase from *Erwinia carotovora* ssp. carotovora. J. Biol. Chem. *273*, 24660–24664.
54. Springer, T.A., Jing, H., and Takagi, J. (2000). A novel Ca^{2+}-binding β-hairpin loop better resembles integrin sequence motifs than the EF-hand. Cell *102*, 275–277.
55. Bork, P., and Bairoch, A. (1995). Extracellular protein modules: a proposed nomenclature. Trends Biochem. Sci. Suppl. *2*.

Accession Numbers

The coordinates and reflection data have been deposited in the Protein Data Bank under accession code 1L0Q.