



HAL
open science

Caf1A usher possesses a Caf1 subunit-like domain that is crucial for Caf1 fibre secretion

Xiaodi Yu, Ganeshram R. Visweswaran, Zoe Duck, Srisailam Marupakula, Sheila Macintyre, Stefan D. Knight, Anton V. Zavialov

► **To cite this version:**

Xiaodi Yu, Ganeshram R. Visweswaran, Zoe Duck, Srisailam Marupakula, Sheila Macintyre, et al.. Caf1A usher possesses a Caf1 subunit-like domain that is crucial for Caf1 fibre secretion. *Biochemical Journal*, 2009, 418 (3), pp.541-551. 10.1042/BJ20080992 . hal-00479027

HAL Id: hal-00479027

<https://hal.science/hal-00479027>

Submitted on 30 Apr 2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Caf1A usher possesses a Caf1 subunit-like domain that is crucial for Caf1 fibre secretion

Xiaodi Yu^{*1}, Ganeshram R. Visweswaran^{*1}, Zoe Duck^{†1}, Srisailam Marupakula^{*}, Sheila MacIntyre[†], Stefan D. Knight^{*2}, and Anton V. Zavialov^{*2}

^{*}Department of Molecular Biology, Uppsala Biomedical Centre, Swedish University of Agricultural Sciences, Box 590, SE-753 24 Uppsala, Sweden, [†]School of Biological Sciences, University of Reading, Reading, RG6 6AJ, UK

Short title: Usher domain, a dummy fibre subunit

Abbreviations used: Caf1, subunit of *Yersinia pestis* capsule; Caf1M and Caf1A, chaperone and outer membrane usher protein, respectively, involved in *Yersinia pestis* capsule assembly; UND, usher N-terminal domain; UMD usher middle domain; OM, outer membrane, TM, transmembrane; UTBD, usher TM β -barrel domain; UCD, usher C-terminal domain; DSC, donor strand complementation; DSE, donor strand exchange; GdmCl, guanidinium hydrochloride; ΔG_{N-U} , free energy of unfolding, ΔH_m , van't Hoff enthalpy of melting; T_m , melting temperature; WT, wild type.

¹These authors contributed equally to the paper

²Address correspondence to: Anton V. Zavialov (E-mail anton.zavialov@molbio.slu.se) or S.D.K. (E-mail stefan.knight@molbio.slu.se)

Atomic coordinates for Caf1A middle domain (232-320 amino acid region) have been deposited in Protein Databank under accession code XXXX.

SYNOPSIS

The chaperone/usher pathway controls assembly of fibres of adhesive organelles of Gram-negative bacteria. The final steps of fibre assembly and fibre translocation to the cell surface are coordinated by outer membrane proteins, ushers. Ushers consist of several soluble periplasmic domains and a single transmembrane β -barrel. Here we report isolation and structural-functional characterization of a novel middle domain of the Caf1A usher (UMD) from *Yersinia pestis*. The isolated UMD is a highly soluble monomeric protein capable of autonomous folding. A 2.8 Å resolution crystal structure of UMD revealed that this domain has an immunoglobulin-like fold similar to that of donor strand complemented Caf1 fibre subunit. Moreover these proteins displayed significant structural similarity. Although UMD is in the middle of the predicted amphipathic β -barrel of Caf1A, the usher still assembled in the membrane in the absence of this domain. UMD did not bind Caf1M-Caf1 complexes, but its presence was shown to be essential for Caf1-fibre secretion. The study suggests that UMD may play the role of a subunit-substituting protein (dummy subunit), plugging or priming secretion through the channel in the Caf1A usher. Comparison of isolated UMD with the recent structure of the corresponding domain of PapC usher revealed high similarity of the core structures suggesting a universal structural adaptation of FGL and FGS chaperone/usher pathways for the secretion of different types of fibres. The functional role of two topologically different states of this plug domain suggested by structural and biochemical results is discussed.

Key words: chaperone/usher pathway, assembly, secretion, fimbrial adhesins, protein structure

INTRODUCTION

The chaperone/usher pathway [1, 2] represents one of the most common mechanisms for assembly of surface located fibrillar organelles in Gram negative bacteria. The surface fibres assembled via this pathway frequently play an important role in bacterial virulence by mediating adhesion to and potentially invasion into host cells, and in some cases also by protecting bacteria from the host's immune system [1, 2].

The fibrillar structures assembled by the chaperone/usher pathway are constructed from Ig-like β -sandwich subunits, the pilin subunits. However, in contrast to an Ig domain, a circular permutation in the pilin sequence positions the sequence corresponding to the seventh, C-terminal G β -strand of a canonical Ig domain at the N-terminus of the polypeptide chain [3-6]. In a typical Ig fold, the 'top' edge of the sandwich, defined by the A and F strands, is capped by the C-terminal G strand, which is hydrogen bonded to the F strand and provides hydrophobic residues to the core of the fold. Because it does not have this strand, a pilin subunit cannot cap its AF edge and a closed hydrophobic core cannot be formed by a single pilin subunit [3-7]. The N-terminal extension of pilins does not contribute to the subunit's globular fold, but deletions or mutations in this region block assembly of pilin subunits into fibers [3-8]. The structure of a Caf1M-(Caf1)₂ ternary complex, with the minimal *Y. pestis* F1-antigen fiber ((Caf1)₂) bound to the Caf1M chaperone, revealed that fiber subunits are linked together by insertion of the N-terminal extension of one subunit into the hydrophobic cleft of the second subunit [6, 7]. The inserted N-terminal segment adopts a β -strand conformation running anti-parallel to the F strand, with hydrophobic side chains bound in acceptor cleft sub-pockets, hence completing the Ig fold of the subunit. This mode of binding, termed 'donor strand complementation' (DSC), had previously been predicted for type 1 and P pilus fibers [3-5], and is likely to be present in all surface polymers assembled through the chaperone/usher pathway. The resulting linear fibres are composed of globular modules each having an intact Ig topology generated by DSC. Depending on subunit composition and mode of subunit-subunit interactions, secreted linear fibres coil into different structures such as thick rigid monoadhesive pili with a diameter of 7-8 nm, thin flexible polyadhesive pili with a poorly defined diameter (2-4 nm) and thin flexible polyadhesive fibers [1, 2].

Biogenesis of surface fibres is assisted by periplasmic chaperones and outer membrane (OM) ushers. As fibre subunits emerge in the periplasm, periplasmic chaperones bind them, ensure correct folding and deliver the folded subunits to the usher. The ushers then act as assembly platforms co-ordinating subunit polymerization and translocation to the cell surface.

The L-shaped periplasmic chaperones comprise two Ig-like domains at $\sim 90^\circ$ angle, with a large cleft between the two domains. The F₁ and G₁ β -strands in the 1st, N-terminal, domain are connected by a long and flexible loop. This loop harbours a conserved motif of hydrophobic residues that is critical for subunit binding [9]. Subunits bind in the cleft between the two chaperone domains with the subunit C-terminal carboxyl group anchored by two strictly conserved positively charged residues at the bottom of the subunit-binding cleft. In chaperone-subunit complexes, the A and F edge strands are hydrogen bonded to

the A₁ and G₁ strands of the chaperone, which creates a super-barrel of β -strands from both the subunit and the chaperone [6, 7]. In the super-barrel, large hydrophobic residues from the G₁ donor strand of the chaperone are inserted between the two sheets of the subunit β -sandwich and become an integral part of the bound subunit's hydrophobic core. Based on the lengths of the subunit binding motifs, periplasmic chaperones are grouped in two subfamilies, FGL and FGS (E₁G₁ long and E₁G₁ short, respectively) [1, 2]. Chaperones from these subfamilies operate with subunits displaying significant structural differences and assemble functionally and morphologically different organelles.

At the usher, chaperone-subunit interactions are replaced by subunit-subunit interactions in a process called 'donor strand exchange' (DSE). No energy input from external sources is required to convert periplasmic chaperone-subunit preassembly complexes into free chaperones and secreted fibres. Instead, assembly is driven by subunit folding energy conserved by the chaperone [6, 7]. Whereas the fibre-incorporated subunit adopts a very stable, compact structure, the chaperone bound subunit is in an open, molten-globule like, high-energy conformation. The large difference in stability between the chaperone-bound, partially folded conformation and the fully folded native fibre conformation of pilins creates a free energy potential that drives fibre formation. Ushers may accelerate donor strand exchange by binding and positioning subunit-chaperone complexes in an orientation favourable to attack of the subunit complex at the base of a growing fibre [6, 10-12]. The exchange of the chaperone G₁ donor strand together with the A₁ 'accessory' strand of the chaperone by the G_d donor strand of the subunit and accompanying collapse of the subunit core appear not to occur through separate steps of complete dissociation and association, but rather by partial displacement in a process termed zip-out-zip-in DSE [6, 11]. Directing zip-out-zip-in steps is another plausible mechanism for usher-mediated catalysis of DSE [6].

Structural information is essential to advance our understanding of the function of ushers. Ushers are large integral OM proteins of about 800 amino acid residues in length. A recent electron microscopy study revealed that the PapC usher involved in assembly of P pili exists as a dimeric complex, with each usher monomer containing apparent channels of ≈ 2 nm diameter [13]. This pore size would be wide enough for translocation of individual folded subunits or their polymers from the periplasm to the cell surface. The type-1 pilus usher FimD was shown to possess an N-terminal domain, comprising residues 1-139 (UND-FimD) [14]. Isolated UND-FimD specifically binds chaperone-subunit complexes with micromolar affinities [15]. The NMR structure of isolated UND-FimD revealed that this domain consists of a flexible, N-terminal 'tail', a core structure of a novel α/β fold and a potential hinge segment that connects the structured core to the rest of the usher [15]. UND-FimD was co-crystallized with the complex between the FimC chaperone and the pilin domain of the FimH subunit (FimH_p) [15]. The crystal structure of this ternary complex revealed that the flexible N-terminal tail of UND-FimD specifically interacts with both the chaperone and subunit, acting as a sensor for loaded FimC molecules. Following targeting to UND-FimD, FimCH chaperone-adhesin complexes form stable interactions with the FimD C-terminus and induce conformational changes in FimD, possibly priming the usher for pilus biogenesis [16, 17]. Structural information on the usher C-terminal domain (UCD) is currently not available.

Capitani et al. [18] predicted a 22-stranded TM β -barrel in the centre of the FimD usher. Importantly, this study also predicted the existence of a novel periplasmic domain formed by the most conserved region of the usher. In the prediction, the sequence of this 'middle domain' was flanked by TM β -strands of the usher TM β -barrel domain (UTBD). This model was recently confirmed on publication of the low resolution crystal structure of the TM segment of PapC usher [19]. In this structure, residues 130-640 of PapC form a 24-stranded β -barrel with the central predicted 'middle domain' located within the central barrel and lying across the channel.

Both FimH and PapC belong to the functional family of FGS chaperone associated ushers [1, 2, 20]. These two ushers assemble relatively rigid, mono-adhesive pili of complex subunit composition, including a 2-domain terminal adhesin subunit [1, 2, 20]. In contrast, little is known about FGL chaperone associated ushers involved in assembly of functionally different organelles, the thin poly-adhesive fibres of simple subunit composition [1]. In this study, using the Caf1A usher from the F1 capsule system of *Y. pestis* as a prototype member of the FGL family, we present the first structural detail of the usher of any member of this family.

Sequence analysis of Caf1A with the HMM-B2TMR program [21] predicts a structural organization very similar to that predicted for FimD (Figure 1). To check the prediction experimentally, we expressed a series of sequences to locate the core region of the usher middle domain (UMD) and investigated the secondary structure, oligomeric state, and stability of isolated UMD. The study revealed an UMD of about 80 amino acid residues capable of autonomous folding into a stable β -structural unit. X-ray structure determination revealed that UMD has a pilin-like fold with significant 3D structural similarity to donor strand complemented Caf1. This study has also demonstrated the absolute requirement of this domain for Caf1 export, although Caf1A itself can assemble in its absence. Based on this data and some topological differences between the isolated Caf1A UMD and channel inserted PapC UMD, a model is presented of how this dummy subunit (UMD) might act as a plug in the Caf1A β -barrel. Finally, our comparison of UMD of Caf1A and PapC suggests a universal structural adaptation of FGL and FGS chaperone/usher pathways for the secretion of different types of fibres and highlights some potentially important differences between these translocation systems.

MATERIALS AND METHODS

Plasmid construction and deletion mutagenesis

To create expression plasmids, encoding various lengths of Caf1A UMD, DNA encoding the amino acid fragments from residues 198, 213, 222, 232, 242, 243, 244, and 246 to residue 320 were amplified from the p12R plasmid [22] by PCR using forward F-PMD (1-8) and reverse R-PMD primers (Table 1S, the supplemental material). The amplified fragments were cloned into pCR-T7/CT-TOPO-TA (198-, 213-, 222-, and 232-320) or pET101D-TOPO (242-, 243-, 244-, and 246-320) expression vectors (Invitrogen) essentially as described in Invitrogen cloning procedures. The genes flanked with the Met and stop codons were positioned downstream of the *T7* promoter and ribosome binding site

provided in the vector. The resulted plasmids were transformed into expression strains, *Escherichia coli* BL21-AI or BL21(DE3). Deletion mutants *cafIA*Δ1-4 were created in p12R by inverse PCR [9] using reverse primers Caf1AΔm1, Caf1AΔm2, Caf1AΔm3 or Caf1AΔm4 and forward primer Caf1AΔm-F (Table 1S). Following mutagenesis PCR products were digested with *Bgl*III, religated and transformed into *E. coli* DH5α. Mutants were confirmed by *Bgl*III digestion of isolated plasmid and DNA sequencing of the complete operon.

Expression-purification of UMD constructs

E. coli BL21-AI or BL21(DE3) transformants growing at 37 °C, were induced for protein expression with 0.2% L-arabinose or 1 mM isopropyl β-D-1-thiogalactopyranoside, respectively. The cells were harvested by centrifugation, resuspended in a 20 mM bis-Tris-HCl pH 6.0 buffer (A), and lysed by sonication with Vibra cell (Sonics & Materials). The crude extract obtained was precipitated by slowly adding solid ammonium sulphate (Merck) to 40% saturation followed by incubation for one hour at 4°C. The precipitate was dissolved in the A buffer and dialyzed overnight against the same buffer. The sample was loaded onto an anion exchange Mono Q HR 10/10 column (Pharmacia Biotech). Proteins were eluted with a 0-150 mM gradient of NaCl. The UMD-containing fractions were pooled and protein in the sample was concentrated to 1.6-2 g/l on a Vivaspin device (Vivascience) with molecular weight cut-off of 5 kDa. The sample was loaded onto a HiLoad Superdex 75 prep grade size exclusion column (Pharmacia Biotech) pre-equilibrated with a 20mM HEPES, 150mM NaCl pH 7.5 buffer. UMD was eluted in a single peak at flow rate 1.5 ml/min. The peak fractions were collected and concentrated to 3-30 g/l for analysis or to 30-60 g/l for crystallization experiments.

Expression of *cafIA* deletion constructs, cell fractionation and surface immunofluorescence

Plasmid p12R encodes all 3 genes required for assembly of surface F1 – *cafIM* chaperone, *cafIA* usher and *cafI* subunit together with a truncated *cafIR* regulator. The operon is repressed with 0.6% glucose and induced in the absence of glucose. Overnight cultures of *E. coli* DH5α/ p12R or *E. coli* DH5α carrying the appropriate plasmid (p12R-A-mΔ1 etc) were washed and subcultured 1 in 10 in LB containing 100 mg/l ampicillin and 0.6 % (w/v) glucose, or no glucose as appropriate, and grown for 2 h at 37°C with shaking at 250 rpm. Crude outer membrane preparations (OMs) were isolated from washed, induced cells by differential centrifugation and washing at 20000 g for 30 min at 4°C following lysis with a French Press (Aminco) at 16,000 lbs/in² and removal of any unlysed whole cells. OMs were re-suspended at a final ratio of 0.5 ml PBS to 100 OD_{600nm} equivalent units of bacterial cells (where 1 OD_{600nm} equivalent unit corresponds to the total cells recovered from 1 ml bacterial culture with an OD_{600nm} of 1.0). OMs were further purified by equilibrium density centrifugation on 30 – 65 % (w/w) sucrose gradients in phosphate buffered saline, pH 7.4 with centrifugation at 116000 g for 28h at 15 °C. Co-equilibration of Caf1A constructs with outer membranes was confirmed by immunoblotting for Caf1A and the integral outer membrane protein OmpA.

Periplasmic fractions were isolated from induced cells by osmotic shock [22], using 3 OD_{600nm} equivalent units to produce 100 μl periplasmic extract. Surface exposed Caf1 was

quantitated in a surface immunofluorescence assay [9] using 1:200 dilution of rabbit anti-Caf1SC antibody and 1:100 dilution of goat anti-rabbit IgG fluorescein conjugate (Sigma). Average fluorescence per 1 OD_{630nm} bacterial cells was calculated from three dilutions of each sample and triplicate cultures of each strain were quantitated to derive the average fluorescence, with standard deviation, per OD_{630nm} bacterial cell. *E. coli* DH5a/pUC19 was used as a background control.

Crystallization and structure determination

Crystallization was performed by the hanging-drop vapour-diffusion method at 293 K. For U232-320, large rhomboidal or more complex multi-face crystals were obtained in drops with either 20% PEG3350 in 0.2 M NaSCN, pH 6.9, or 10% PEG8000 in 0.2 M MgCl₂, 0.1 M Tris-HCl, pH 7.0. The crystals belong to space group C222₁ with unit cell dimensions $a=30.0$ Å, $b=112.0$ Å, $c=115.1$ Å, and two copies of the UMD per asymmetric unit. Diffraction data were collected under liquid-nitrogen cryoconditions at 100 K. To avoid damage on freezing, crystals were soaked for 30-60 s in cryoprotection solution prepared by mixing 1 part of precipitant solution with 1 part of 25% PEG400. Crystals were flash-cooled by rapidly moving them into the cold nitrogen stream or by dipping them in liquid nitrogen. Native X-ray diffraction data were collected on beam-line ID14eh1 ESRF, France, using an ADSC Q4 CCD-based detector (ADSC). To obtain a derivative for heavy metal phasing, crystals were incubated with precipitant solution containing 2.5 mM KAuCl₄ for 24 h. Au SAD experiments were performed at BM14U using an XFlash Detector (Bruker). Data were collected, processed, and reduced using MOSFLM and SCALA [23]. Using PHENIX [24], heavy-atom parameters were obtained and refined, and initial phases calculated. The initial model was constructed using PHENIX and O [25]. Positional, bulk solvent, and isotropic B factor refinement was performed using REFMAC5 [26]. NCS restraints were applied during early cycles of modelling and refinement. Progress of refinement and selection of refinement schemes were monitored by the R_{free} for a test set comprising 4.6% of the data. Data collection and refinement statistics are given in Table 2.

Gel analysis

SDS-gel electrophoresis was performed using 20% polyacrylamide for analysis of UMD fragments, 14% for subunit and 8.5% for Caf1A usher. Samples were incubated for 5 min at 95°C in sample buffer containing 2% w/v SDS, unless otherwise indicated. Immunoblots were developed with 1:5000 dilution of rabbit anti-Caf1A antibody and 1:20000 dilution of goat anti-rabbit peroxidase conjugate using an ECL detection kit (Amersham).

CD spectroscopy

CD spectra were recorded on a J-810 dichrograph (Jasco) equipped with a PTC-423S temperature control system (Jasco) in 0.1 cm quartz cuvettes. Typically, five spectra recorded at a scanning speed of 50 nm/min with band width set to 1 or 2 nm were accumulated.

GdmCl- and temperature dependent folding transitions

To study unfolding of UMD, the native protein (1.4 g/l) was diluted 10-times with 20 mM phosphate buffer (pH 6.4) (B) containing different concentrations of GdmCl and incubated at 21°C for 16 hours. To study refolding, UMD was first denatured in 4.1 M GdmCl and incubated for one hour at 21°C. Unfolded UMD (1.1 g/l) was then diluted 10-times with the B buffer containing different concentrations of GdmCl and incubated at 21°C for 2-16 hours. The GdmCl-induced unfolding transition was measured by following the change of ellipticity at 217 nm, evaluated according to the two state model of folding using a six-parameter fit [27], and normalized. Free energy for unfolding of UMD in the absence of GdmCl was determined by linear extrapolation of the dependence of the free energy of unfolding on GdmCl concentration to 0 M [27]. To study temperature denaturation, UMD at a concentration of 0.15 g/l in the B buffer was heated from 20°C to 95°C in 5°C steps at a heating rate of 1°C/min. Spectra were recorded between each 5°C heating interval. To study the reversibility of temperature denaturation, the sample of UMD preheated to 95°C was allowed to cool to 21°C, followed by recording of a far-UV CD spectrum. Progress of protein melting was calculated from the change of ellipticity at 206 nm. Equilibrium constants in the transition region were evaluated based on a two state folding model. Van't Hoff enthalpy of melting was obtained from the temperature dependence of the equilibrium constants as in [28].

Binding assay

10 µl Ni-NTA agarose beads (QIAGEN) were incubated for 30 min with 75 µg Caf1M-Caf1 complex, carrying an N-terminal 6His tag on Caf1, in 100 µl binding buffer (5 mM imidazole, 25 mM NaH₂PO₄, 150 mM NaCl, pH 8.00). The beads were washed twice with 500 µl loading buffer to remove unbound complexes. To study binding between Caf1M-Caf1 and UMD or UND, the beads with bound Caf1M-Caf1 were added to 50 µl binding buffer containing 1 g/l UMD or UND and incubated for 30 min at 24°C in a shaker. The beads were recovered by centrifugation and washed twice with 500 µl loading buffer. Complexes were eluted from the beads with 30µl elution buffer (125 mM imidazole, 25 mM NaH₂PO₄, 150 mM NaCl, pH 8.0) and were analyzed with 20% SDS PAGE.

Multiple sequence alignment and structural superposition

The Caf1A-UMD sequence was used as query for an NCBI Position-Specific Iterated BLAST (Psi-BLAST) search using Expect and Psi-BLAST thresholds 8 and 0.003, respectively, against the non-redundant protein sequences database (nr) to identify a large number (>1000) sequences with E-values significantly better than the cutoff. A vast majority of these sequences were annotated as usher sequences. As this set contained many identical or near-identical sequences, a set of 147 unique sequences was extracted from the top 1000 hits and aligned using ALIGNX (InforMax).

Structures were superimposed using the INDONESIA program (D. Madsen, G. Kleywegt, and P. Johansson, unpublished [<http://xray.bmc.uu.se/~dennis/>]).

RESULTS AND DISCUSSION

Isolation and characterization of the middle domain of Caf1A

Initially the 213-320 fragment of Caf1A (U213-320) corresponding to the entire predicted UMD was expressed and purified from the *E. coli* cytoplasm. U213-320 was eluted in a single peak from a size exclusion column (Figure 1S supplemental data). Comparison of the retention coefficient (K_{av}) for U213-320 with those for globular proteins of known molecular weight suggested an apparent molecular weight of 16.1 kDa (Figure 1S, right panel). This value is 1.3 times higher than the sequence deduced weight for the U213-320, indicating that U213-320 is likely to exist as a monomer with abnormally large gyration radius in solution. CD measurement of purified samples of U213-320 revealed a typical for β -structural proteins far UV CD spectrum with a single ellipticity minimum of -6000 deg cm² dmol⁻¹ (per dmole of residue) at 217 nm (Figure 2a). Heating U213-320 to 95°C or adding 3 M GdmCl dramatically changed the spectrum, suggesting loss of the native structure. The spectrum was completely restored by sample cooling or 10 times dilution, respectively. The maximal CD changes caused by thermal and chemical denaturation were observed at 206 and 217 nm, respectively. These wavelengths were chosen to study thermal and chemically induced conformational transitions in U213-320 (Figure 2b and c). Both temperature and GdmCl dependence of the CD signal have the characteristic sigmoidal shape of cooperative transitions. Analysis of the transitions using the two-state model allowed estimation of thermodynamic parameters for U213-320 stabilization (Table 1). U213-320 is a temperature resistant protein ($T_m=70.4\pm 0.5^\circ\text{C}$, $\Delta H_m=313\pm 15$ kJ mol⁻¹) with thermodynamic stability at physiological conditions ($\Delta G_{295}=17.4\pm 1$ kJ mol⁻¹) typical for single domain globular proteins. Taken together, these results show that U213-320 is a stable β -structural monomeric unit capable of autonomous folding.

To test for the presence of unstructured regions in U213-320, we subjected purified U213-320 to limited proteolysis with trypsin and chymotrypsin. Proteolysis of U213-320 was observed at chymotrypsin concentrations below 1 $\mu\text{g/ml}$ and SDS-PAGE revealed concentration dependent disappearance of the intact protein with simultaneous appearance of a single truncated species (Figure 2d). Precise molecular weights for the proteolytic products, 9751 and 8658 Da for the trypsin and chymotrypsin proteolysis, respectively, were determined in a parallel analysis of the digestion reaction by mass spectroscopy. These values correlated with the masses of sequences Ile229-318 (9748 kDa) and Tyr239-318 (8655 kDa) of Caf1A usher, and cleavage of UMD between Lys228-Ile229 and Tyr238-Tyr239, for trypsin and chymotrypsin respectively. Both sites are close to the N-terminus of the recombinant UMD protein. Hence, these results identified an unstructured sequence at the N-terminus of U213-320.

We considered two possibilities: either U213-320 contains an extra sequence at the N-terminus, which does not participate in formation of the core structure of UMD, or, on the contrary, it lacks an essential sequence stabilizing its N-terminus and completing the UMD structure. To test these alternatives we created three additional constructs expressing U198-320, U222-320, and U232-320. The first construct includes an extra 15 amino acid-

sequence while the other two lack 9 and 19 residues, respectively (Figure 1). The gel filtration analysis of the constructs revealed a slight increase in mobility for U198-320 and mobility decrease for U222-320 and especially for U232-320, yielding apparent molecular weights of 16.5, 14.2, and 9.1 kDa for the constructs, respectively (Figure 1S). The constructs showed similar CD spectra with some increase of the positive signal below 205 nm, e.g. from about 0 for U198-320 to 4000 deg cm² dmol⁻¹ for U232-320 (Figure 3a). Thermal denaturation showed similar transitions for all the constructs, revealing slightly higher stability for U232-320, which melts at 2-3°C higher temperature than the other constructs (Figure 3b). Hence, the shortest version still contains the core structure of UMD. The slowest mobility during gel filtration and elevated thermal stability observed for this construct could be explained by the lack of flexible sequence at the N-terminus.

Structure of the middle domain of Caf1A

Crystallization screening for the shortest construct (U232-320) quickly revealed two conditions producing well-diffracting crystals. The crystal structure of U232-320 was solved to 2.8 Å resolution using Au SAD (Table 2). The structure consists of a symmetric domain-swapped dimer of similar UMD monomers (r.m.s.d. 0.69 Å for 69 Cα atom positions) (Figure 4a). The U232-320 monomer has the classical s-type immunoglobulin (Ig) fold [29], forming a β-sandwich with the participation of the N-terminus of the second monomer as one of the strands.

To model a non-swapped UMD domain, the N-terminal β strand of one monomer was built using coordinates of atoms of the corresponding segment in the second monomer of the dimer (Figure 4b). The resulting model consists of one three-stranded (A, B, and E) and one four-stranded (D, C, F, and G) β sheet (Figure 4c). No electron density was found for the first seven residues of U232-320. The A β strand starts after Pro 239 and consists of only four residues. The last, G, β strand ends at Tyr 310. The core structure is followed by an extension of five residues, Asp 311-Ala 316, that are involved in crystallographic contacts with a neighbouring dimer. The last four residues of U232-320 are unstructured. The β-sandwich is partially open at the beginning of the B, D, and F and ending of the A, C, E, and G β strands. The long loop connecting E and F β strands fills in this opening, contributing two residues into the hydrophobic core of the protein. The other loops are short, consisting of three-four residues.

The sequence U232-320 contains nearly 50% of the most conserved residues of usher proteins ([18], Figure 1). Most of these residues play clear structural roles as hydrophobic core (Val, Leu, Ile, Phe) or sharp turn (Gly) forming residues (Figure 4b). A highly conserved glutamic acid residue 298 links the two sheets and caps one end of the β-sandwich. Four highly conserved surface exposed residues, Ile 252, Arg 263 (Gln in most UMDs), Glu 291, and Asp 300 (Figure 4b), have no apparent structural role and may instead be important for UMD functionality.

N-terminal boundary of the middle domain of Caf1A

The high susceptibility of the A β strand to chymotrypsin digestion (Figure 2d, Figure 1c) is indicative of elevated mobility. This is further strengthened by the observation of the A β strand swapping in the UMD dimer (Figure 4a). To check whether this sequence is

crucial for protein folding we created a shorter version of UMD, U242-320, lacking the first half of the A β strand including two buried residues, Pro 239 and Tyr 241 (Figure 1c). U242-320 accumulated in cells in soluble form (Figure 3c). Analysis of the protein by gel filtration revealed that, as with the other purified usher fragments, U242-320 behaves as a monomer (Figure 1S supplemental data). U242-320 displayed a CD spectrum very similar to that of U232-320 (Figure 3a). However, the deletion caused a drop of more than 10 K in the melting temperature (Figure 3b). Further truncation of the N-terminal sequence by one additional residue, the surface exposed Gln 242, significantly decreased cytoplasmic levels of the expressed domain, whereas inclusion of the next buried residue, Trp 243, in the deletion or deletion of the entire A β strand (U246-320) completely abolished expression of the isolated UMD (Figure 3c). Hence, the A β strand is important for UMD folding, and despite its apparently elevated mobility should be viewed as a part of the core structure of this domain.

Structural similarity between the middle domain and F1 capsular subunit Caf1

The Caf1 donor strand complemented (Caf1dsc) fibre subunit (PDB accession code 1P5U) and UMD show similar overall topology (Figure 4c). To check the structural similarity between these proteins we superimposed their structures. Despite the clear difference in size of the proteins, superposition revealed significant structural similarity (Figure 4d). Sixty of seventy-one core C α atoms of UMD were superimposed with the corresponding atoms on Caf1dsc with an r.m.s.d. of 1.78 Å. Interestingly, the structurally similar part of Caf1dsc is in the top half of the subunit, as viewed in a vertically positioned fibre. The finding of high similarity with a specific part of the subunit indicates that UMD may represent a case of molecular mimicry, playing a role as a subunit-substituting protein, a 'dummy subunit'.

Membrane assembly of Caf1A constructs lacking the middle domain

Topology analysis (Figure 1) suggested a periplasmic location for UMD and that the UMD is bound on either side by membrane spanning β strands. As the UMD was shown to be an autonomously folding, soluble domain we reasoned that UMD could be deleted leaving an amphipathic β -barrel. Hence, to investigate the requirement of UMD for F1 assembly, a series of four internal deletions were created within the *caf1A* gene in the plasmid p12R. One deletion, *caf1A* Δ S236-V315 (A Δ 4) precisely removed the immunoglobulin like fold of UMD, while a second *caf1A* Δ Y215-V315 (A Δ 2) removed the entire predicted UMD domain leaving 2 residues following an absolutely conserved Gly at the end of the fourth predicted TM strand. As TM predictions for residues upstream of this fourth predicted TM strand were quite high, complicating identification of the fourth TM strand, *caf1A* Δ I205-V315 (A Δ 1) and *caf1A* Δ Q192-V315 (A Δ 3) were also created to ensure removal of the complete periplasmic middle domain and to remove the fourth TM strand, respectively. A protein of approximately the correct size (expected sizes of 78.856 kDa, 80.067 kDa, 77.414 kDa and 82.334 kDa for A Δ 1-4, respectively) was present in OM preparations from each of these constructs and notably the SDS-PAGE profile of A Δ 4 was very similar to that of wt (Figure 5a). Analysis of the mobility of Caf1A and Caf1A Δ 4 on SDS-PAGE following pre-incubation in sample buffer at different temperatures confirmed that Caf1A Δ 4 mimicked the heat-modifiable profile of wt Caf1A (Figure 5b). At 25°C and 55°C, three species were present, the fully denatured monomeric form, a faster migrating

doublet migrating approximately 20kDa faster than the fully denatured form and an oligomeric form, consistent with that of a dimer (181.4 kDa for wt Caf1A and 164.7 kDa for Caf1A Δ 4). PapC usher was reported to function as a dimeric twin-pore oligomer [13]. Thus, presence and resistance of this oligomeric Caf1A to heat denaturation (resistant at 55 °C, denatured at 100 °C) is good evidence that Caf1A Δ 4 is assembled in the membrane. The other deletion mutants also formed the heat sensitive oligomeric species although less of the faster migrating doublet, suggesting that they are also assembled but possibly in a less native conformation.

Presence of the middle domain is essential for Caf1 secretion

Ability of Caf1A Δ 2 and Caf1A Δ 4 to support surface assembly of F1 polymer was assessed by an *in vivo* immunofluorescence assay. In contrast to *E. coli* cells expressing the wild-type operon (p12R) and even cells grown under conditions where the wild-type operon is essentially repressed (*E. coli* p12R grown in 0.6% glucose), no Caf1 could be detected on the surface of *E. coli* cells expressing the mutated Caf1A protein, from either p12R-A Δ 2 or p12R-A Δ 4 (Figure 6a). Instead, accumulation of periplasmic Caf1, bound to chaperone, was observed with all four mutants (Figure 6b).

The middle domain does not bind chaperone-subunit complexes

Based on the observation of structural similarity between UMD and fibre inserted Caf1, we considered the possibility that UMD might be involved in orientating subunit within the pore by directly binding subunit and or chaperone. We have previously shown that the N-terminal domain of Caf1A (UND, residues 1-137) participates in recognition of Caf1M-Caf1 complexes by Caf1A (unpublished results). To directly compare binding of Caf1M-Caf1 complexes by UMD and UND we investigated the ability of agarose bound Caf1M-Caf1 complexes to trap soluble UMD or UND. In this system, we used a previously designed version of Caf1 subunit with the N-terminal donor sequence replaced by a 6-His tag. This modification arrests Caf1 polymerization, allowing accumulation of stoichiometric Caf1M-Caf16H (binary) complexes [6, 8]. Caf1M-Caf16H complexes were bound to Ni-NTA agarose beads (QIAGEN) via the 6-His tag. The Caf1M-Caf1 beads were then incubated with UMD or UND-containing solutions, recovered and bound protein eluted with imidazole. Analysis by SDS PAGE revealed a band of UND (Figure 6c, right panel), but no band of UMD co-eluting with Caf1M-Caf16H complex (Figure 6c, left panel). No binding between UMD and larger Caf1M-Caf1_n complexes could be observed in our ion-exchange-based binding tests (not shown).

Functional role of the middle domain

Our results suggest that UMD is not involved in directly binding Caf1M:Caf1 chaperone-subunit complexes but that its presence is required for assembly. There could be several different explanations for the requirement of UMD for export. The similar structure and cross-sectional dimensions of UMD and fiber inserted subunit suggest: that UMD might function as a dummy subunit, mimicking a fiber-incorporated subunit and plugging the translocation pore prior to initiation of assembly and secretion. This is consistent with the recent demonstration of the location of the PapC UMD across the central barrel of PapC usher [19]. Plugging the pore might be important to prevent leakage of large periplasmic components through the outer membrane. However, if this were the only role of UMD one

might expect Δ UMD-Caf1A to be functional in assembly and secretion. As this was not the case and the accumulated Caf1 in the periplasm was not accessible to external antibody, UMD appears to also fulfill some other role, possibly in maintaining the very large usher β -barrel in a translocation-competent conformation.

Structural comparison of middle domain from FGS and FGL chaperone/usher translocation systems

Caf1A and PapC usher proteins belong to different functional families of chaperone/usher secretion systems, the FGL and FGS chaperone/usher pathways, which assemble structurally and functionally different organelles. Subunits of these two types of organelles show no sequence homology and display significant structural differences [1, 2]. To elucidate structural adaptation of FGL and FGS chaperone/usher translocation systems for the secretion of different types of fibers we compared the available structures of the middle domains from Caf1A (the isolated UMD) and PapC (the domain located within the usher barrel).

Structural superposition of the two middle domains immediately revealed a single principal difference in their fold (Figure 7a): whereas the isolated Caf1A UMD has a classical seven-stranded Ig fold (Figure 4c), PapC UMD, lacks the A β strand and is an incomplete Ig fold with a six-stranded β sandwich, consisting of β sheets BE and DCFG. To compensate for the absent A β strand, B and E β strands in PapC are shifted by 1.4-2 Å in the direction of the edge β strand G of the opposite β sheet, creating a more compact structure. This conformational difference is reflected in elevated RMSD between corresponding C α atoms in B and E β strands of Caf1A and PapC (1.64 Å), contrasting with a low RMSD value estimated for strands of the opposite β sheet, DCFG (0.72 Å). While this could be a real topological difference between FGL and FGS ushers, it may reflect the two different states of UMD studied, with important functional implications, as discussed below. Another large difference between middle domains of Caf1A and PapC is found in the conformation of the loop between E and F β strands. In PapC, the central residue in the EF loop, Val 302, is involved in the hydrophobic core formation. To be able to interact with hydrophobic core residues, Val 302 pulls the central part of the EF loop towards the core structure. In contrast, the corresponding region in Caf1A consists of glycines, which are placed 3-5 Å higher above the core structure of the domain. This conformation is partially stabilized by the hydrogen bonding of the following Ser 289 with the G β strand, which in Caf1A is longer than that in PapC.

Despite the topological difference, the core segments of both UMDs (excluding the N terminal sequence) show high structural similarity: sixty-two of seventy-one core C α atoms of Caf1A UMD (residues 248-286 and 289-311) were superimposed with the corresponding atoms on PapC UMD (263-301 and 303-325) with an RMSD of only 1.2 Å (pairs of atoms with distance exceeding 3.7 Å were excluded from superposition). Donor strand complemented Caf1 and PapE subunits, transported via Caf1A and PapC ushers, respectively, are structurally much more dissimilar, displaying RMSD of 1.94 Å for 87 C α atoms superposed within 3.7 Å. The large difference between the UMD and subunit conservation suggests that the channel is universally adapted for the transport of organelle subunits of different structure and shape.

Identification of conserved UMD-channel wall contacts

Caf1A and PapC not only belong to the functionally different secretion systems (FGL and FGS, respectively), but they are also members of distantly related phylogenetic families of “classical” ushers (γ and π , respectively, [20]). Hence, the structural comparison of these distantly related ushers that share only 25 identical residues (33%) in their most conserved, middle domain sequences could be applied to identify conserved surface residues, contributing to the UMD-channel wall contact.

Remaut and colleagues [19] identified several ionic interactions, which they proposed may influence the position of the PapC UMD inside the translocation channel. Surprisingly, according to our superposition, none of the PapC residues involved in these ionic interactions (Arg 281, 303, 305, and 316) are identical or even similar to those in corresponding positions in Caf1A. Nevertheless, one such ionic interaction might still be formed in Caf1A. In PapC, Arg 305, within the middle domain, forms a salt bridge with Glu 467 on the channel wall, which in turn interacts with Arg 237 located within a β 5-6 hairpin of the wall. In place of Arg 305, Caf1A contains a highly conserved Glu 291 (Figures 4b and 7a). At the same time, instead of the channel wall glutamate, Caf1A contains several lysines in this region (KPKNK) that could form electrostatic interactions with Glu 291 and an aspartate residue replacing the PapC Arg 237 within the conserved hairpin sequence (FRS). In PapC, most of the ionic interactions are formed between UMD and either a small α -helix or β -hairpin extending from the channel wall. While alignments indicate presence of the similar structural features in Caf1A, differences in arrangement of ionic interactions may reflect localized differences in the structure of the channel and movement of UMD.

Caf1A and PapC middle domains contain a limited number of identical residues on their surface. Interestingly, a majority of these residues, including the highly conserved Ile 252 and Gly 276 (Figure 4b) and relatively conserved Thr 255, Pro 274, and Pro 277 are closely positioned in space (Figure 7). Together with Leu 272, which in PapC is replaced by Met 287, these residues form a hydrophobic patch on the surface (Figure 7b). The patch is further extended by the aliphatic part of side chains of residues Asp 300 in Caf1A and Asn 314 in PapC (Figure 7), which belong to a highly conserved Asp/Asn populated position in the usher alignment (Figures 1c and 4b). In PapC, all these residues form extensive hydrophobic interactions with the inner surface of the β barrel, dominated by the side chains of tyrosines 540, 601, and 622 or aliphatic parts of side chains of some polar residues (not shown). The high conservation of these residues as well as conformational similarity of their side chains in distantly related ushers suggests that this hydrophobic patch is likely to be involved in formation of UMD-channel wall contact in all “classical” usher proteins. These hydrophobic interactions, at the ‘tip’ of the middle domain, are likely to play a more important role in the lateral positioning of UMD inside of the translocation channel than the less conserved electrostatic interactions.

Hypothesis: Six/seven β stranded conformation switching and capping/uncapping of the usher channel

Our results showed that the A β strand is important for the stability of isolated UMD of Caf1A. They also inferred some mobility of this sequence. Threading of Caf1A and PapC sequences on the 6 and 7-stranded β sandwiches of UMD of PapC and Caf1A, respectively, suggested that both sequences are compatible with both folds (not shown). Thus, this raises the fascinating possibility that the 7-stranded topology presented here represents an alternate state of UMD common to all ushers. In this scenario, when located within the translocation channel the 6-stranded UMD would be stabilised by surface contact(s) with the channel wall; whereas, the released 7-stranded structure would be independently stable. In the PapC 6-stranded structure [19], the sequence corresponding to the additional Caf1A UMD β -strand plus the upstream sequence form a long flexible strand. However, both this sequence and the C-terminus of UMD are more or less fixed by trans-membrane β -strands. Thus, on formation of the additional β -strand, switching from the 6 to 7-stranded β sandwich, the N-terminal linker would both contract in length and change its tethering position leading to tension on the UMD and re-orientation of the entire domain, as schematically illustrated in Figure 8. Such a mechanism would allow a cooperative switching of Caf1A from a closed (plugged by UMD) pore conformation to an open one where UMD has moved out of the pore or docked inside the channel in a vertical orientation [19], Figure 8b. The critical requirement of residues in the PapC N-terminal linker sequence for pilus biogenesis [30] supports involvement of this sequence in opening of the channel. Our ongoing structural studies of the open conformation of the full-length usher will help elucidate the validity of this hypothesis.

ACKNOWLEDGEMENTS

This work was supported by grants from the Swedish Research Council (to AVZ and SDK) and BBSRC, UK (to SM).

REFERENCES

- 1 Zavialov, A. V., Zav'yalova, G. A., Korpela, T. and Zav'yalov, V. P. (2007) FGL chaperone-assembled fimbrial poly-adhesins: anti-immune armament of Gram-negative bacterial pathogens. *FEMS Microbiol. Rev.* **31**, 478-514
- 2 Thanassi, D. G., Saulino, E. T. and Hultgren, S. J. (1998) The chaperone/usher pathway: a major terminal branch of the general secretory pathway. *Curr. Opin. Microbiol.* **1**, 223-231
- 3 Choudhury, D., Thompson, A., Stojanoff, V., Langermann, S., Pinkner, J., Hultgren, S. L. and S. D. Knight (1999) X-ray structure of the FimC-FimH chaperone-adhesin complex from uropathogenic *Escherichia coli*. *Science* **285**, 1061-1066
- 4 Sauer, F. G., Futterer, K., Pinkner, J. S., Dodson, K. W., Hultgren, S. J. and Waksman, G. (1999) Structural basis of chaperone function and pilus biogenesis. *Science* **285**, 1058-1061

- 5 Sauer, F. G., Pinkner, J. S., Waksman, G. and Hultgren, S. J. (2002) Chaperone priming of pilus subunits facilitates a topological transition that drives fiber formation. *Cell* **111**, 543-551
- 6 Zavialov, A. V., Berglund, J., Pudney, A. F., Fooks, L. J., Ibrahim, T. M., MacIntyre, S. and Knight, S. D. (2003) Structure and biogenesis of the capsular F1 antigen from *Yersinia pestis*: preserved folding energy drives fiber formation. *Cell* **113**, 587-596
- 7 Zavialov, A. V., Tischenko, V. M., Fooks, L. J., Brandsdal, B. O., Aqvist, J., Zav'yalov, V. P., Macintyre, S. and Knight, S. D. (2005) Resolving the energy paradox of chaperone/usher-mediated fibre assembly. *Biochem. J.* **389**, 685-694
- 8 Zavialov, A. V., Kersley, J., Korpela, T., Zav'yalov, V. P., MacIntyre, S. and Knight, S. D. (2002) Donor strand complementation mechanism in the biogenesis of non-pilus systems. *Mol. Microbiol.* **45**, 983-995.
- 9 MacIntyre, S., Zyrianova, I. M., Chernovskaya, T. V., Leonard, M., Rudenko, E. G., Zav'yalov, V. P. and Chapman, D. A. G. (2001) An extended hydrophobic interactive surface of *Yersinia pestis* CafIM chaperone is essential for subunit binding and F1 capsule assembly. *Mol. Microbiol.* **39**, 12-25
- 10 Vetsch, M., Erilov, D., Moliere, N., Nishiyama, M., Ignatov, O. and Glockshuber, R. (2006) Mechanism of fibre assembly through the chaperone-usher pathway. *EMBO Rep.* **7**, 734-738
- 11 Remaut, H., Rose, R. J., Hannan, T. J., Hultgren, S. J., Radford, S. E., Ashcroft, A. E. and Waksman, G. (2006) Donor-strand exchange in chaperone-assisted pilus assembly proceeds through a concerted beta strand displacement mechanism. *Mol. Cell* **22**, 831-842
- 12 Nishiyama, M., Ishikawa, T., Rechsteiner, H. and Glockshuber, R. (2008) Reconstitution of pilus assembly reveals a bacterial outer membrane catalyst. *Science* **320**, 376-379
- 13 Li, H., Qian, L., Chen, Z., Thibault, D., Liu, G., Liu, T. and Thanassi, D. G. (2004) The outer membrane usher forms a twin-pore secretion complex. *J Mol Biol* **344**, 1397-1407
- 14 Nishiyama, M., Vetsch, M., Puorger, C., Jelesarov, I. and Glockshuber, R. (2003) Identification and characterization of the chaperone-subunit complex-binding domain from the type 1 pilus assembly platform FimD. *J. Mol. Biol.* **330**, 513-525
- 15 Nishiyama, M., Horst, R., Eidam, O., Herrmann, T., Ignatov, O., Vetsch, M., Bettendorff, P., Jelesarov, I., Grutter, M. G., Wuthrich, K., Glockshuber, R. and Capitani, G. (2005) Structural basis of chaperone-subunit complex recognition by the type 1 pilus assembly platform FimD. *EMBO J.* **24**, 2075-2086
- 16 Saulino, E. T., Thanassi, D. G., Pinkner, J. S. and Hultgren, S. J. (1998) Ramifications of kinetic partitioning on usher-mediated pilus biogenesis. *EMBO J.* **17**, 2177-2185
- 17 So, S. S. and Thanassi, D. G. (2006) Analysis of the requirements for pilus biogenesis at the outer membrane usher and the function of the usher C-terminus. *Mol. Microbiol.* **60**, 364-375
- 18 Capitani, G., Eidam, O. and Grutter, M. G. (2006) Evidence for a novel domain of bacterial outer membrane ushers. *Proteins* **65**, 816-823

- 19 Remaut, H., Tang, C., Henderson, N. S., Pinkner, J. S., Wang, T., Hultgren, S. J.,
Thanassi, D. G., Waksman, G. and Li, H. (2008) Fiber formation across the
bacterial outer membrane by the chaperone/usher pathway. *Cell* **133**, 640-652
- 20 Nuccio, S. P. and Baumler, A. J. (2007) Evolution of the chaperone/usher assembly
pathway: fimbrial classification goes Greek. *Microbiol Mol Biol Rev* **71**, 551-575
- 21 Martelli, P. L., Fariselli, P., Krogh, A. and Casadio, R. (2002) A sequence-profile-
based HMM for predicting and discriminating beta barrel membrane proteins.
Bioinformatics **18 Suppl 1**, S46-53
- 22 Chapman, D. A. G., Zavialov, A. V., Chernovskaya, T. V., Karlyshev, A. V.,
Zav'yalova, G. A., Vasiliev, A. M., Dudich, I. V., Abramov, V. M., Zav'yalov, V.
P. and MacIntyre, S. (1999) Structural and functional significance of the FGL
sequence of the periplasmic chaperone Caf1M of *Yersinia pestis*. *J. Bacteriol.* **181**,
2422-2429
- 23 CCP4 (1994) The CCP4 suite: programs for protein crystallography. *Acta.*
Crystallogr. D Biol. Crystallogr. **50**, 760-763
- 24 Adams, P. D., Grosse-Kunstleve, R. W., Hung, L. W., Ioerger, T. R., McCoy, A. J.,
Moriarty, N. W., Read, R. J., Sacchettini, J. C., Sauter, N. K. and Terwilliger, T. C.
(2002) PHENIX: building new software for automated crystallographic structure
determination. *Acta. Crystallogr. D Biol. Crystallogr.* **58**, 1948-1954
- 25 Jones, T. A., Zou, J. Y., Cowan, S. W. and Kjeldgaard, M. (1991) Improved
methods for building protein models in electron density maps and the location of
errors in these models. *Acta. Crystallogr. A* **47 (Pt 2)**, 110-119
- 26 Murshudov, G. N., Vagin, A. A. and Dodson, E. J. (1997) Refinement of
macromolecular structures by the maximum-likelihood method. *Acta. Crystallogr.*
D Biol. Crystallogr. **53**, 240-255
- 27 Santoro, M. M. and Bolen, D. W. (1988) Unfolding free energy changes
determined by the linear extrapolation method. 1. Unfolding of
phenylmethanesulfonyl alpha-chymotrypsin using different denaturants.
Biochemistry **27**, 8063-8068
- 28 Marky, L. A. and Breslauer, K. J. (1987) Calculating thermodynamic data for
transitions of any molecularity from equilibrium melting curves. *Biopolymers* **26**,
1601-1620
- 29 Bork, P., Holm, L. and Sander, C. (1994) The immunoglobulin fold. Structural
classification, sequence patterns and common core. *J. Mol. Biol.* **242**, 309-320
- 30 Henderson, N. S., So, S. S., Martin, C., Kulkarni, R. and Thanassi, D. G. (2004)
Topology of the outer membrane usher PapC determined by site-directed
fluorescence labeling. *J. Biol. Chem.* **279**, 53747-53754

FIGURE LEGENDS

Figure 1 Prediction of the middle domain of Caf1A

(a) Prediction of Caf1A topology using the HMM-B2TMR program [21]. Peaks on the plot (above the sequence scheme) predict trans-membrane β strands. Regions of the sequence that are involved in formation of trans-membrane β barrel and soluble domains are indicated. The graph on the top shows the similarity profile for the alignment of 147 usher proteins (see *Materials and Methods*). The graph was generated with ALIGNX with

fraction of consensus set to 0.99. **(b)** Model for the domain structure of the Caf1A usher. UTBD, usher trans-membrane β -barrel domain; UND, usher N-terminal domain; UMD, usher middle domain; UCD, usher C-terminal domain. **(c)** Constructs designed to identify the core structure (domain constructs) and function of UMD (Caf1A deletion constructs). All domain constructs had the same last amino acid residue G320 of the mature Caf1A (shown in red), but different residues at the beginning. Construct start residues are shown in red and labelled with U198, U213, U222, U232, and U242 for U198-320, U211-320, U222-320, U232-320, and U242-320 respectively. The smaller constructs with low (U243-320) or negligible (U244-320, and U246-320) levels of expression are not indicated. Triangle and arrow indicate the major cleavage sites for chymotrypsin and trypsin, respectively. Residue positions with 98% or more sequence identity in the alignment of 147 usher proteins are shown in bold. A conserved surface exposed hydrophobic position populated by either Ile or Val and conserved negatively or positively charged residues are indicated with h, (-), and (+) symbols, respectively. β sheets are shown with green arrows. The sequence corresponding to that shown to be critical for the function of the PapC usher [30] is underlined. Caf1A deletion constructs: deletions start with Q192 (A- Δ 3), I205 (A- Δ 1), Y215 (A- Δ 2), and S235 (A- Δ 4) (indicated in turquoise) and finish with V315 (turquoise). In each construct the two amino acids ArgSer, encoded by *Bg*/III, replace the deleted sequence, The dashed, underlined sequence indicates the last predicted transmembrane β strand adjacent to the N-terminus of UMD.

Figure 2 UMD is a stable, autonomous folding unit

(a) Far-UV CD spectra of U213-320 at 20°C (black continuous line), 95°C (red continuous line), cooled from 95 to 20°C (red dotted lines), in 3 M GdmCl (blue continuous lines), and refolded from GdmCl (blue dotted lines). **(b)** and **(c)** Thermal **(b)** and GdmCl-induced **(c)** unfolding of U213-320 followed by CD measurements at 206 and 217nm, respectively (normalized plots). Lines are best fits for a two-state transition. All measurements were performed at pH 6.4. **(d)** Limited proteolysis of the U213-320 construct. Denaturing SDS-PAGE of purified U213-320 before and after incubation with 0.1, 1.2, 2.3 $\mu\text{g ml}^{-1}$ chymotrypsin (lanes on the left) and 1.2, 2.3, 47.6 $\mu\text{g ml}^{-1}$ trypsin (lanes on the right). Digestion was performed in 20 mM phosphate buffer (pH 6.4) at 25°C for 30 min. The reaction was stopped by adding PMSF to a final concentration of 6.2 mM and by subsequent boiling. The molecular weights of the initial protein and major proteolytic fragments determined by mass spectrometry analysis on an Ultraflex TOF/TOF mass spectrometer (Bruker Doultonics) are indicated. The major cleavage sites deduced from the comparison of the data with the U213-320 sequence are shown in Figure 1.

Figure 3 The 239-242 amino acid sequence is important for the UMD folding

(a) Far-UV CD spectra of U198-320 (solid line), U213-320 (shot dashes), U222-320 (dots), U232-320 (dash-dot), and U242-320 (long dashes) at pH 6.4 and 20°C. **(b)** Thermal unfolding followed by CD measurements at 206 nm (normalized plot). U198-320, squares; U213-320 triangles; U222-320, diamonds; U232-320, open circles; U242-320, filled circles. **(c)** Coomassie blue stained SDS polyacrylamide gel of soluble cytoplasmic protein from the control BL21(DE3) cells (C) and BL21(DE3) cells expressing U242-320, U243-320, U244-320, and U246-320. The arrow indicates position of the band of UMD.

Figure 4 Structure of the UMD

(a) Crystal structure of U232-320 dimer (cartoon diagram). Two monomers in the dimer are shown in different colours. (b) Model of the U232-320 structure obtained by reverted beta-strand swapping (cartoon diagram, stereo view). The crystal structure of the U232-320 dimer was used as a basis for modelling. One of the monomers in the dimer is shown in magenta. To revert the beta-strand swapping, the N-terminal (A) beta strand (shown in cyan) was modelled using coordinates of atoms of the corresponding segment in the second monomer of the dimer. The loop connecting A and B beta strands was modelled in the O program [25]. The invariant and highly conserved residues (Figure 1c) with and without clearly identifiable structural function are shown with spheres in yellow and green, respectively. (c) Topology diagrams of UMD and Caf1dsc. Arrows indicate β -strands; the single helix (α A) in Caf1 is shown as a rectangle. Only major β -strands are shown. β -sheets ABE and DCFG of UMD and corresponding β -sheets ABED and CFG in Caf1 are shown in yellow and blue colour, respectively. (d) Superposition of UMD and Caf1dsc (PDB accession code 1P5U, fibre subunit) structures (stereo view). UMD, Caf1, and the Gd donor strand are shown in cyan, magenta, and red, respectively.

Figure 5 Outer membrane assembly of Caf1A mutants lacking UMD

(a) Immunoblot of Caf1A in OM preparations from *E. coli* Dh5 α carrying p12R (A-wt) or p12R encoding Caf1A with deleted UMD sequences: A- Δ 1 (Δ I205-V315), A- Δ 2 (Δ Y215-V315), A- Δ 3 (Δ Q192-V315), A- Δ 4 (Δ S235-V315). Samples pre-treated 100 °C for 5 min. (b) Heat modifiability of Caf1A mutants. Immunoblot of Caf1A mutants in OMs from *E. coli* strains as indicated and pre-treated in sample buffer for 5 min at stated temperature. Wild-type Caf1A, solubilised at 55°C to show oligomeric form, is shown for reference (a and b). Blots of SDS-8.5% polyacrylamide gels. White and black block arrows, denatured and oligomeric Caf1A, respectively; star, faster migrating doublet predominating in wt and Caf1A Δ 4.

Figure 6 Requirement of UMD for fibre secretion

(a) Surface immunofluorescence of *E. coli* cells expressing wt *caf* operon from p12R and grown in LB with no glucose (induction of *caf* operon), in LB with 0.6% glc (p12R glc, represses *caf* operon) or expressing *caf1A* Δ 2 or *caf1A* Δ 2 together with wt *caf1M* and *caf1* from the *caf* operon in p12R- Δ 2 or p12R- Δ 4. Surface Caf1 polymer was quantitated as described in Methods and expressed as fluorescence per OD600 bacterial cells. Error bars show the SD of independent triplicate cultures. (b) Stained SDS- 14% polyacrylamide gel of periplasmic extracts from *E. coli* p12R expressing wt *caf1A*, *caf1A* Δ 1 (205-315), - Δ 2 (215-315), - Δ 3 (192-315) or Δ 4 (235-315) or carrying pUC19 plasmid only. Arrow, accumulated Caf1 subunit. (c) UMD does not bind to chaperone-subunit complexes. Ni-NTA agarose beads carrying Caf1M-Caf1 complexes were added to UMD (U213-320) or UND solutions and after 30 min recovered by centrifugation. Complexes were washed from the beads with an imidazole containing buffer and analyzed with SDS-PAGE (see *Procedures* for details). *Right panel*: Sample of purified U213-320 (lane 2) and bead wash before (lane 1) and after incubation (lane 3) with U213-320. *Left pane*: Sample of purified UND (lane 1) and bead wash before (lane 2) and after incubation (lane 3) with UND. M, Caf1M; MF, Caf1M-Caf1; F, Caf1.

Figure 7 Comparison of Caf1A and PapC middle domains

(a) Structural superposition of U232-320 and 252-330 amino acid fragment of PapC (PDB accession number, 2VQI) [19] (stereo view). U232-320 and PapC-252-330 are in magenta and cyan, respectively, except for the A strand in U232-320 and the corresponding segment in PapC-252-330, which are shown in orange and green, respectively. Beta strands are labeled according to the immunoglobulin fold classification. N- and C-terminal ends of structured fragments are indicated. Core interacting residues in the A strand (Pro 239, Tyr 241, and Trp 243), selected surface residues in U232-320 and PapC-252-330 (Ile 252/267, Thr 255/270, Pro 274/289, Pro 277/292, Asp 300/Asn 314, and Glu 291/Arg 305), and Val 302 of PapC are shown in ball-and-stick. (b) Molecular surface of the region of U232-320 predicted to play key role in positioning the domain in the lateral orientation within the trans-membrane channel. Residues contributing to the hydrophobic patch on the surface of U232-320 are labeled. In (a) and (b), carbon is yellow; nitrogen is blue; oxygen is red.

Figure 8 Hypothetical role of the N-terminal conformational change in the capping/uncapping of the usher channel

(a) Model for UMD movement upon switching between 6 and 7 β stranded sandwich conformations. The model was created by superimposing the C α atoms of N and C terminal amino acids (spheres) of the experimental (cyan) and beta strand swap-reverted (magenta) structures of UMD. The conformations with an extended N-terminal region and the N-terminal sequence forming the A β strand are shown in cyan and magenta, respectively. (b) Contraction of the N-terminal region resulting in 7 β stranded sandwich may switch the usher from closed to two possible open conformations, UMD channel inserted (P) and channel free (P') proposed in [19]. The N-terminal region linking UMD with the trans-membrane β barrel is shown as a pink line (channel closed conformation) or an arrow indicating the A β strand (channel open conformation). The patch of surface residues identical in Caf1A and PapC, which in PapC form hydrophobic interactions with the inner wall of the trans-membrane β barrel, is indicated in yellow colour.

Table 1 Energetics of stabilization for UMD (213-320 amino acid fragment)

ΔH_m , kJ/mol	T_m , °C	$\Delta G_{N.U.}$, kJ mol ⁻¹	m , kJ mol ⁻¹ M ⁻¹	[GdmHCl] _m , M
313±15	70.4±0.5	17.4±1	13±0.7	1.30±0.05

Accepted Manuscript

THIS IS NOT THE VERSION OF RECORD - see doi:10.1042/BJ20080992

Table 2 Data and refinement statistics

Data Collection		
Data set	Native	Au derivative, peak
Wavelength (Å)	0.934	0.861
Resolution range (Å)	50-2.70 (2.85-2.7)	58-3.2 (3.37-3.2)
No. unique reflections	5685 (792)	3544 (510)
Multiplicity	4.9 (5.1)	19.9 (21)
Completeness	100.0 (100.0)	100.0 (100.0)
R_{merge}	0.065 (0.573)	0.122 (0.43)
$\langle I/\sigma(I) \rangle$	20.6 (2.2)	24.9 (5.5)

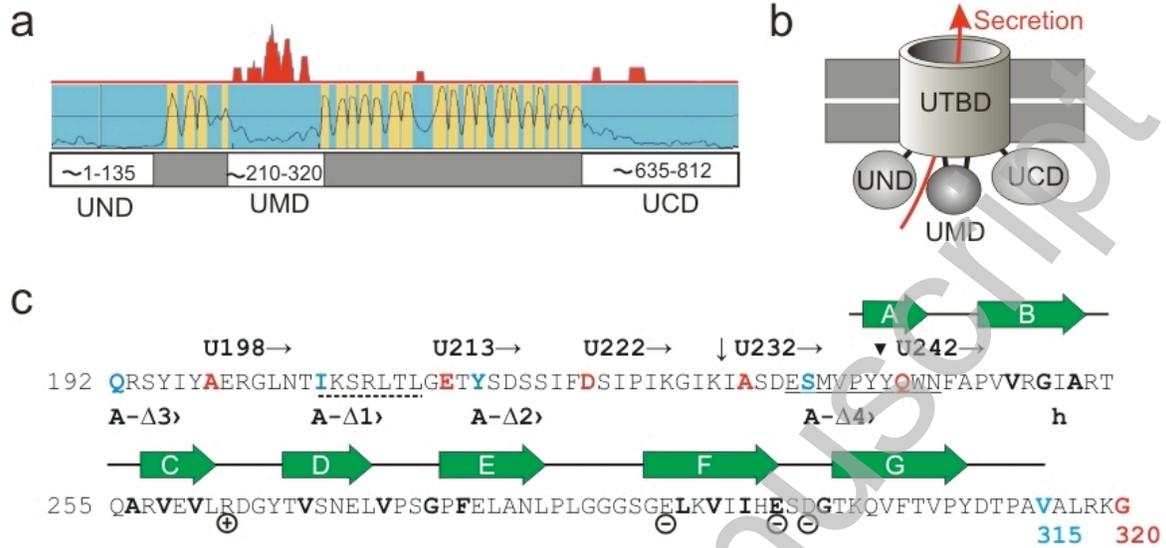
Refinement

Dataset	Native
Resolution (Å)	20-2.80
Number of reflections	
total	4613
work set	4401
test set	212
R_{work}	0.26
R_{free}	0.29
Number of atoms	
protein	1123
solvent	7
Average B-factor (Å ²)	54
R.m.s.d. stereochemistry	
bonds (Å)	0.005
angles (°)	0.849
R.m.s.d. B-factors (Å ²)	
main chain	0.173
side chain	0.249
Coordinate error* (Å)	0.333
Ramachandran plot	
Residues in most favoured regions (%)	85.2
Residues in additional allowed regions (%)	14.7

Values for high resolution shells in parenthesis. $R_{\text{merge}} = \frac{\sum_h \sum_i |I(h)_i - \langle I(h) \rangle|}{\sum_h \sum_i I(h)_i}$, where $I(h)$ is the intensity of a reflection h , \sum_h is the sum over all reflections and \sum_i is the sum over i measurements of reflection h . $R_{\text{work}} = \frac{\sum |F_o - F_c|}{\sum F_o}$ where F_o and F_c are the observed and calculated structure factors respectively. R_{free} is calculated for a test set of reflections randomly excluded from refinement. R.m.s.d. stereochemistry is the deviation from ideal values. R.m.s.d. B-factors is deviation between bonded atoms.

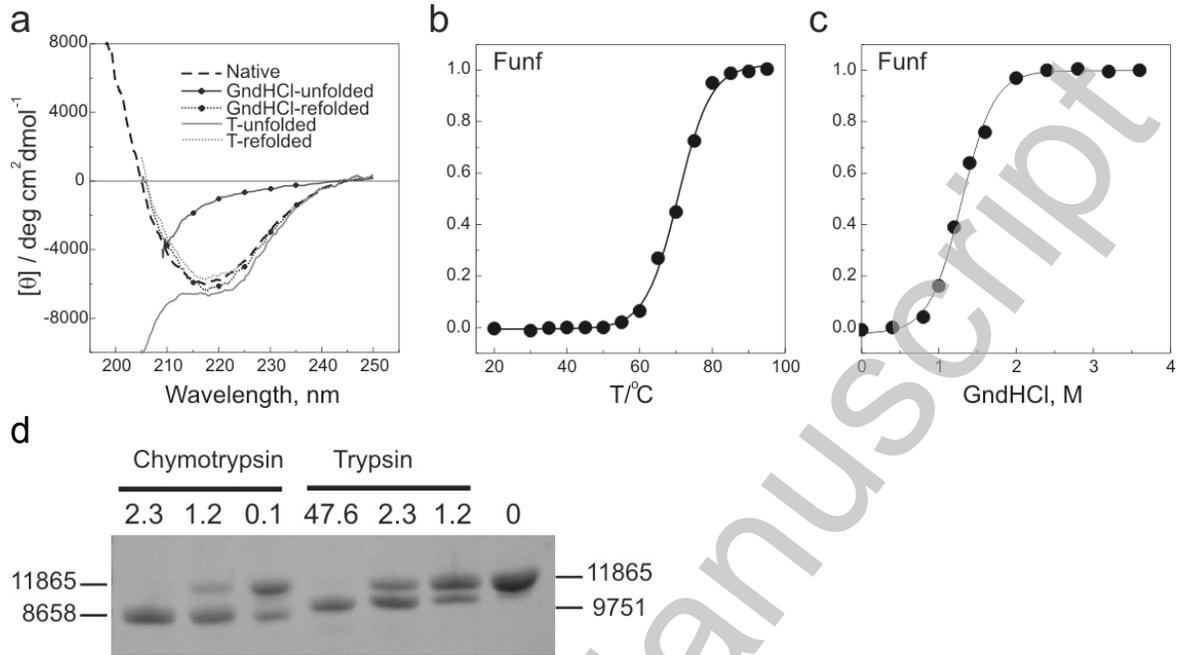
*Based on maximum likelihood.

Figure 1



THIS IS NOT THE VERSION OF RECORD - see doi:10.1042/BJ20080992

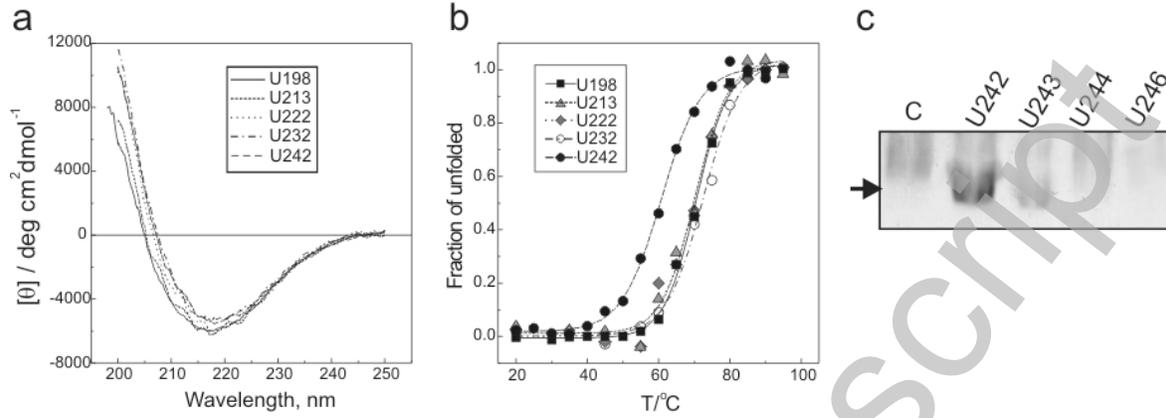
Figure 2



THIS IS NOT THE VERSION OF RECORD - see doi:10.1042/BJ20080992

Accepted Manuscript

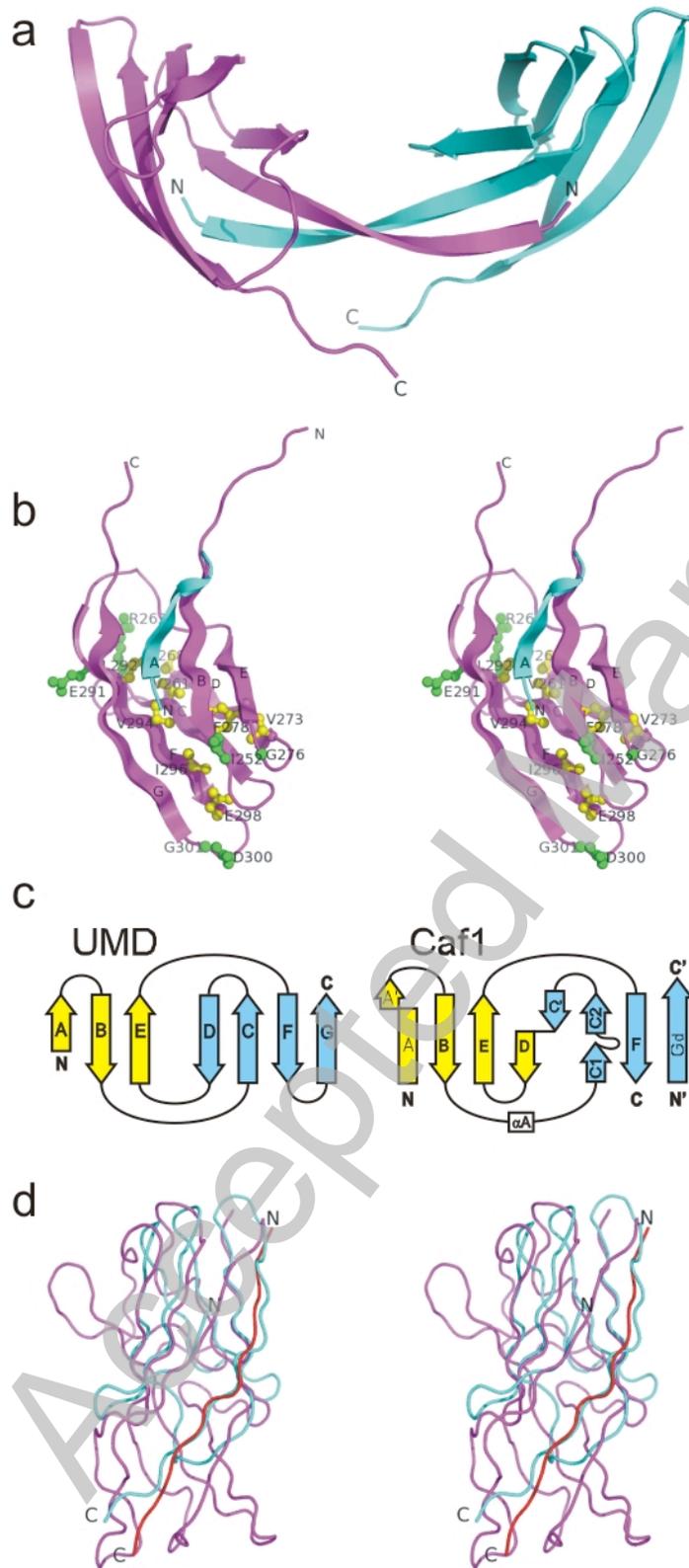
Figure 3



Accepted Manuscript

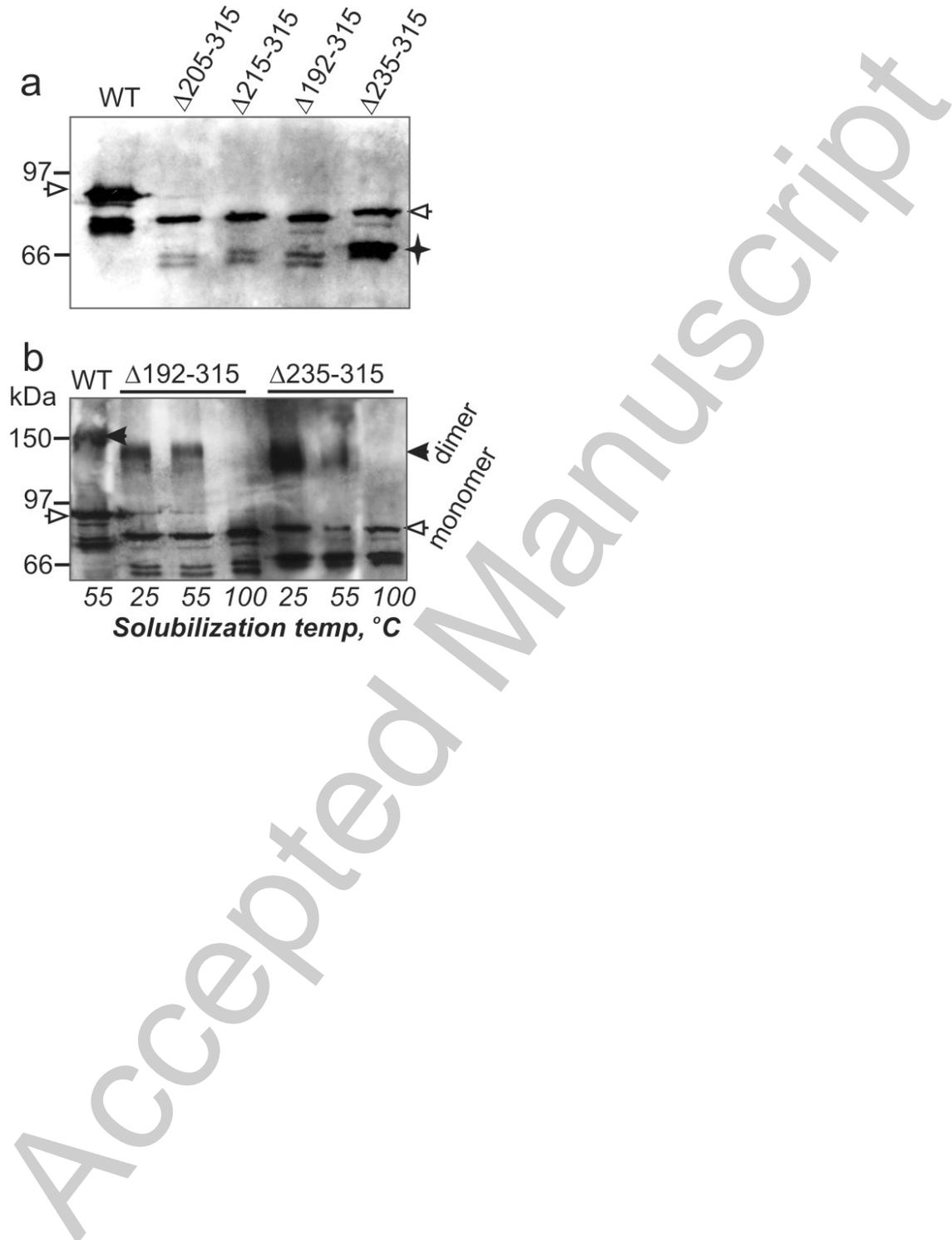
THIS IS NOT THE VERSION OF RECORD - see doi:10.1042/BJ20080992

Figure 4



THIS IS NOT THE VERSION OF RECORD - see doi:10.1042/BJ20080992

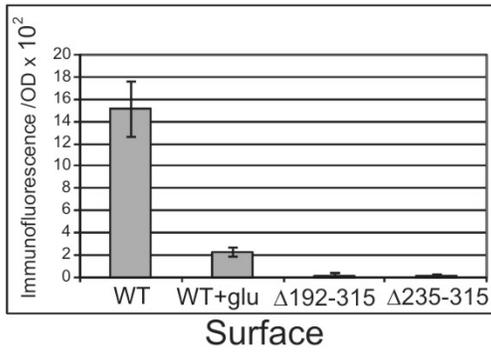
Figure 5



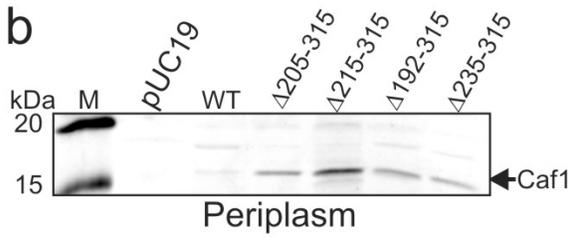
THIS IS NOT THE VERSION OF RECORD - see doi:10.1042/BJ20080992

Figure 6

a



b



c

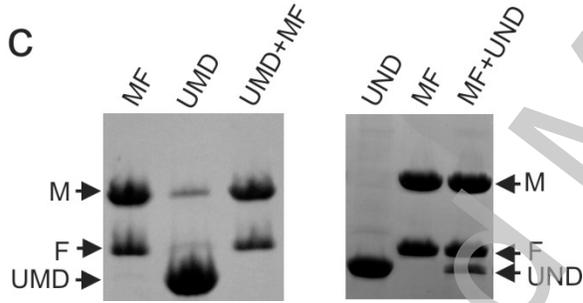
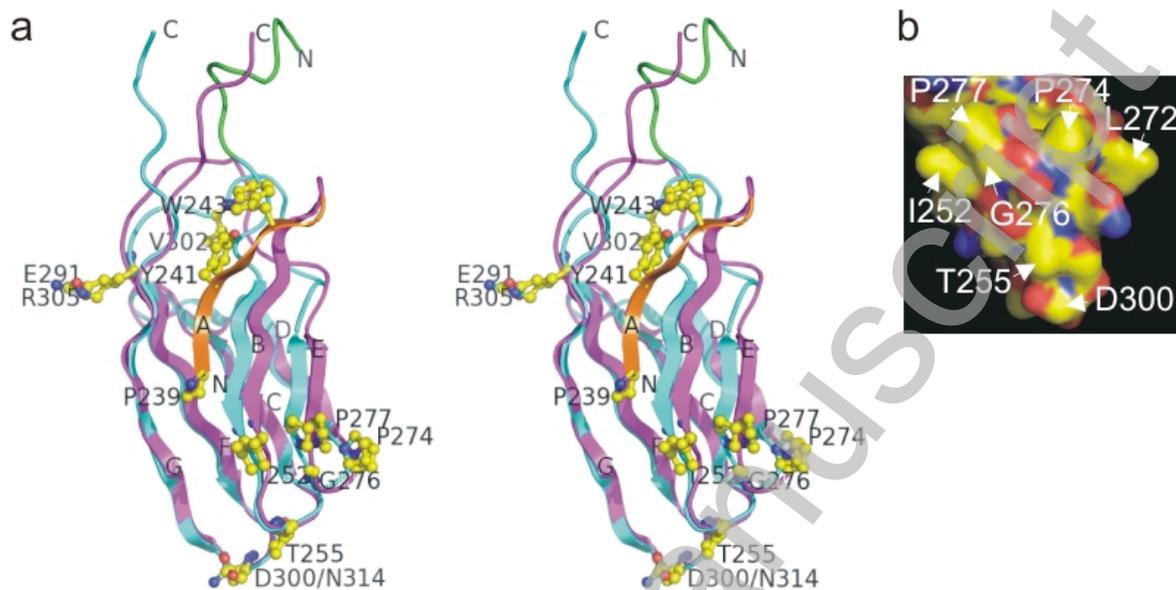
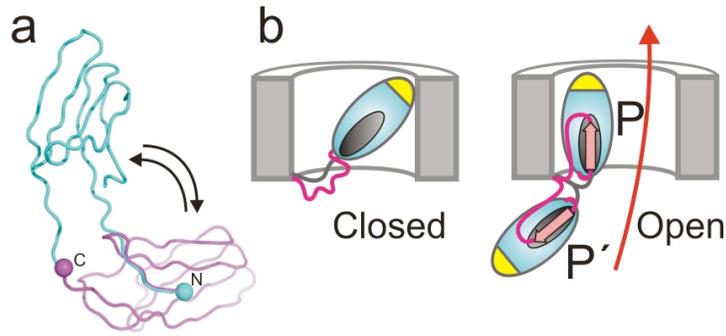


Figure 7



THIS IS NOT THE VERSION OF RECORD - see doi:10.1042/BJ20080992

Figure 8



Accepted Manuscript

THIS IS NOT THE VERSION OF RECORD - see doi:10.1042/BJ20080992