Structure of the *Escherichia coli* TolB protein determined by MAD methods at 1.95 Å resolution

Chantal Abergel[1]*, Emmanuelle Bouveret[2], Jean-Michel Claverie[1], Kieron Brown[3], Alain Rigal[2], Claude Lazdunski[2] and Hélène Bénédetti[2†]

Background: The periplasmic protein TolB from *Escherichia coli* is part of the Tol–PAL (peptidoglycan-associated lipoprotein) multiprotein complex used by group A colicins to penetrate and kill cells. TolB homologues are found in many Gram-negative bacteria and the Tol–PAL system is thought to play a role in bacterial envelope integrity. TolB is required for lethal infection by *Salmonella typhimurium* in mice.

Results: The crystal structure of the selenomethionine-substituted TolB protein from *E. coli* was solved using multiwavelength anomalous dispersion methods and refined to 1.95 Å. TolB has a two-domain structure. The N-terminal domain consists of two α helices, a five-stranded β-sheet floor and a long loop at the back of this floor. The C-terminal domain is a six-bladed β propeller. The small, possibly mobile, contact area (430 Å^2) between the two domains involves residues from the two helices and the first and sixth blades of the β propeller. All available genomic sequences were used to identify new TolB homologues in Gram-negative bacteria. The TolB structure was then interpreted using the observed conservation pattern.

Conclusions: The TolB β-propeller C-terminal domain exhibits sequence similarities to numerous members of the prolyl oligopeptidase family and, to a lesser extent, to class B metallo-β-lactamases. The α/β N-terminal domain shares a structural similarity with the C-terminal domain of transfer RNA ligases. We suggest that the TolB protein might be part of a multiprotein complex involved in the recycling of peptidoglycan or in its covalent linking with lipoproteins.

Addresses: [1]Information Génétique et Structurale, CNRS-UMR 1889, [2]Laboratoire d'Ingénierie et de Dynamique des Systèmes Membranaires, CNRS-UPR 9027 and [3]Architecture et Fonction des Macromolécules Biologiques, CNRS-UPR 9039, Institut de Biologie Structurale et Microbiologie 31 Chemin Joseph Aiguier, Marseille, 13402 Cedex 20, France.

[†]Present address: Centre de Biophysique Moléculaire, CNRS-UPR 4301, rue Charles-Sadron, 45071 Orléans Cedex 2, France.

*Corresponding author.
E-mail: chantal@igs.cnrs-mrs.fr

Key words: crystal structure, sequence comparison, TolB

Received: 12 July 1999
Revisions requested: 9 August 1999
Revisions received: 26 August 1999
Accepted: 16 September 1999

Published: 1 October 1999

Structure October 1999, 7:1291–1300
http://biomednet.com/elecref/0969212600701291

0969-2126/99/$ – see front matter

Introduction

The cell envelope of Gram-negative bacteria includes the inner membrane, the outer membrane, and the murein (peptidoglycan) layer that lies within the periplasmic space. The outer membrane forms a strong permeability barrier that shields bacteria like *Escherichia coli* from the detergent action of bile salts and degradation by digestive enzymes while allowing nutrients of less than 600 Da to diffuse into the periplasmic space through protein pores. The three dimensional (3D) structures of several outer-membrane proteins have been determined [1]. The murein layer interacts with many outer-membrane proteins, including some lipoproteins, as well as with inner-membrane proteins [2]. The complex organization and assembly mechanisms of the cell envelopes of Gram-negative bacteria are still poorly understood.

The Tol–PAL (peptidoglycan-associated lipoprotein) system consists of several proteins encoded by two operons located at 17 minutes on the chromosomal map of *E. coli* [3,4]. These proteins form two complexes. One is located in the cytoplasmic membrane and involves the TolA, TolQ and TolR proteins interacting with each other by their transmembrane segments [5–8]. The other complex is composed of the interacting proteins TolB (periplasmic) and PAL (outer membrane) and is associated with the outer membrane [9,10]. The Tol–PAL complex seems to play a role in cell-envelope integrity. Mutations in the corresponding genes make the cells hypersensitive to drugs and detergent, leaky for periplasmic proteins [3,11], and result in the formation of outer-membrane vesicles [12]. This latter phenotype suggests a defect in cell-envelope assembly similar to that observed in *lpp* mutants [13].

The precise function of the Tol–PAL system remains unknown; it must, however, play an important role in the biosynthesis and/or structure of Gram-negative bacterial envelopes. The system has also been characterized in *Haemophilus influenzae*, *Brucella abortus*, *Pseudomonas aeruginosa*, *Pseudomonas putida*, and *Actinobacillus pleuropneumoniae* (for a review see [11]). Using all available nucleotide sequences from the numerous ongoing bacterial genome projects we found additional TolB-like proteins (i.e. homologous to both the N- and C-terminal domains) in

proteobacteria of the gamma, beta, and alpha subdivisions (see the Materials and methods section).

Recently, it has been shown that TolB interacts with Lpp and OmpA, and that PAL also interacts with OmpA [14]. It therefore seems that the TolB and PAL proteins may participate in a larger complex involved in anchoring the outer membrane to the peptidoglycan. In this regard, the interaction of PAL with the peptidoglycan is certainly crucial for the function of the Tol–PAL system. Finally, TolB interacts with trimeric porins of the outer membrane, as is the case with TolA [11].

The TolQRAB proteins are required for the uptake of group A colicins into the cell, and the TolQRA proteins are necessary for the entry of filamentous phage DNA into the cytoplasm [3,11]. The C-terminal domain of TolA interacts with colicins A and El [11] and with g3p [15]. It has been shown that TolB can interact with the N-terminal domains of colicin E3 and of colicin A [11]. Moreover, a short region called the 'TolB box', which is required for this interaction, has been determined in these colicins and may be present in each TolB-dependent group A colicin [11]. This interaction plays a critical function in colicin uptake because when the interaction is prevented by a point mutation in the TolB box, colicin uptake no longer occurs [11]. Therefore TolB, which is located in the periplasmic space, occupies a strategic position that enables it to interact with many different partners: PAL, OmpA, Lpp, porins and the N-terminal translocation domains of colicins E3 and A. Finally, TolB has recently been shown to be required for the lethal infection of mice by *Salmonella typhimurium* [16].

TolB is produced in a precursor form with a 21-residue signal peptide that is cleaved upon export to the periplasm [10]. Some TolB protein always fractionates with the outer membrane [10], which was explained by the interaction of TolB with PAL [9]. The mature TolB protein consists of 409 residues and comprises several tandem-repeat regions [17] predicted to form a β-propeller structure [18].

The 3D structure of TolB has been solved to help investigate the function of the protein and to understand its interaction with numerous partners. The TolB protein has a distinct two-domain structure, an α/β N-terminal domain (residues 1–145), and a large C-terminal domain with a six-bladed β-propeller structure (Figures 1a,b). The amino acid conservation pattern among the available TolB sequences was analyzed in the context of the 3D structure in order to predict the regions involved in the many interactions described above. The structural and sequence similarity of the TolB C-terminal domain with the β-propeller domain of prolyl oligopeptidases [19] and a motif shared with class B metallo-β-lactamases [20] led us to suggest that TolB is involved in the recycling of peptidoglycan or its covalent linking with lipoproteins.

Results and discussion

Overall structure

Figure 1b presents a ribbon diagram of the two-domain architecture of the protein. The N-terminal α/β domain consists of a central β-sheet floor flanked on one side by two helices and, on the other side, by a 'flap-like' loop centred on a tryptophan residue (W70). The second domain corresponds to a canonical β-propeller fold. It consists of six β sheets (numbered I–VI in Figures 1a,b), each of which is made of four antiparallel β strands running from the inside to the outside, that are arranged in a propeller-like fashion around an approximate sixfold axis. This organization is common to β-propeller-containing proteins, which often exhibit clearly separate N-terminal (or C-terminal) catalytic or functional domains [19,21–24]. A nine-residue linker (T136–T144) joins the two domains.

The N-terminal domain (D1)

The N-terminal domain consists of five β strands forming two β sheets, an 11-residue helix (α1) and a 16-residue amphipathic helix (α2). The first β sheet is made of two parallel strands (β1, β2) and the second comprises three long antiparallel β strands (β3–β5). These two β sheets form a hydrophobic floor sandwiched between the two helices and a long loop centered on a tryptophan residue (W70) that forms a flap.

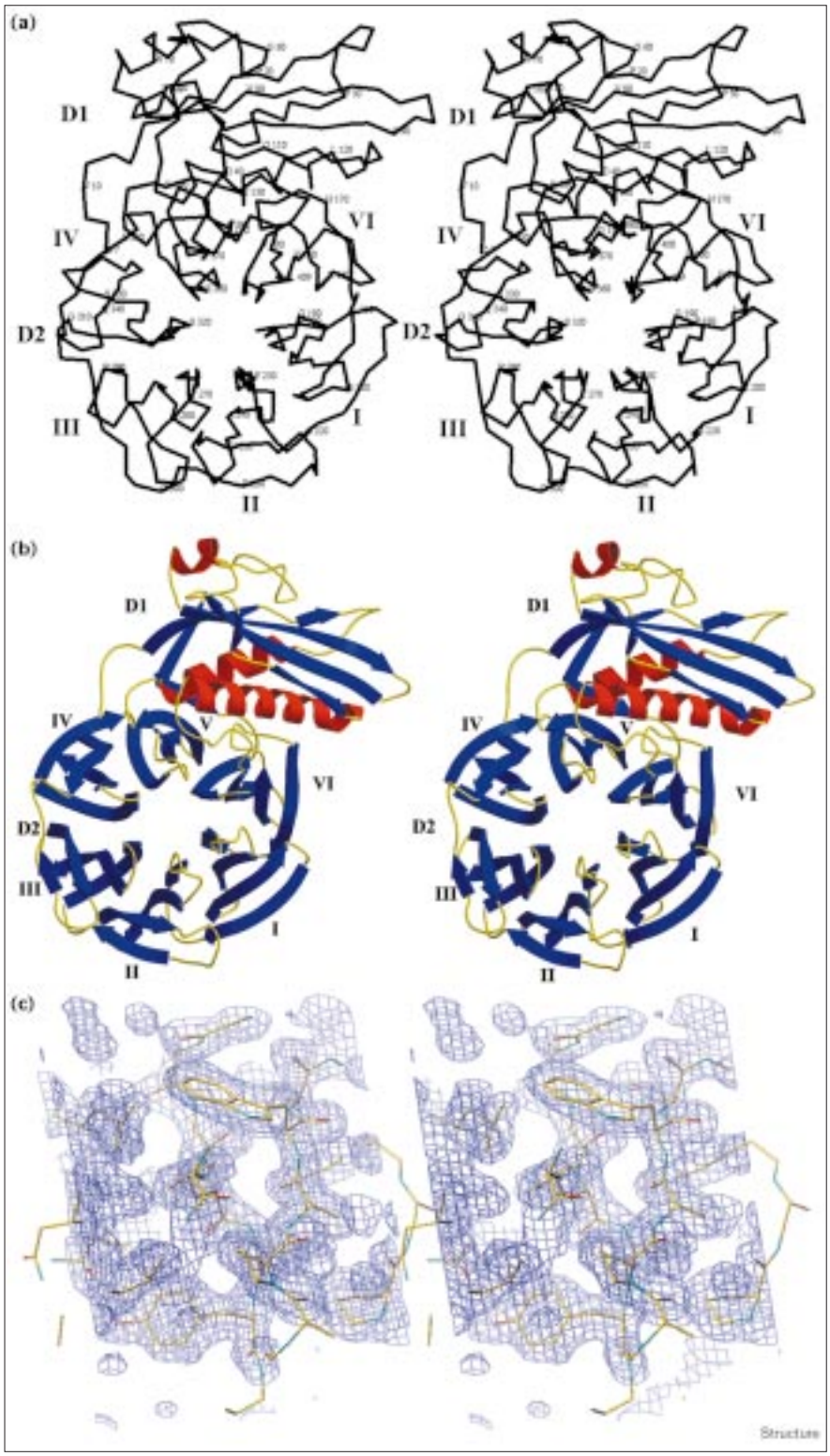
Structural homologues of this domain were searched for using the program DALI [25]. Surprisingly, the best structural matches in the Protein Data Bank (PDB) [26] are the C-terminal domains of three related transfer RNA (tRNA) ligases (glycyl-tRNA synthetase: 1ati; histidyl-tRNA synthetase: 1adj; threonyl-tRNA synthetase: 1qf6). The TolB N-terminal domain has the same topology as the C-terminal domains of class II synthetases, and superimposes with a root mean square deviation (rmsd) of 2.8–2.9 Å for a segment of 76–82 residues (with 14–17% identical amino acids). This C-terminal domain is thought to be the tRNA-binding domain in synthetases [27]. The next best DALI match (with the same topology) is a porphobilinogen deaminase (1pda) with a 2.7 Å rmsd over a 74-residue segment, but only 8% of amino acids are identical in the two protein segments.

A multiple alignment of representative TolB sequences (Figure 2a) reveals 20 (out of 140) strictly conserved positions in proteobacteria from the gamma, beta, and alpha subdivisions. These positions are mostly involved in hydrophobic interactions and hydrogen bonds in the D1 domain or between the N-terminal and β-propeller domains.

Contact area between the N-terminal and β propeller domains

The buried surface area between the N-terminal and β-propeller domains was estimated from the difference in solvent accessibility computed for the entire molecule and

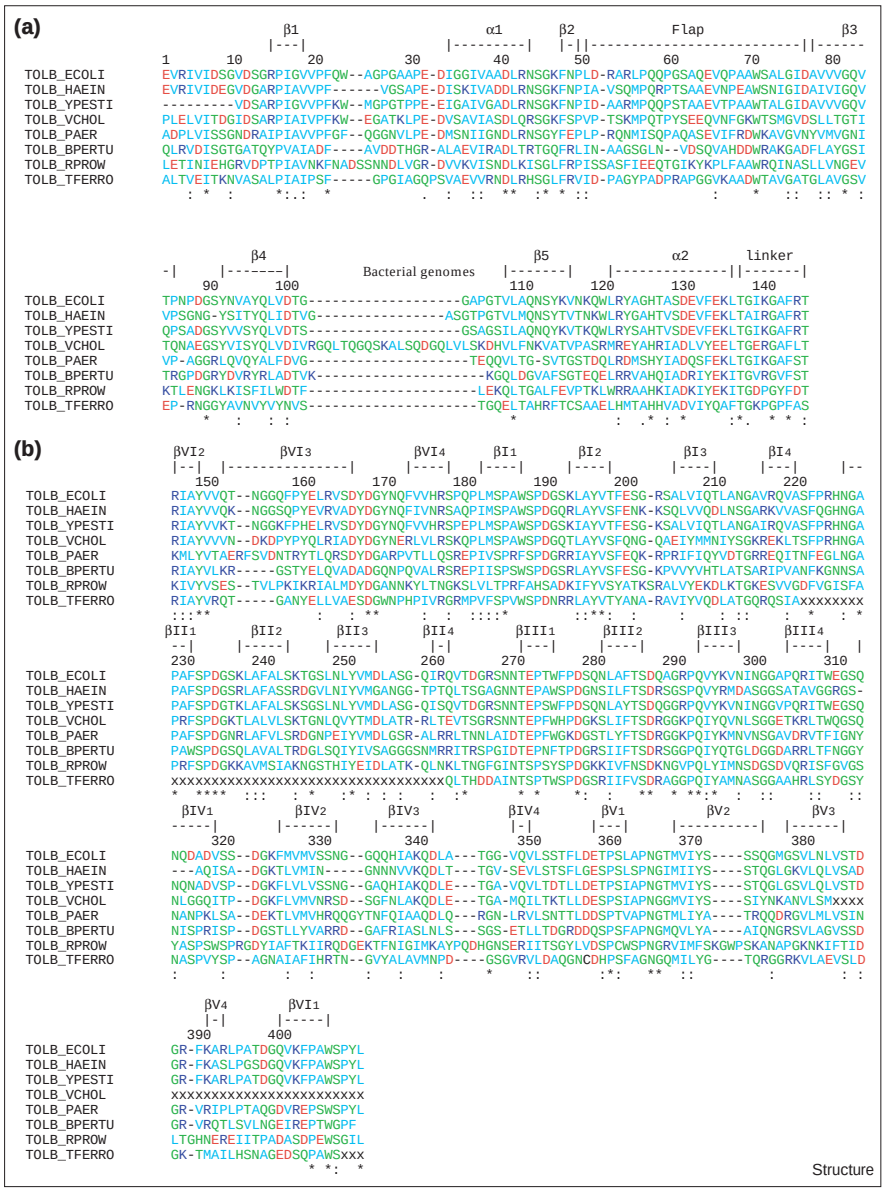
Figure 1



Stereoviews of the TolB protein structure. (a) Stereoview of a the Cα trace of the TolB protein structure (residues 7–409). The figure was prepared using TURBO–FRODO [55]. (b) Stereoview of a ribbon diagram of the TolB protein structure (residues 7–409). The blue arrows represent β strands, the red ribbons represent α helices and turns and loops are coloured in yellow. The figure was prepared using MOLSCRIPT [60]. (c) Stereoview of the electron-density map of the central cavity region of the β-propeller structure of TolB. The $2F_o–F_c$ electron-density map is contoured at 1.5σ. Atoms are coloured by type: carbon, yellow; oxygen, red; nitrogen, cyan. The figure was prepared using TURBO-FRODO [55].

for each domain taken separately. The total contact area between the two domains is ~430 Å^2. The interaction involves a nine-residue segment (T136–T144) making contacts between the amphipathic helix of the N-terminal

Figure 2



Multiple alignment and secondary-structure assignment of nine bacterial TolB sequences in complete and incomplete genomes. The *Rickettsia prowazekii* sequence is from [61]. Negatively charged residues are coloured in red, positively charged residues are blue, polar residues are green and hydrophobic residues are cyan. Secondary-structure elements are marked by dashed lines (β for β strands and H for helices). Numbers indicate every tenth residues on the TolB *E. coli* sequence. The D1 domain (a) and the D2 domain (b) are shown.

domain (R121-Y122-H125-D129) and blade VI of the β propeller (D165-Y166-D167-Y169-N170-P359-S406-L409). Most of these positions are strictly conserved throughout our panel of sequences (Figure 2). The histidine H125 is hydrogen bonded with two conserved aspartic acid residues, one from the second helix (D40) and the other from blade VI of the β propeller (D167). The last α1 residue (N43) and the following turn (S44–K46) are in contact with blade V of the β propeller (N364-T366-M367) and correspond with a well-conserved sequence segment. The ten other α1 residues delineate, with the propeller edge (blades V and VI), a cavity filled with 35 solvent molecules. The rather small interface area (430 Å^2) is unlikely to be sufficient to maintain a very stable interaction, and might therefore allow a conformational change upon binding with another macromolecule [28].

The β-propeller domain (D2)

The β-propeller domain, composed of six repeated motifs (Figures 1a,b), has a cylindrical external shape (approximately 34 Å in height × 43 Å in diameter). The six blades, each composed of four-stranded antiparallel β sheets, are twisted and radially arranged around a central channel. The central cavity is not cylindrical over its entire length, but assumes a funnel-like shape. The cylindrical section of the channel (16 Å long) has a 9–10 Å diameter on the N-terminal side and narrows to a collar of 7–8 Å in diameter before widening to 40 Å in its conical section.

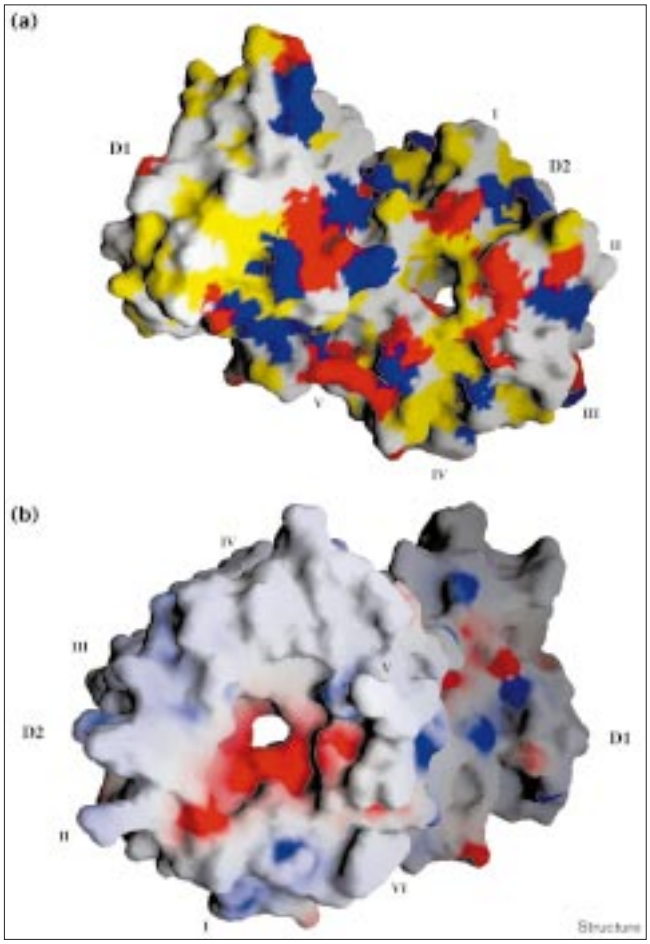
Except for blade III, the blades of the β-propeller domain are built up regularly, with the polypeptide chain progressing outwards from the central axis in four-stranded antiparallel β sheets. A variation is seen in blade III, where the fourth strand does not assume a canonical extended-strand geometry.

The overall symmetry of the β propeller was assessed by superimposing the six blades. The rmsd values range from 1.1 Å (blades IV and VI on I) to 1.8 Å (blades III and V on I). Blades I–V are joined in succession around the central pseudo-sixfold axis. In blade VI, the ring closure ('velcro') is achieved by forming the fourth antiparallel β sheet from both termini of the β-propeller domain: the C terminus provides the innermost strand, which is bonded (eight hydrogen bonds) to the three antiparallel β sheets from the N terminus (Figure 1a,b).

In contrast to other β-propeller domains, in the TolB β-propeller domain the six-residue sequences flanking the central channel (innermost strands) are very similar in five of the six blades (see below). On each blade, the last exposed residue before entering the channel is a proline residue (P188, P232, P276, P363, P407), with the exception of the fourth blade where the last exposed residue is a serine (S319). On the other side, the penultimate amino acid before the funnel widens is again a proline residue (P184, P228, P272, P359, P403), except in the fourth atypical blade where this residue is an alanine (A316). The two extremities of the β-propeller central channel are therefore circled by two proline rings. The distances across the channel, between the residues facing each other, are 9.27 Å for P_I–S_{IV}, 10.61 Å for P_{II}–P_V and 12.13 Å for P_{III}–P_{VI}. In addition, aromatic residues constitute a median ring around the β-propeller channel (W186, F230, W274, W405 on strand 1; Y148, Y195, Y370, on strand 2 and Y250, Y294 on strand 3). Most of the amino acids making up these three rings are evolutionarily conserved (Figure 2b) and seem to have a stabilizing role in the TolB structure. For instance, the sidechains (highly ordered, see Figure 1c) of the conserved aromatic residues assume the same orientation and point away from the channel to fill each blade concavity. In the central cavity, the residue sidechains are pointing towards the next blade or outwards (away) from the central cavity, giving the central channel a circular and rigid aspect.

The accessible surface computed by GRASP [29] reveals a peculiarity of TolB (Figure 3); throughout its entire length, the central channel area can be clearly divided into a hydrophobic side (encompassing two-thirds of the area) and a negatively charged (D315–D317) side (Figure 3a). As a result, the channel is only partially filled by ordered water molecules (35 can be detected). As can be seen in Figure 3b, a strongly electronegative electrostatic potential characterizes both the wide entrance and the interior

Figure 3



Molecular surfaces of the TolB protein structure. **(a)** Surface representation of the TolB structure colour-coded according to the surface property: dark blue corresponds to positively charged surfaces, red to negatively charged surfaces and yellow to hydrophobic surfaces. The figure was generated with the program GRASP [29]. **(b)** Solvent-accessible surface coloured according to electrostatic potential using GRASP defaults. The electrostatic potential is contoured in the range from $-10k_BT$ (red) to $+10k_BT$ (blue). The figure is rotated by 180° relative to (a).

of the TolB β-propeller channel. This potential might serve to funnel a positively charged compound through the channel.

Finally, we noticed a pseudo catalytic triad on the edge of the β propeller, (H336, βIV3; S351 and D356, turn IV4–V1). However, in the crystal structure these residues are not in a canonical conformation. Furthermore, they are not conserved among TolB sequences (Figure 2b).

Analysis of known mutations

The previously described TolB mutants [14,30] can now be analyzed in the context of the 3D structure. Expression of TolB mutants with a stop codon at positions V150

(βVI_2) or T307 (βIII_4) cannot be detected *in vivo*. The same phenotype was observed after the deletion of the D88–P104 segment (forming the β3–β4 turn, β4, and the β4–β5 turn in D1). The S384F mutation (βV3) also resulted in undetectable TolB expression. This suggests that a phenylalanine residue could not be accommodated at this location without interfering with correct folding. Finally, the null phenotype is also observed with the H125D mutation, suggesting that this residue is also critical to the folding of the TolB protein. Indeed, this conserved position is at the centre of a number of interactions between the β propeller (D2), the linker and the D1 domain.

The insertion of two residues after T136 (EL) produced a detectable protein exhibiting a loss of interaction with PAL and to a lesser extent with OmpA [14]. The A227V mutation produced a protein unable to interact with OmpA and Lpp, while the interaction with colicin was maintained. Thus, the OmpA and Lpp proteins might interact with the TolB protein via the β-propeller domain, whereas the colicin protein seems to interact with the D1 domain.

Inserting two residues (SS) after P87 (in β3–β4 turn) did not seem to disturb the TolB–PAL interaction, while the colicin protein could not penetrate so efficiently into the bacteria. This again suggests an interaction between the D1 domain and the colicin. A severely truncated form of the TolB protein (residues 1–363, thus lacking most of blade V and VI) still interacts with colicins [31]. Moreover, the addition of a 14-residue peptide after residue 363 did not prevent this interaction.

Sequence analysis

The *H. influenzae* TolB is the only homologue of *E. coli* TolB recorded in the Swiss-Prot database [32]. The two sequences have ~60% residue identity over their entire length. We further identified TolB homologues in seven other bacterial species of the gamma subdivision (*Salmonella typhi*, *Yersinia pestis*, *Actinobacillus actinomycetemcomitans*, *Vibrio cholerae*, *P. aeruginosa*, *Shewanella putrefaciens* and *Thiobacillus ferrooxidans*), one in the beta subdivision (*Bordetella pertussis*) and two in the alpha subdivision (*R. prowazekii* and *Caulobacter crescentus*). A multiple alignment of the most informative TolB sequences is presented in Figure 2. A putative *T. ferrooxidans* TolB homologue is included, even though it lacks a stretch of 40 residues, to display the amino acid changes required to adapt the protein to an acidic medium (pH ~2). Three charged residues in the D1 domain of TolB are conserved throughout the bacterial sequences, except for in *T. ferrooxidans* where they are replaced by polar ones: D99→N, R121→Q and E133→Q. These residues are hydrogen bonded with residues in the β propeller (VI_3–VI_4) and D1 (β4–β5, α1) domains.

Even though the N-terminal D1 domain is more conserved than the C-terminal β-propeller domain, sequence analyses could not suggest a possible function for this domain. β-Propeller domains do not often exhibit significant primary sequence similarity, suggesting that the overall shape of the molecule is not affected by amino acid changes. As predicted previously [18] the TolB β propeller is comprised of six repeating sequence units of 43–46 residues: 145–190, 191–234, 235–278, 279–322, 323–365, 366–409. Each repeat does not constitute a separate blade; it contributes three strands to one blade and the innermost strand to the next blade (Figures 1b,2b). A distinctive feature in the TolB repeats is the strong conservation of the last seven residues in each repeat, according to the prototype: P[AS][WF][AS]PDG, where the residues in between the two prolines constitute the wall of the central channel. Such short sequences seem interchangeable, and are shuffled around in different bacteria. Amidst the variability of the rest of the repeat unit, this conservation suggests that there is a strong evolutionary pressure on the geometry and accessibility to the central channel, possibly to accommodate a specific ligand. A noticeable exception is repeat 4, which exhibits a much higher variability and in which the central aromatic residue is often replaced by a large hydrophobic [LIV] one.

In a first approach to identifying remote functional homologies to the TolB β-propeller domain, we derived a minimal TolB-specific repeat signature from the most conserved position in the successive TolB repeats of *E. coli*, and other gamma-subdivision bacteria (Figure 2b): P<.3>P[DN][GS]<.2>[LVI].[WFY]<.3>[EQRK]<.5><.?2>[LIV], where [ABC] indicates a choice of residue, a dot '.' indicates any amino acid (fixed place holder) and <.?n> represents the possibility of 0–n optional unspecified residues (gap). This signature has a probability of random occurrence (per position) of 7.10^{-8}. The regular expression was then used to scan the nonredundant (nr) database (www.ncbi.nlm.nih.gov; 368,976 sequences, for a total of 108,724,348 positions); ~seven random matches were expected. The signature was found 31 times in 19 different sequences, including eight TolB-related sequences, three class B β-lactamases [20], a predicted acylaminoacyl-peptidase from *Bacillus subtilis* (Embl: CAB15213), one dipeptidyl peptidase, three hypothetical/unknown proteins, and two neurotoxins. Given the periplasmic location of the TolB protein, and its relationship with peptidoglycan, the matches with the β-lactamases (Swiss-Prot: P26918, and two close homologues from different *Aeromonas* species) caught our attention. Moreover, the match with the TolB signature overlapped with two essential residues of the class B β-lactamase active site [20]. Class A, C and D β-lactamases are active-site serine enzymes believed to have evolved from cell-wall synthetic DD-peptidases without retaining significant sequence similarity [33–35]. The class B β-lactamases are periplasmic

metalloenzymes of unknown evolutionary origin, with no relationship to the β-propeller fold. All β-lactamases share the ability to hydrolyze β-lactam compounds (such as penicillin), which have structural homology to the D-alanyl-D-alanine moiety. β-lactam antibiotics specifically block the transpeptidase reaction necessary for the cross-linking of peptidoglycan [36]. The TolB repeat signature (corresponding to blades I–III, and V) might thus correspond to the sequence determinant required for the recognition of the molecular intermediate (e.g. the D-alanyl-D-alanine moiety) of the peptidoglycan transpeptidase reaction.

To expand on this finding, we used Phi-Blast [37] to evaluate the overall sequence similarity of TolB-signature-containing target proteins and the TolB sequence. Again, the most significant match, besides known TolB homologues, was found with the predicted acylaminoacyl-peptidase from *B. subtilis* (EMBL: CAB15213) (p value < 3.10^{-9}). This protein most probably consists of a seven-bladed β-propeller domain associated with an acylaminoacyl-peptidase C-terminal catalytic domain.

The TolB potential relationship with protein of the prolyl oligopeptidase family [19] is further supported by the result of a standard similarity search [38], where again the sole significant matches (although with low E values of ~0.003) with the β-propeller domain, corresponded to various acylaminoacyl peptidases and dipeptidyl peptidases. For instance, a 50% identity match extending over the entire blade II sequence was found with a prolyl-dipeptidyl peptidase (PIR: S66261, from the green sulphur bacteria *Flavobacterium meningosepticum*).

Potential interaction sites

The β-propeller fold adopted by the TolB C-terminal domain brings about many possible association sites with its known interactors: PAL, OmpA, Lpp, porins and colicins. Structural domains (e.g. blades) or specific residues involved in making contact with other proteins are expected to be more conserved. Interaction with the N-terminal domain involves a stretch of residues protruding from the β propeller (P363–T366, V_1–V_2; T385–F389, V_3–V_4; S406–L409, VI_1–C-terminus; Y166–Q171, VI_3–VI_4). These residues are conserved throughout the gamma-subdivision sequences. Sequence conservation suggests other potential interacting areas. The first one involves D1 and D2 together in the cavity between the two domains (Y166–Q171, VI_3–VI_4; L393–G398, V_4–VI_1). Around the β-propeller channel, on the side of the D1 domain but away from the two domains interaction area, both turns and loops are conserved and might be potential interactors for proteins or complex chemical structures. These conserved positions in turns and loops of the D1 domain, involve blades I and II (S187–S191, I_1–I_2; L210–G213, I_3–I_4; S231–S235, II_1–II_2; L254–S256, II_3–II_4;), and blades III and IV (P276–S279, III_1–III_2; N299–G301, III_3–III_4; D321–K323, IV_1–IV_2). On the other side of the β propeller, exposed residues are conserved in blades VI–I, VI–II, and II (P178–P180, VI_4–I_1; F198–S201, VI_4–II_1; K243, II_2–II_3; S220–N225, II_1–II_3), as well as in blades II–III and III (G265–N268, II_4–III_1; D286–291, III_2–III_3). These conserved residues in turns and loops constitute prime candidate sites for macromolecular interaction.

Biological implications

The 3D structure of the TolB protein reveals a β-propeller fold associated with an N-terminal α/β domain. Although TolB D1 domain exhibits no significant sequence similarity to a characterized protein, two types of evidence establish a link between the TolB C-terminal domain D2 and proteins of the oligopeptidase family: significant local sequence similarities with individual acylaminoacyl peptidases and dipeptidyl peptidases; and a β-propeller fold. TolB might, therefore, be involved in the recognition of peptidic structures. In addition, a sequence similarity between the central channel of D2 and the active site of class B β-lactamases suggests that TolB might interact with the D-alanyl-D-alanine moiety. Interestingly, this compound is central to the biosynthesis and recycling of peptidoglycan, and to its ligation with lipoproteins. However, the TolB D2 β-propeller domain seems to lack the flexibility required to perform a catalytic function [19]. This role might be performed by the N-terminal D1 domain, or by another protein of the Tol–PAL system.

Recent experimental evidence (data not shown) indicates that the TolB protein might directly interact with peptidoglycan. Yet it is unlikely that a direct implication of the Tol–PAL system in the biosynthesis of peptidoglycan — a well known pathway — would have been overlooked [39]. However, the TolB protein might still be involved in less well understood pathways, such as those responsible for the recycling of peptidoglycan (involving a number of unknown periplasmic transglycosylases, endopeptidases and L,D-carboxypeptidases) [40], or the assembly of murein lipoproteins [41].

It did not escape our attention that $TolB^-$ and Lpp^- mutants have very similar 'vesicle-forming' phenotypes [12,13]. TolB might, therefore, be part of a complex carrying the elusive murein lipoprotein–peptidoglycan ligase activity, covalently attaching the murein lipoprotein via a D-alanyl-D-alanine transpeptidation followed by peptide-bond formation between the ε-amino group of a C-terminal lysine residue and a carboxyl group of the diaminopimelate of the murein chain [41]. If this were the case, the strong electronegative electrostatic potential converging towards the TolB β-propeller channel (Figure 3b) might play a role in the binding of the C-terminal Tyr-Arg-Lys sequence required for the peptidoglycan–lipoprotein ligation [42].

A similar patch of electronegative potential is found in histone acetyltransferases for which a lysine ε-NH2 is also the final target of the reaction [43]. The 3D structure of the TolB protein will now be used to help design the experiments needed to validate or invalidate the above-mentioned hypotheses.

Materials and methods

Protein preparation and crystallization

The met⁻ strain of *E. coli* B834(DE3) (Novagen; [44]) was transformed with pTolBHis [45]. The produced TolBHis derivative consists of the mature TolB protein fused at its N terminus with six histidines. The B834(DE3) strain allowed the incorporation of selenomethionine residues in place of methionine residues, permitting the macromolecular crystal structure to be determined using the multiwavelength anomalous diffraction (MAD) methods [46]. Cell were grown, harvested and treated as described previously [17]. Protein purification from the cell extract was carried out by affinity chromatography with a cobalt resin (Talon TM, Clontech). The resulting preparation was concentrated and dialyzed against 10 mM Tris buffer, pH 7.5, containing 1 mM TCEP (molecular probe). The concentration of the selenomethionyl TolBHis final preparation was 2 mg/ml^{-1}. The SAmBA software [47] was used to determine the crystallization conditions of the selenomethionine-substituted TolB protein based upon the crystallization conditions of the native protein as a startup [48]. The crystallization conditions were as described earlier except for a higher pH value (7.5 instead of 6.5). The produced TolB protein was analyzed both by mass spectroscopy to confirm the selenomethionine incorporation and by N-terminal sequencing. Three truncated forms of TolB were present and the analysis of dissolved crystals of the selenomethionine-substituted TolB protein revealed a major component corresponding to the sequence of the mature TolB protein.

Table 1

Refinement statistics.

Resolution (Å)	20.0–1.95
Number of observed reflections	360,642
Number of unique reflections	25,351
Used reflections	25,183
Multiplicity	7.2 (2.6)
$<I>/\sigma(I)$	7.9 (2.1)
Completeness (%)	94.5 (84.8)
R_{merge}	6.1 (22.8)
R factor (%)	18.8
R_{free} (10% test set) %	23.8
Rmsd from ideality	
Bonds (Å)	0.009
Angle (°)	1.7
Improper (°)	0.9
Dihedral (°)	25.8
Average B factor (Å^2)	
Protein	20.9
Solvent	31.85

Values in parentheses are for the outer resolution shell. Space group $P2_{(1)}$, a = 63.2 b = 40.5 c = 76.3 β = 110.9.
R factor = $\Sigma_h \big| |F_{obs}(h)| - |F_{calc}(h)| \big| / \Sigma_h |F_{obs}(h)|$ where F_{obs} and F_{calc} are the observed and calculated structure factor amplitudes for the reflection with Miller indices *h* = (h,k,l).

Data collection and processing

All data were collected from a single crystal on BM14 at the ESRF synchrotron (Grenoble, France) with tunable synchrotron radiation in order to perform a MAD experiment. The crystal was soaked for 1 min in a cryoprotectant solution consisting of 8 µl of a 50% glycerol solution and 10 µl of the reservoir solution made up of 0.1 M MES pH 6.5, 15% (w/v) PEG 8000, 0.1 M NaCl and 5% glycerol. The crystal was flash frozen to 105K in a cold nitrogen gas stream and the images were collected on a Mar charge-coupled device (CCD) camera. The X-ray fluorescence and transmission from the crystal were measured as functions of the incident X-ray energy in the vicinity of the Se–K edge. Two energies were chosen near the Se adsorption edge; 12,668 eV (λ = 0.97873) and 12,665 eV (λ = 0.97896) corresponding to the maximum *f″* and the minimum *f′* respectively. A third remote energy was selected for the third (remote I) and the forth (remote II) data set, 14,000 eV (λ = 0.88561), the last one being collected to get higher-resolution data. Data were collected with the crystal in an arbitrary setting after optimizing the oscillation range using STRATEGY [49] and a complete data set was collected before changing the incident photon energy. The limited extent of the fourth data set was due to time constraints, which prevented the completion of the experiment. The diffraction data were indexed and integrated using DENZO [50] from the CCP4 suite [51] and the orientation matrix was then applied to the remaining energies. The four datasets were scaled to the remote dataset using SCALA [51] and structure-factor amplitudes were calculated using TRUNCATE [51]. Statistics of the processed data are listed in Table 1.

MAD phasing

SCALEPACK [50] was also used in order to use the SOLVE program [52]. Phases were calculated on the three wavelengths in the 20 and 2.5 Å resolution range. A single solution was found with five sites and a mean figure of merit of 0.70 for all the data between 20 and 2.5 Å.

The three datasets were scaled to the remote high-resolution dataset (remote II) and the MAD phases given by the SOLVE program were included in this dataset, which was then used for the refinement. Final statistics of the MAD data collection are presented in Table 2. The phases obtained were improved using the solvent-flattening [53] and histogram-matching techniques [54] as implemented in program DM

Table 2

Final statistics of the MAD data collection.

Dataset	λ (Å)	Resolution (Å)	R_{sym} (%)	R_{anom} (%)	I/σ(I)	Completeness (%)	Anomalous completeness (%)	Redundancy
f max	0.9787	2.5	3.2 (8.8)	4.2 (7)	7.6 (1.0)	98.3 (96.6)	90.9 (85.8)	3.1 (2.9)
f min	0.9789	2.5	3.2 (7.7)	3 (6)	11.9 (5.0)	98.5 (96.9)	91 (85.7)	3.1 (2.9)
Remote I	0.8856	2.25	3.6 (9.2)	3.8 (8)	12.2 (4.9)	91.5 (89.8)	79.7 (75.0)	2.9 (2.6)
Remote II	0.8856	1.93	5.2 (21.9)	5.2 (18)	10.3 (2.1)	74.3 (74.3)	57.5 (54.2)	2.6 (2.3)

Values of R_{sym} and R_{anom} in parentheses are for the outer resolution shell. $R_{sym} = \Sigma \big| I - <I> \big| / \Sigma(I)$, where I = observed intensity and <I> = average intensity. $R_{anom} = \Sigma \big| <I^+> - <I^-> \big| / \Sigma (<I^+> + <I^->)$.

and the electron-density maps were readily interpretable and were used to construct the mainchain of the molecules.

Model building and refinement

The C_α mainchain was built using option TPPR in the TURBO-FRODO program [55] and two selenomethionine residues separated by a residue in the TolB sequence (M325, M327) were used as preliminary markers. After interpreting the entire electron-density map, the model consisted of eight fragments (G13–A24; P30–D51; R54–P85; Y91–F198; R202–V218; A227–L241; L247–I298; P303–L409) with sidechains clearly identifiable, allowing us to position them correctly into the TolB sequence. The TPPR were then replaced by the proper residues and their sidechains.

Preliminary refinement was performed using CNS [56] between 20 and 1.95 Å. After a few rounds of minimization including bulk-solvent correction, simulated annealing and B-factor refinement, we were able to connect all fragments with the exception of the N-terminal residues still not visible in the electron-density maps. Water molecules were then included using both REFMAC/ARP [51] and CNS [56]. The water molecules were added to peaks over 3σ in the F_o–F_c map and were assigned unit occupancy. Water molecules with B values greater than 50 were removed. After several iterations of minimization, B-factor refinement, bulk-solvent and anisotropic corrections, and model building, the final R_{work} and R_{free} were 18.8% and 23.8% respectively. The model includes 403 residues and 458 water molecules. No density was observed for residues A24, G25, P222 and G245, and their occupancy was set to zero. The quality of the model is summarized in Table 1. PROCHECK [57] indicates that 88.4% of the residues are located in the most-favourable regions of the Ramachandran plot, and 10.9% in additional allowed regions and 0.6% in generously allowed regions.

Sequence analysis

Similarity searches were performed using the gapped-Blast, Psi-Blast [38], and Phi-Blast [37] programs using default parameters, filtering procedure and scoring matrix, as implemented on the NCBI server (www.ncbi.nlm.nih.gov), or on the Marseille Blast server (igs-server.cnrs-mrs.fr). Multiple alignments of the TolB sequences were generated using ClustalW [58]. The Real/Lookfor package [59] was used to design the optimal TolB repeat signature, and search for it in sequence databases. Additional TolB sequences were retrieved from complete and partial bacterial genomic sequences provided by the Institute for Genomic Research website (http://www.tigr.org) for *C. crescentus, V. cholera, T. ferrooxidans,* and *S. putrefacians,* the Sanger center (http://www.beowulf.org.uk/home.htm) for *Y. pestis*, *S. typhi*, and *B. pertussis*, and the *Pseudomonas* Genome Project (http://www.pseudomonas.com/), the University of Oklahoma's Advanced Center for Genome Technology (http://www.genome.edu) for *A. actynomycetemcomitans*.

Accession numbers

Crystallographic coordinates as well as experimental amplitudes of the TolB structure have been deposited in the PDB with the accession code 1CRZ.

Acknowledgements

We thank Yves Bourne and Christian Cambillau for helpful discussions and advice and Gordon Leonard for invaluable assistance during the data collection and solving process. We thank Françoise Gosse for careful reading of the manuscript.

References

1. Nikaido, H. (1999). Microdermatology: cell surface in the interaction of microbes with the external world. *J. Bacteriol.* **181**, 4-8.
2. Leduc, M., Ishidate, K., Shakibai, N. & Rothfield L. (1992). Interactions of *Escherichia coli* membrane lipoproteins with the murein sacculus. *J. Bacteriol*. **174**, 7982-7988.
3. Webster, R.E. (1991). The *tol* gene products and the import of macromolecules into *Escherichia coli*. *Mol. Microbiol.* **5**, 1005-1011.
4. Vianney, A., Muller, M.M., Clavel, T., Lazzaroni, J.C., Portalier, R. & Webster, R.E. (1996). Characterization of the *tol-pal* region of *Escherichia coli* K-12: translational control of *tolR* expression by TolQ and identification of a new open reading frame downstream of pal encoding a periplasmic protein. *J. Bacteriol.* **178**, 4031-4038.
5. Derouiche, R., Bénédetti, H., Lazzaroni, J.C., Lazdunski, C. & Lloubès, R. (1995). Protein complex within *Escherichia coli* inner membrane – TolA N-terminal domain interacts with TolQ and TolR proteins. *J. Biol. Chem.* **270**, 11078-11084.
6. Lazzaroni, J.C., *et al.,* & Geli, V. (1995). Transmembrane α-helix interactions are required for the functional assembly of the *Escherichia coli* Tol complex. *J. Mol. Biol.* **246**, 1-7.
7. Germon P., Clavel, T., Vianney, A., Portalier, R. & Lazzaroni, J.C. (1998). Mutational analysis of the *Escherichia coli* K-12 TolA N-terminal region and characterization of its TolQ-interacting domain by genetic suppression. *J. Bacteriol.* **180**, 6433-6439.
8. Journet, L., Rigal, A., Lazdunski, C. & Bénédetti, H. (1999). Role of TolR N-terminal, central, and C-terminal domains in its dimerization and interaction with TolA and TolQ. *J. Bacteriol.* **181**, 4476-4484.
9. Bouveret, E., Derouiche, R., Rigal, A., Lloubès, R., Lazdunski, C. & Bénédetti, H. (1995). Peptidoglycan-associated lipoprotein–TolB interaction. *J. Biol. Chem*. **270**, 11071-11077.
10. Isnard, M., Rigal, A., Lazzaroni, J.C., Lazdunski, C. & Lloubès R. (1994). Maturation and localization of the TolB protein required for colicin import. *J. Bacteriol.* **176**, 6392-6396.
11. Lazdunski, C.J., Bouveret, E., Rigal, A., Journet, L., Lloubes, R. & Bénédetti, H. (1998). Colicin import into *E. coli* cells. *J. Bacteriol.* **180**, 4993-5002.
12. Bernadac, A., Gavioli, M., Lazzaroni, J.C., Raina, S. & Lloubès, R. (1998). *Escherichia coli* tol-pal mutants form outer membrane vesicles. *J. Bacteriol.* **180**, 4872-4878.
13. Fung, T., MacLister, T.J. & Rothfield, L.I. (1978). Role of murein lipoprotein in morphogenesis of the bacterial division septum: phenotype similarity of *ikyd* and *ipo* mutants. *J. Bacteriol.* **133**, 1467-1471.
14. Clavel, T., Germon, P., Vianney, A., Portalier, P. & Lazzaroni, J.-C. (1998). TolB protein of *E. coli* K12 interacts with the outer membrane peptido-glycan associated proteins Pal, Lpp and OmpA. *Mol. Microbiol.* **29**, 359-367.
15. Lubkowski, J., Hennecke, F., Plückthun, A. & Wlodawer, A. (1999). Filamentous phage infection: crystal structure of g3p in complex with its coreceptor, the C-terminal domain of TolA. *Structure* **7**, 711-722.
16. Bowe, F., Lipps, C.J., Tsolis, R.M., Gooisman, E., Heffron, F. & Kusters, J. (1998). At least four per cent of the *Salmonella typhimurium* genome is required for fatal infection of mice. *Infect. Immunol.* **66**, 3372-3377.
17. Bouveret, E. (1998). Mécanisme d'importation des colicines via le complexe Tol-PAL chez *Escherichia coli*. Thèse de Doctorat. Université Aix-Marseille II.
18. Ponting, C.P. & Pallen, M.J. (1999). A β-propeller domain within TolB. *Mol. Microbiol.* **31**, 739-740.
19. Fülöp, V., Böcskei, Z. & Polgar, L. (1998). Prolyl oligopeptidase: an unusual β-propeller domain regulates proteolysis. *Cell* **94**, 161-170.
20. Massidda, O., Rossolini, G.M. & Satta, G. (1991). The *A. hydrophila* cphA gene: molecular heterogeneity among class B metallo-β-lactamases. *J. Bacteriol.* **173**, 4611-4617.
21. Fülöp, 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.
22. Luo, Y., Li, S.C., Chou, M.Y., Li, Y.T. & Luo, M. (1998). The crystal structure of an intramolecular trans-sialidase with a NeuAc alpha2→3Gal specificity. *Structure* **6**, 521-530.
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. Smith, T.F., Gaitatzes, C., Saxena, K. & Neer, E.J. (1999). The WD repeat: a common architecture for diverse function. *Trends Biochem. Sci.* **24**, 181-185.
25. Holm, L. & Sander, C. (1999). Protein folds and families: sequence and structure alignments. *Nucleic Acids Res.* **27**, 244-247.
26. Sussman, J.L., *et al.,* & Abola, E. E. (1998). Protein data bank (PDB): database of three-dimensional structural information of biological macromolecules. *Acta Crystallogr. D* **54**, 1078-1084.
27. Logan, D.T., Mazauric M.-H., Kern, D. & Moras, D. (1995). Crystal structure of glycyl-tRNA synthetase from *Thermus thermophilus*. *EMBO J.* **14**, 4156-4167.

28. Janin, J., Miller, S. & Chothia, C. (1988). Surface, subunit interface and interior of oligomeric protein. *J. Mol. Biol.* 204, 155-164.
29. Nicholls, A., Sharp, K.A. & Honig, B. (1991). Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins* 11, 281-296.
30. Sun, T.P. & Webster, R. (1986). *fii* a bacterial locus required for filamentous phage infection and it's relation to colicin tolerant TolA and TolB. *J. Bacteriol.* 165, 107-115.
31. Bénédetti, H., Lazdunski, C. & Lloubès R. (1991). Protein import into *Escherichia coli:* colicins A and E1 interact with a component of their translocation system. *EMBO J.* 10, 1989-1995.
32. Bairoch, A. & Apweiler, R. (1999). The SWISS-PROT protein sequence data bank and its supplement TrEMBL in 1999. *Nucleic Acids Res.* 27, 49-54.
33. Kelly, J.A., *et al.,* & Ghuysen, J.M. (1986). On the origin of bacterial resistance to penicillin: comparison of a beta-lactamase and a penicillin target. *Science* 231, 1429-1431.
34. Samraoui, B., Sutton, B.J., Todd, R.J., Artymiuk, P.J., Waley, S.G. & Phillips, D.C. (1986). Tertiary structural similarity between a class A β-lactamase and a penicillin-sensitive D-alanyl carboxypeptidase-transpeptidase. *Nature* 320, 378-380.
35. Kelly, J.A. & Kuzin, A.P. (1995). The refined crystallographic structure of a DD-peptidase Penicillin-target enzyme at 1.6 Å resolution. *J. Mol. Biol.* 254, 223-236.
36. Tipper, D.J. & Strominger, J.L. (1965). Mechanism of action of penicillins: a proposal based on their structural similarity to acyl-D-alanyl-D-alanine. *Proc. Natl Acad. Sci. USA* 54, 1133-1141.
37. Zhang, Z., *et al.,* & Altschul, S.F. (1998). Protein sequence similarity searches using patterns as seeds. *Nucleic Acids Res.* 26, 3986-3990.
38. Altschul, S.F., *et al.,* & Lipman, D.J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25, 3389-3402.
39. van Heijernoort, J. (1996). Murein synthesis. In *E. coli and Salmonella: Cellular and Molecular Biology.* (Neidhardt, F.C., ed.), pp. 1025-1034, ASM Press, Washington DC, USA.
40. Park, J.T. (1996). The murein sacculus. In *E. coli and Salmonella: Cellular and Molecular Biology.* (Neidhardt, F.C., ed.), pp. 48-57, ASM Press, Washington DC, USA.
41. Wu, H.C. (1996). Biosynthesis of lipoproteins. In *E. coli and Salmonella: Cellular and Molecular Biology.* (Neidhardt, F.C., ed.), pp. 1005-1014, ASM Press, Washington DC, USA.
42. Zhang, W.Y. & Wu, H.C. (1992). Alterations of the carboxy-terminal aminoacid residues of *E. coli* lipoprotein affect the formation of murein-bound lipoprotein. *J. Biol. Chem.* 267, 19560-19564.
43. Lin, Y., Fletcher, C.M., Zhou, J., Allis, C.D. & Wagner G. (1999). Solution structure of the catalytic domain of GCN5 histone acetyltransferase bound to coenzyme A. *Nature* 400, 86-89.
44. Studier, F.W. & Moffat, B.A. (1986). Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *J. Mol. Biol.* 189, 113-130.
45. Bouveret, E., Rigal, A., Lazdunski, C. & Bénédetti, H. (1998). Distinct regions of the colicin A translocation domain are involved in the interaction with TolA and TolB proteins upon import into *Escherichia coli. Mol. Microbiol.* 27, 143-157.
46. Hendrickson, W.A., Horton, J.R. & LeMaster, D.M. (1990). Selenomethionyl proteins produced for analysis by multiwavelength anomalous diffraction (MAD): a vehicle for direct determination of three-dimensional structure. *EMBO J.* 9, 1665-1672.
47. Audic, S., Lopez, F., Claverie, J.-M., Poirot, O. & Abergel C. (1997). SAmBA: an interactive software for optimizing the design of protein crystallization experiments. *Proteins* 29, 251-256.
48. Abergel, C., *et al.,* & Bénédetti, H. (1998). Crystallization and preliminary crystallographic study of a component of the *Escherichia Coli* Tol system: TolB. *Acta Crystallogr. D* 54, 102-104.
49. Ravelli, R.B.G., Sweet, R.M., Skinner, J.M., Duisenberg, A.J.M. & Kroon, J. (1997). STRATEGY: a program to optimize the starting spindle angle and scan range for X-ray data collection. *J. Appl. Crystallogr.* 30, 551-554.
50. Otwinowski, Z. (1993). DENZO: oscillation data reduction program. In *Data Collection and Processing.* (Sawyer, L., Isaacs, N.W. & Bailey, S. eds), pp. 56-63, DLSCI/R34 Daresbury Laboratory, UK.
51. Collaborative Computational Project, Number 4. (1994). The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* 50, 760-766.
52. Terwilliger, T.C. & Berendzen, J. (1999). Automated structure solution for MIR and MAD. *Acta Crystallogr. D* 55, 849-861.
53. Wang, B.C. (1985). Resolution of phase ambiguity in macromolecular crystallography. *Methods Enzymol.* 115, 90-112.
54. Zhang, K.Y.J. & Main, P. (1990). The use of Sayre's equation with solvent flattening and histogram matching for phase extension and refinement of protein structures. *Acta Crystallogr. A* 46, 41-46.
55. Roussel, A. & Cambillau, C. (1991). The TURBO-FRODO Graphics Package. In *Silicon Graphics Geometry Partners Directory* (Silicon Graphics, ed.), vol. 81, Mountain View, USA.
56. Brünger, A.T., *et al.,* & Warren, G.L. (1998). Crystallography and NMR system (CNS): a new software system for macromolecular structure determination, *Acta Crystallogr. D* 54, 905-921.
57. Laskowski, R., MacArthur, M., Moss, D. & Thornton, J. (1993). Procheck: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* 26, 91-97.
58. Higgins, D.G., Thompson, J.D. & Gibson, T.J. (1996). Using CLUSTAL for multiple sequence alignments. *Methods Enzymol.* 266, 383-402.
59. Abergel, C., Robertson, D.L. & Claverie, J.-M. (1999). Hidden dUTPase sequence in human HIV type 1 gp120. *J. Virol.* 73, 751-753.
60. Kraulis, P.J. (1991). MOLSCRIPT: a program for producing both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.* 24, 946-950.
61. Andersson, S.G., *et al.,* & Kurland, C.G. (1998). The genome sequence of *Rickettsia prowazekii* and the origin of mitochondria. *Nature* 396, 133-140.

Because *Structure with Folding & Design* operates a 'Continuous Publication System' for Research Papers, this paper has been published on the internet before being printed (accessed from http://biomednet.com/cbiology/str). For further information, see the explanation on the contents page.