Crystal structure of the type IV secretion system component CagX from *Helicobacter pylori*

Jin Zhang, Fei Fan, Yanhe Zhao, Lifang Sun, Yadan Liu, Ronan M. Keegan, Michail N. Isupov and Yunkun Wu

IUCr Journals

CRYSTALLOGRAPHY JOURNALS ONLINE

Crystal structure of the type IV secretion system component CagX from *Helicobacter pylori*

Jin Zhang,[a] Fei Fan,[b] Yanhe Zhao,[a] Lifang Sun,[a] Yadan Liu,[a] Ronan M. Keegan,[c,d] Michail N. Isupov[e] and Yunkun Wu[a]*

[a]State Key Laboratory of Structural Chemistry, Fujian Institute of Research on the Structure of Matter, Chinese Academy of Science, Fuzhou 350002, People's Republic of China, [b]Fujian Health College, Fuzhou 350101, People's Republic of China, [c]CCP4, Research Complex at Harwell, Rutherford Appleton Laboratory, Harwell Oxford, Didcot OX11 0FA, England, [d]Institute of Integrative Biology, University of Liverpool, Liverpool L69 7ZB, England, and [e]The Henry Wellcome Building for Biocatalysis, Biosciences, College of Life and Environmental Sciences, University of Exeter, Exeter EX4 4QD, England. *Correspondence e-mail: wuyk@fjirsm.ac.cn

Helicobacter pylori, a Gram-negative bacterial pathogen prevalent in the human population, is the causative agent of severe gastric diseases. An *H. pylori* type IV secretion (T4S) system encoded by the cytotoxin-associated gene pathogenicity island (*cag*PAI) is responsible for communication with host cells. As a component of the *cag*PAI T4S system core complex, CagX plays an important role in virulence-protein translocation into the host cells. In this work, the crystal structure of the C-terminal domain of CagX (CagXct), which is a homologue of the VirB9 protein from the VirB/D4 T4S system, is presented. CagXct is only the second three-dimensional structure to be elucidated of a VirB9-like protein. Another homologue, TraO, which is encoded on the *Escherichia coli* conjugative plasmid pKM101, shares only 19% sequence identity with CagXct; however, there is a remarkable similarity in tertiary structure between these two β-sandwich protein domains. Most of the residues that are conserved between CagXct and TraO are located within the protein core and appear to be responsible for the preservation of this domain fold. The studies presented here will contribute to our understanding of different bacterial T4S systems.

1. Introduction

Helicobacter pylori is a Gram-negative bacterial pathogen which lives in the stomachs of more than half of the population of the world and causes chronic gastritis, peptic ulcers, gastric adenocarcinoma and gastric lymphoma (Blaser, 1997; Blaser & Atherton, 2004; Covacci *et al.*, 1999; Parsonnet *et al.*, 1994). The cytotoxin-associated gene pathogenicity island (*cag*PAI) contains a 40 kb foreign DNA region which is considered to be the main genetic determinant of more virulent *H. pylori* strains (Backert *et al.*, 2002; Blaser & Atherton, 2004). The *cag*PAI carries 27 genes which encode several effector proteins and one type IV secretion (T4S) system. T4S systems are important nanomachines in Gram-negative bacteria which play important roles in various biological processes from the transfer of virulence factors into eukaryotic cells to the conjugative delivery of genetic material and the uptake or release of DNA (Cascales & Christie, 2003). The T4S system of *H. pylori* translocates the major effector protein CagA into gastric epithelial cells. Subsequent phosphorylation of intracellular CagA leads to epithelial cell elongation and disruption of tight junctions, which results in the pathogenicity of *H. pylori* (Backert *et al.*, 2002; Blaser & Atherton, 2004; Odenbreit *et al.*, 2000; Segal *et al.*, 1997; Backert & Selbach, 2008). The T4S system has been visualized at the surface between *H. pylori* and gastric epithelial cells, forming a
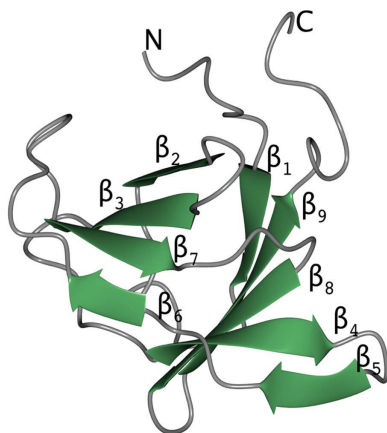
Table 1
Cloning of expression constructs.

Source organism	*H. pylori*
DNA source	cDNA isolated from *H. pylori* 26695
Forward primer (FCagXct-EcoRI)	CCGGAATTCCCCGTGCCTAGAAACTACAACTAC
Reverse primer (NCagXct-XhoI)	CCGCTCGAGTTATGTCAATGGATTTTTCCCAT-AGCC
Cloning vector	pET-32a
Expression vector	pET-32a
Expression host	*E. coli* BL21 (DE3)
Complete amino-acid sequence of the construct product	PVPRNYNYYQAPEKRSKHIMPSEIFDDGTFTY-FGFKNITLQPAIFVVQPDGKLSMTDAAIDP-NMTNSGLRWYRVNEIAEKFKLIKDKALVTV-INKGYGKNPLT

needle-like structure crossing the inner and outer membranes (Zanotti & Cendron, 2014). It has been proposed that the transmembrane core complex of the *H. pylori cag*PAI T4S system is formed by CagY, CagT and CagX, while the external pilus is thought to be composed of a large number of copies of CagC and CagL (Fischer, 2011; Terradot & Waksman, 2011).

As an essential component of the *cag*PAI T4S system core complex, CagX plays a critical role in CagA translocation into the host cell. Genetic and functional studies have indicated that the ability of *H. pylori* to translocate CagA into gastric cells is abrogated by inactivation of CagX (*hp0528*; Akopyants *et al.*, 1998; Censini *et al.*, 1996; Li *et al.*, 1999; Fischer *et al.*, 2001). The C-terminal domain of CagX has been shown to be responsible for its interaction with CagT by co-immuno-precipitation, MBP pull-down and yeast two-hybrid assays (Gopal *et al.*, 2015).

CagX is presumed to be structurally and functionally homologous to VirB9 from the well studied VirB/D4 T4S of the plant pathogen *Agrobacterium tumefaciens*, although the two proteins have low sequence identity (Bayliss *et al.*, 2007;

Christie *et al.*, 2005; Censini *et al.*, 1996). While VirB9 forms a heterotrimer with VirB7 and VirB10 in the VirB/D4 T4S system, it is proposed that CagX associates not only with CagT and CagY, but also with CagM and Cagδ, to form a trans-membrane core complex in the *cag*PAI T4S system (Kutter *et al.*, 2008; Pinto-Santini & Salama, 2009).

Here, we report the expression, purification, crystallization, structure determination and analysis of the C-terminal domain of the CagX protein (CagXct). This important component of the *cag*PAI T4S system folds into a β-sandwich domain containing nine β-strands. It is the first structure of a component of the transmembrane core complex of the *cag*PAI T4S system to be determined, and is only the second three-dimensional structure of a VirB9 homologue. The crystal structure was determined by the molecular-replacement method and refined at a resolution of 1.4 Å. Comparison of the CagXct structure with that of the C-terminal domain of another VirB9 homologue, TraO, which is a part of the outer
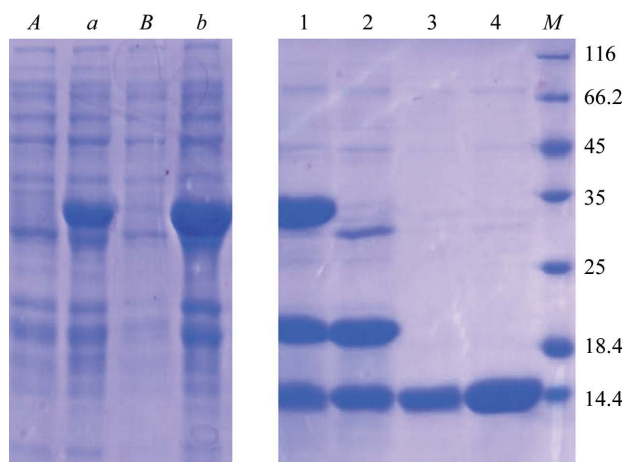

Figure 1
15% SDS–PAGE analysis of purified CagXct stained with Coomassie Brilliant Blue. Lane *A*, non-induced expression strains. Lane *a*, induced expression strains. Lane *B*, pellet fractions. Lane *b*, supernatant fractions. Lane 1, eluted with 500 m*M* imidazole; the band containing CagXct with a Trx tag is around 35 kDa. Lane 2, after cleavage by TEV; the band containing CagXct is under 14.4 kDa. Lane 3, flowthrough of the second nickel column. Lane 4, the sample before size-exclusion chromatography, showing high purity. Lane *M* contains molecular-mass marker (labelled in kDa).
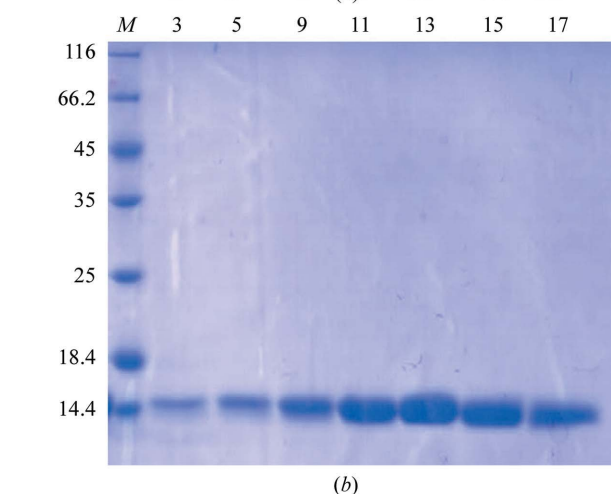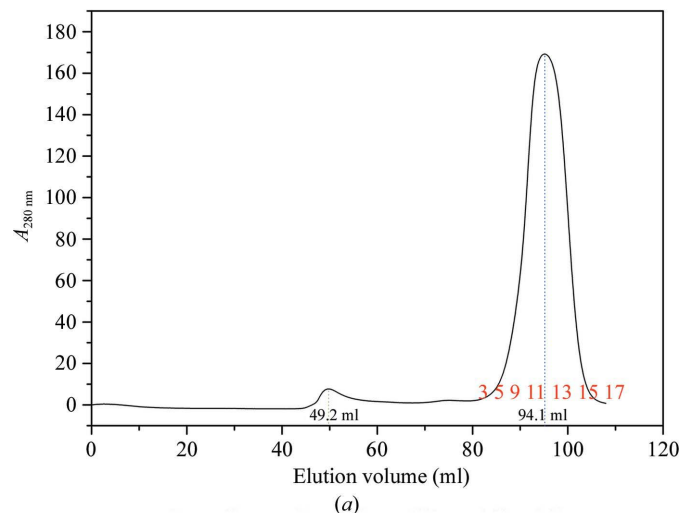

Figure 2
(*a*) Size-exclusion chromatography trace (HiLoad 16/600 Superdex 200 75 pg; GE Healthcare). CagXct elutes at 80–110 ml, which is in agreement with its monomer molecular mass of 11.4 kDa. The peak intensity is 192 mAU. Samples between 80 and 110 ml labelled 3–17 in red were picked for 15% SDS–PAGE analysis. (*b*) 15% SDS–PAGE analysis of the purified CagXct samples labelled 3, 5, 9, 11, 13, 15 and 17 in (*a*). Lane *M* contains molecular-mass marker (labelled in kDa).

Table 2
Data collection and processing.

Values in parentheses are for the outer shell.

Diffraction source	BL17U1, SSRF
Wavelength (Å)	0.97946
Temperature (K)	100
Detector	ADSC Q315R
Rotation range per image (°)	1
Total rotation range (°)	180
Exposure time per image (s)	0.8
Space group	$P2_1$
a, b, c (Å)	33.1, 61.6, 48.4
α, β, γ (°)	90.00, 90.23, 90.00
Resolution range (Å)	38.07–1.40 (1.44–1.40)
Total No. of reflections	130565
No. of unique reflections	34978
Completeness (%)	91.7 (56.0)
$\langle I/\sigma(I)\rangle$	12.3 (3.0)
R_{merge} (%)†	5.0 (34.7)
Multiplicity	3.7 (3.4)
$CC_{1/2}$	0.997 (0.803)

† $R_{merge} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$, where $I_i(hkl)$ are the intensities of the individual measurements of a given reflection hkl and $\langle I(hkl)\rangle$ is the average intensity over all replicates of that reflection, $\sum_{hkl}$ is the sum over all reflections and $\sum_i$ is the sum over i measurements of the reflection.

membrane complex encoded by the *Escherichia coli* conjugative plasmid pKM101, suggests the possible conservation of some protein–protein interactions between the pKM101 T4S system and the *cag*PAI T4S system.

2. Materials and methods

2.1. Cloning of expression constructs and protein expression

The gene encoding the soluble fragment of Hp0528 (amino acids 396–498; CagXct) was amplified from *Hp26695* genomic DNA using the primer pair FCagXct/NCagXct and subcloned into pET-32a vector *via* EcoRI and XhoI (Table 1). The constructed pET-32a-CagXct was transformed into *E. coli* BL21 (DE3) cells. Expression of CagXct was performed in LB medium containing 100 μg ml^{-1} ampicillin and the culture was incubated at 37°C and 220 rev min^{-1} until an OD_{600} of 0.6–0.8 was reached. 0.3 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) was added to induce the expression of recombinant CagXct and the culture was left to shake for 12 h at 16°C and 180 rev min^{-1}. The expression levels of CagXct were monitored by SDS–PAGE (Fig. 1).

2.2. Protein purification and crystallization

The cells were harvested by centrifugation (7000 rev min^{-1}, 5 min), resuspended in a lysis buffer consisting of 50 mM Tris–HCl pH 7.0, 500 mM NaCl, 5%(v/v) glycerol, 1 mM phenylmethylsulfonyl fluoride (PMSF), 1%(v/v) Tween 20 and then sonicated on ice (Scientz-IID). The recombinant protein was purified from the supernatant by immobilized nickel-affinity column chromatography (GE Healthcare) and digested with TEV protease to remove the His tag and Trx tag (Fig. 1b). The fractions were analyzed by SDS–PAGE (Fig. 1). The purified CagXct was then concentrated to 5 mg ml^{-1} and applied onto a size-exclusion chromatography column (HiLoad 16/600 Superdex 200 75 pg; GE Healthcare) in 25 mM Tris–HCl pH

7.0, 150 mM NaCl, 5% glycerol. The trace showed that most of the CagXct eluted in a monomeric form (Fig. 2a). Fractions were assessed by 15% SDS–PAGE (Fig. 2b).

The purified CagXct was concentrated and crystallized using the sitting-drop vapour-diffusion method in 96-well Intelli-Plates (Art Robbins Instruments). The crystallization trials were set up with an ARI robot (Art Robbins Instruments) using the following screens: Crystal Screen and Crystal Screen 2 (Hampton Research) and The PEGs, JCSG and Classics Suites (Qiagen). The drops had a total volume of 1 μl and consisted of a 1:1 ratio of protein solution to precipitant. The best crystals of CagXct grew from protein solution concentrated to 70 mg ml^{-1} with a precipitant consisting of 27% polyethylene glycol 4000, 100 mM HEPES pH 7.5, 10% 2-propanol.

2.3. Data collection and processing

Data were collected at 100 K on the BL17U1 beamline of the Shanghai Synchrotron Radiation Facility (SSRF) at a wavelength of 0.9792 Å. The data were initially processed with *HKL*-2000 (Otwinowski & Minor, 1997) and the *CCP*4 suite (Winn *et al.*, 2011) in the orthorhombic space group $P2_12_12_1$, with unit-cell parameters a = 33.1, b = 61.6, c = 48.4 Å, $\alpha = \beta = \gamma = 90°$. Owing to problems in molecular repacement (MR) and refinement in this space group, the data were later reprocessed as triclinic using *DIALS* (Waterman *et al.*, 2016) in the *xia*2 pipeline (Winter *et al.*, 2013). The overall *DIALS* R_{merge} value decreased from 0.106 in $P2_12_12_1$ to 0.050 in $P1$.

After the correct space group had been established by the downstream refinement as monoclinic $P2_1$ with unique b = 61.6 Å, β = 90.23°, the data were reprocessed to 1.4 Å resolution in this space group with *XDS* (Kabsch, 2010) in the *xia*2 pipeline (Table 2).

2.4. Structure solution and refinement

A promising MR solution was originally found in $P2_12_12_1$ with the *MrBUMP*/*Phaser* pipeline (Keegan & Winn, 2008; McCoy *et al.*, 2007) using the structure of chain *B* (TraO) of the *E. coli* pKM101 plasmid-encoded outer membrane complex as a search model (Chandran *et al.*, 2009; PDB entry 3jqo; 19% sequence identity). The model, which contained one CagXct molecule per asymmetric unit (41% solvent), was refined using *REFMAC*5 (Murshudov *et al.*, 2011; Winn *et al.*, 2001) and rebuilt using *Coot* (Emsley *et al.*, 2010). Missing protein side chains were clearly visible in the electron density; however, the rebuilt structure could not be refined to an R_{free} value of below 0.38. An inspection of systematic absences proved to be inconclusive regarding the screw/rotation character of the crystal axes, therefore *MoRDa* (Vagin & Lebedev, 2015) was used to conduct MR and refinement in every possible orthorhombic space group. Surprisingly, high-scoring solutions were found in several space groups; however, the $P2_12_12_1$ solution appeared to be the best since the *MoRDa* solutions in other space groups had higher R factors.

Intensity statistics (*L*-test; Padilla & Yeates, 2003) implemented in *POINTLESS* (Evans, 2011) suggested that the

CagXct crystal belonged to a lower symmetry space group and was twinned. A distinct crystallographic dyad could not be determined by the processing statistics, since the merging R factors were close to 10% for each crystal axis.

To establish the true space group, the data were reprocessed in $P1$. The MR solution (four protein monomers) in this space group was found using *MoRDa* with PDB entry 3jqo chain B as the model. The solution had a *MoRDa* Q-factor of 0.71, a probability of correct solution of 99% and refined to an R_{free} of 0.40 without any model rebuilding, with most of the amino-acid side chains absent. The *CCP*4 program *Zanuda* (Lebedev & Isupov, 2014) was applied to the partially refined $P1$ model, which had an R_{free} of 0.32. The true space group, in which the model refined to the same R values, was monoclinic $P2_1$, with unique $b = 61.6$ Å, $\beta = 90.23°$. The R_{free} values were 0.38 or higher in all other monoclinic and orthorhombic space groups. *SFCHECK* (Vaguine *et al.*, 1999) analysis of data reprocessed in $P2_1$ clearly identified a nonmerohedral twinning operation $(-h, -k, l)$ with an obliquity of $0.23°$ and a twinning fraction of 0.46. Twin isotropic B-factor refinement of the CagXct structure was performed with *REFMAC*. The atomic coordinates and structure factors for the CagXct structure have been deposited in the Protein Data Bank with accession code 5h3v.

3. Results and discussion

3.1. Cloning, overexpression, purification and crystallization

The soluble fragment (residues 396–498) of CagX was successfully cloned and expressed in *E. coli* strain BL21 (DE3) with a His tag and a Trx tag. The tags were removed by limited proteolysis of CagXct bound to a nickel immobilized metal ion-affinity column using TEV protease. The protein was further purified by a second run of nickel immobilized metal ion-affinity chromatography and size-exclusion chromato-
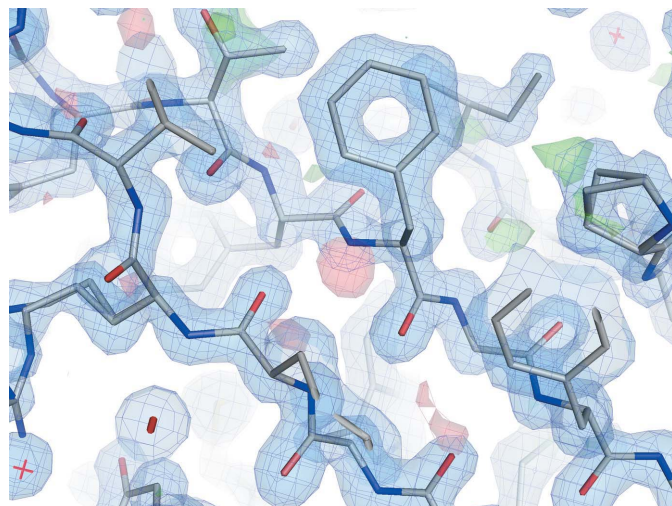

Figure 3
Electron-density maps around β-sheet 2 of CagXct monomer A. The $2F_o - F_c$ map (blue) is contoured at 1.4σ and the $F_o - F_c$ map is contoured at 3.0σ (green) and -3.0σ (red). The difference density ripples (red and green) are likely to be owing to crystal twinning. This figure was prepared using *PyMOL* (Schrödinger).

Table 3
Structure solution and refinement of CagXct.

PDB code	5h3v
Resolution range (Å)	38.07–1.40 (1.43–1.40)
Final R_{cryst}	0.200 (0.291)
Final R_{free}	0.249 (0.442)
No. of twin domains	2
Twin fractions	0.509 (h, k, l), 0.491 ($-h$, $-k$, l)
L-test for twinning	$\langle |L| \rangle = 0.40$, $\langle L^2 \rangle = 0.22$
R.m.s. deviations	
Bonds (Å)	0.012
Angles (°)	1.712
Wilson B factor (Å^2)†	20.2
No. of protein residues	206
No. of solvent atoms	145
Average B-factor values	
Protein (Å^2)	15.0
Solvent (Å^2)	22.6
Ramachandran plot analysis‡	
Most favoured (%)	85.2
Additionally allowed (%)	13.6
Generously allowed (%)	0.6
Disallowed	0.6

† The Wilson B factor was calculated using *SFCHECK* (Vaguine *et al.*, 1999). ‡ The Ramachandran plot analysis was performed by *PROCHECK* (Laskowski *et al.*, 1993).

graphy. CagXct was concentrated to 70 mg ml^{-1} and crystallized by the vapour-diffusion method from PEG and 2-propanol.

3.2. Quality of the model

The CagXct crystal structure was solved by MR and subjected to isotropic B-factor twin refinement in *REFMAC*5 at 1.4 Å resolution after the true space group had been established as monoclinic $P2_1$ with a β angle close to 90°, with the crystal forming a nonmerohedral twin. The quality of the electron-density maps was mostly acceptable, although some ripples were observed in the $F_o - F_c$ map calculated using the detwinned data (Fig. 3). These ripples are likely to be caused by crystal twinning since similar electron-density features have been reported for other twinned and order–disorder crystal structures (Lebedev *et al.*, 2006; Rye *et al.*, 2007). The CagXct model was refined to an R factor of 0.200 and an R_{free} of 0.249 in space group $P2_1$ (Table 3). It contains two protein monomers with all residues built into the electron density, one PEG and two 2-propanol molecules and 129 waters. Many residue side chains were modelled with alternative conformations. The CagXct model contains no residues in the *cis*-conformation. Asp481 in both monomers of CagXct is a Ramachandran plot outlier, while Gly258 in the equivalent position in the TraO structure has similar main-chain torsion angles. The two monomers of CagXct comprising the asymmetric unit can be superimposed with an r.m.s.d. of 0.38 Å over all 103 C$^\alpha$ atoms.

3.3. Overall structure

CagXct folds into a β-sandwich domain formed by two antiparallel β-sheets containing nine β-strands (Fig. 4). β-Sheet 1 contains strands $\beta1$, $\beta4$, $\beta5$, $\beta8$ and $\beta9$ and has Richardson topology 3, 1, $-2x$, -1 (Richardson, 1981). β-Sheet 2 contains strands $\beta2$, $\beta3$, $\beta6$ and $\beta7$ and has topology 1, 2x, -1.
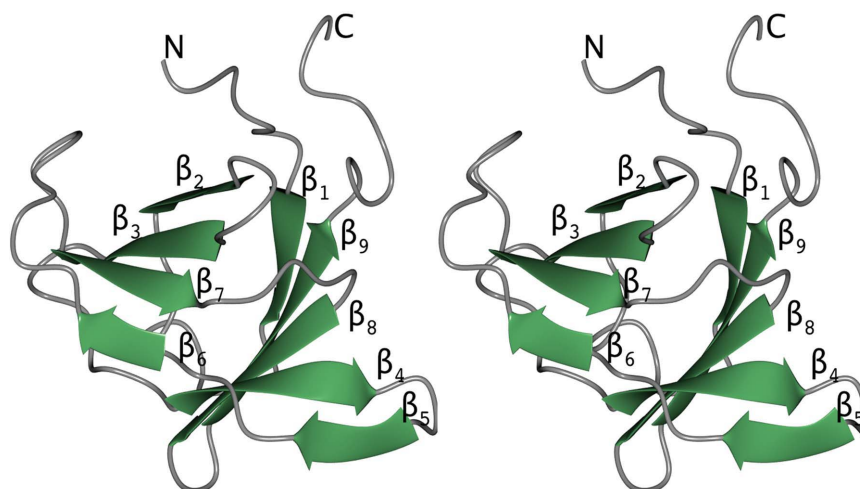
Figure 4
A stereo diagram showing a cartoon representation of CagXct, with the β-strands coloured green and loops coloured grey. The secondary-structure elements of this β-sandwich domain are numbered. Figs. 4 and 6 were prepared using *CCP4mg* (McNicholas *et al.*, 2011).

The two independent CagXct monomers do not form any oligomer between themselves or with their crystal symmetry mates, which is in line with a monomer being the main species in solution, as can be seen from the size-exclusion chromatography trace (Fig. 2*a*).

3.4. Comparison of CagXct with TraO

A *DALI* search (Holm & Rosenström, 2010) reveals that the most similar structure to CagXct is the C-terminal domain of the VirB9 homologue TraO (TraOct) from the crystal structure of the pKM101 plasmid-encoded T4S outer membrane complex (Chandran *et al.*, 2009; PDB entry 3jqo; chain *B* was used as an MR model with a sequence identity of 19%). This ∼0.6 MDa complex represents a 14-fold rotational symmetry ring spanning the outer membrane which is formed by the pKM101 T4S system proteins TraOct, TraN and TraFct. An NMR structure is also available for TraO in complex with another component of the pKM101 T4S system, the VirB7-like TraN (PDB entry 3jqo; Chandran *et al.*, 2009).

CagXct and TraOct are remarkably similar, despite their low sequence identity (Figs. 5 and 6). The region 177–270 of the TraO monomer (PDB entry 3jqo; chain *B*) aligns with the CagXct monomer with an r.m.s.d. of 1.4 Å over 95% of the C$^\alpha$ atoms. Such a high level of structural similarity may explain the success of MR structure solution using the TraO model.

To maintain structural similarity, many buried protein residues which form the protein core of TraO are conserved or substituted by residues with similar properties in CagX. Remarkably, most of these conserved residues retain the same side-chain conformation (Fig. 6). Generally, it would appear that the conservation of the overall fold/shape of the VirB9-like domain is more important for the function of this member of the T4S system than the preservation of specific individual residues that may be involved in interactions with other proteins forming the outer membrane complex of the T4S system.

3.5. Conservation of protein–protein interactions in T4S

The structure of the outer membrane complex of the T4S plasmid pKM101 (Chandran *et al.*, 2009) provides insight into the general architecture of bacterial T4S systems. A TraOct domain in this ring structure interacts with two neighbouring TraO monomers, two TraF monomers and a single TraN monomer. Since relatively little is known about the interactions of CagX with its partners in the *H. pylori* T4S system, it is interesting to map the sequence/structure conservation features between TraO and CagXct onto these monomer–monomer interactions.

TraN is a long peptide which winds around TraO. It adds an additional β-strand to β-sheet 1 of TraO, which is observed in both the binary complex TraO–TraN and in the heterotrimeric outer membrane complex. Most amino-acid residues of strand β1, which runs antiparallel to TraN in TraO, are not conserved in CagX; however, the main chains of the matching residues of
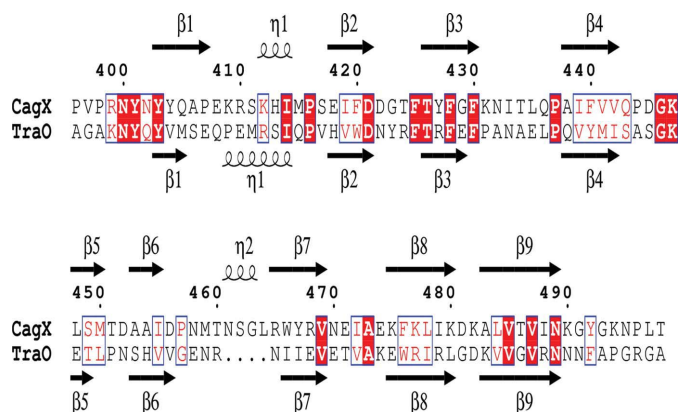

Figure 5
Anmino-acid sequence alignment of CagXct and TraOct. The secondary-structure elements are indicated above and below the alignment, respectively, as β-strands and η-helices (3₁₀-helices). Conserved residues are shown in red boxes; matching amino acids with similar properties are shown in blue boxes. The secondary-structure assignments were carried out and the figure was produced using *ESPript*3 (Robert & Gouet, 2014).
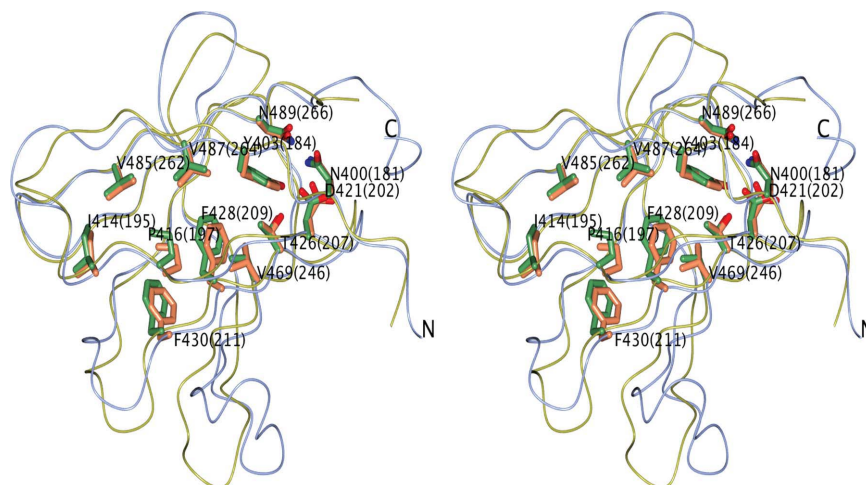
Figure 6
A stereo diagram showing the structural superposition of CagX (ice-blue worm model) and TraO (yellow; PDB entry 3jqo). Side chains of residues forming the conserved protein core are shown with C atoms in coral (CagX) or green (TraO). Residue numbers of TraO are given in parentheses after these of the CagX residues.

both these T4S proteins have the same conformation and solvent accessibility. Thus, it appears likely that the β-sheet interaction between TraO and TraN is reproduced in the Cag–CagT interface.

Interactions between different TraO monomers in the 14-fold ring of the outer membrane complex of T4S plasmid pKM101 are not extensive and the residues involved in these interactions do not appear to be conserved in CagX.

Interactions between TraO and the two monomers of TraF in this complex are more extensive; however, there is little conservation of amino-acid residues in this interface. Interestingly, one of the 2-propanol molecules binds to the main-

chain N atom of the Ramachandran plot outlier Asp481 in monomer A in the CagXct structure. The main-chain N atom of the equivalent Gly258 in TraO forms a hydrogen bond to the main-chain O atom of Gly364 in TraF, with the positions of Gly364 and the 2-propanol ligand O atoms overlapping. This may suggest the preservation of another main-chain interaction in the CagX–CagY interface.

The sequence pattern 'xVxVxN' (where x represents a binding residue) is shared by both structures in the β9 strand in TraO and CagXct, suggesting that the conserved non-binding residues in this pattern may play a role in maintaining the proper distance with the β1 strand in a spatial configuration to ensure binding to a protein partner (Fig. 7). However, the binding residues are not conserved in β9; Leu484, Thr486 and Ile488 in CagXct correspond to Val261, Gly263 and Arg264 in TraO (Figs. 5 and 7). The diversity in binding residues should be determined by their different interacting protein substrates.

CagXct is only the second VirB9 homologue for which a three-dimensional structure has been elucidated. Its structure will provide additional insight into binding and translocation mechanisms of the transmembrane core complex in *H. pylori* and other T4S systems.
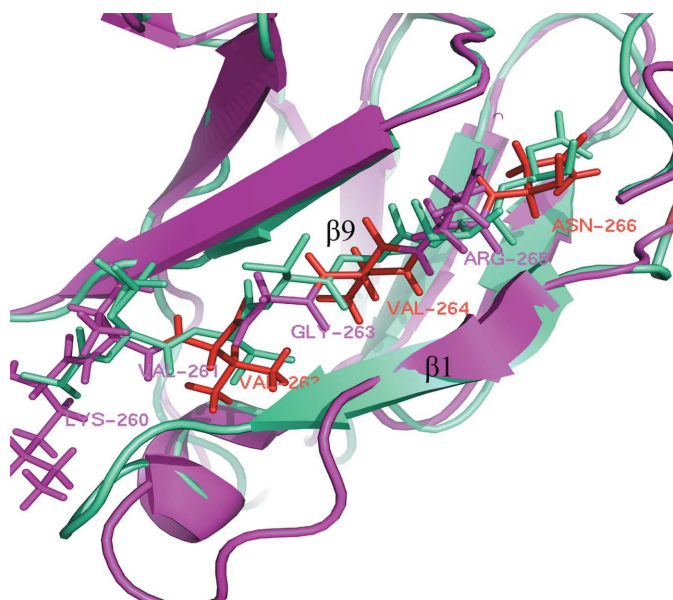
Figure 7
Comparison of the protein-binding region of the β9 strand (shown as sticks; β1 strands are shown as cartoons) in CagXct (cyan) and in TraO (magenta, only the residues of TraO are labelled). Apart from the binding residues, there is a high sequence-similarity pattern: 'xVxVxN' (these residues are shown in red; x represents the binding amino-acid residues).

References

Akopyants, N. S., Clifton, S., Kersulyte, D., Crabtree, J. E., Youree, B. E., Reece, C. A., Bukanov, N. O., Drazek, E. S., Roe, B. A. & Berg, D. E. (1998). *Mol. Microbiol.* **28**, 37–53.

Backert, S., Churin, Y. & Meyer, T. F. (2002). *Keio J. Med.* **51**, Suppl. 2, 6–14.

Backert, S. & Selbach, M. (2008). *Cell. Microbiol.* **10**, 1573–1581.

Bayliss, R., Harris, R., Coutte, L., Monier, A., Fronzes, R., Christie, P. J., Driscoll, P. C. & Waksman, G. (2007). *Proc. Natl Acad. Sci. USA*, **104**, 1673–1678.

Blaser, M. J. (1997). *J. Clin. Invest.* **100**, 759–762.

Blaser, M. J. & Atherton, J. C. (2004). *J. Clin. Invest.* **113**, 321–333.

Cascales, E. & Christie, P. J. (2003). *Nature Rev. Microbiol.* **1**, 137–149.

Censini, S., Lange, C., Xiang, Z., Crabtree, J. E., Ghiara, P., Borodovsky, M., Rappuoli, R. & Covacci, A. (1996). *Proc. Natl Acad. Sci. USA*, **93**, 14648–14653.

Chandran, V., Fronzes, R., Duquerroy, S., Cronin, N., Navaza, J. & Waksman, G. (2009). *Nature (London)*, **462**, 1011–1015.

Christie, P. J., Atmakuri, K., Krishnamoorthy, V., Jakubowski, S. & Cascales, E. (2005). *Annu. Rev. Microbiol.* **59**, 451–485.

Covacci, A., Telford, J. L., Del Giudice, G., Parsonnet, J. & Rappuoli, R. (1999). *Science*, **284**, 1328–1333.

Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). *Acta Cryst.* **D66**, 486–501.

Evans, P. R. (2011). *Acta Cryst.* **D67**, 282–292.

Fischer, W. (2011). *FEBS J.* **278**, 1203–1212.

Fischer, W., Püls, J., Buhrdorf, R., Gebert, B., Odenbreit, S. & Haas, R. (2001). *Mol. Microbiol.* **42**, 1337–1348.

Gopal, G. J., Pal, J., Kumar, A. & Mukhopadhyay, G. (2015). *Biochem. Biophys. Res. Commun.* **456**, 98–103.

Holm, L. & Rosenström, P. (2010). *Nucleic Acids Res.* **38**, W545–W549.

Kabsch, W. (2010). *Acta Cryst.* **D66**, 125–132.

Keegan, R. M. & Winn, M. D. (2008). *Acta Cryst.* **D64**, 119–124.

Kutter, S., Buhrdorf, R., Haas, J., Schneider-Brachert, W., Haas, R. & Fischer, W. (2008). *J. Bacteriol.* **190**, 2161–2171.

Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.

Lebedev, A. A. & Isupov, M. N. (2014). *Acta Cryst.* **D70**, 2430–2443.

Lebedev, A. A., Vagin, A. A. & Murshudov, G. N. (2006). *Acta Cryst.* **D62**, 83–95.

Li, S. D., Kersulyte, D., Lindley, I. J., Neelam, B., Berg, D. E. & Crabtree, J. E. (1999). *Infect. Immun.* **67**, 3893–3899.

McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. (2007). *J. Appl. Cryst.* **40**, 658–674.

McNicholas, S., Potterton, E., Wilson, K. S. & Noble, M. E. M. (2011). *Acta Cryst.* **D67**, 386–394.

Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F. & Vagin, A. A. (2011). *Acta Cryst.* **D67**, 355–367.

Odenbreit, S., Püls, J., Sedlmaier, B., Gerland, E., Fischer, W. & Haas, R. (2000). *Science*, **287**, 1497–1500.

Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.

Padilla, J. E. & Yeates, T. O. (2003). *Acta Cryst.* **D59**, 1124–1130.

Parsonnet, J., Hansen, S., Rodriguez, L., Gelb, A. B., Warnke, R. A., Jellum, E., Orentreich, N., Vogelman, J. H. & Friedman, G. D. (1994). *N. Engl. J. Med.* **330**, 1267–1271.

Pinto-Santini, D. M. & Salama, N. R. (2009). *J. Bacteriol.* **191**, 7343–7352.

Richardson, J. S. (1981). *Adv. Protein Chem.* **34**, 167–339.

Robert, X. & Gouet, P. (2014). *Nucleic Acids Res.* **42**, W320–W324.

Rye, C. A., Isupov, M. N., Lebedev, A. A. & Littlechild, J. A. (2007). *Acta Cryst.* **D63**, 926–930.

Segal, E. D., Lange, C., Covacci, A., Tompkins, L. S. & Falkow, S. (1997). *Proc. Natl Acad. Sci. USA*, **94**, 7595–7599.

Terradot, L. & Waksman, G. (2011). *FEBS J.* **278**, 1213–1222.

Vagin, A. & Lebedev, A. (2015). *Acta Cryst.* **A71**, s19.

Vaguine, A. A., Richelle, J. & Wodak, S. J. (1999). *Acta Cryst.* **D55**, 191–205.

Waterman, D. G., Winter, G., Gildea, R. J., Parkhurst, J. M., Brewster, A. S., Sauter, N. K. & Evans, G. (2016). *Acta Cryst.* **D72**, 558–575.

Winn, M. D. *et al.* (2011). *Acta Cryst.* **D67**, 235–242.

Winn, M. D., Isupov, M. N. & Murshudov, G. N. (2001). *Acta Cryst.* **D57**, 122–133.

Winter, G., Lobley, C. M. C. & Prince, S. M. (2013). *Acta Cryst.* **D69**, 1260–1273.

Zanotti, G. & Cendron, L. (2014). *World J. Gastroenterol.* **20**, 1402–1423.