The extra-membranous domains of the competence protein HofQ show DNA-binding, flexibility and a shared fold with type-I KH-domains

Michael Tarry[a*1], Mari Jääskeläinen[a,b*], Annamari Paino[b], Heidi Tuominen[b], Riikka Ihalin[b#] and Martin Högbom[a#]

a. Center for Biomembrane Research, Department of Biochemistry and Biophysics, Stockholm University, SE-106 91, Stockholm, Sweden

b. Department of Biochemistry and Food Chemistry, FI-20014 University of Turku, Turku, Finland

* These authors contributed equally to this work

To whom correspondence should be addressed:
hogbom@dbb.su.se, tel: +46 8 162110, fax: +46 8 153679
riikka.ihalin@utu.fi, tel: +358 2 333 6851, fax: +358 2 333 6860

1. Current address: Department of Biochemistry, Francesco Bellini Building, McGill University, Montreal, Quebec, H3G 1Y6, Canada

Running Title: Structure and DNA binding of emHofQ

Abstract

Secretins form large oligomeric assemblies in the membrane that control both macromolecular secretion and uptake. Several *Pasteurellaceae* are naturally competent for transformation but the mechanism for DNA assimilation is largely unknown. In *Haemophilus influenzae*, the secretin ComE has been demonstrated to be essential for the DNA uptake. In closely related *Aggregatibacter actinomycetemcomitans*, an opportunistic pathogen in periodontitis, the ComE homolog HofQ is believed to be the outer membrane DNA translocase. Here we report the structure of the extra-membranous domains of HofQ at 2.3 Å resolution by X-ray crystallography. We also show that the extra-membranous domains of HofQ are capable of DNA binding. The structure reveals two secretin-like folds, the first of which is formed via means of a domain swap. The second domain displays extensive structural similarity to KH-domains, including the presence of a GxxG motif, which is essential for the nucleotide binding function of KH-domains, suggesting a possible mechanism for DNA binding by HofQ. The data indicate a direct involvement in DNA acquisition and provides insight into the molecular basis for natural competence.

Keywords: Membrane protein, DNA-uptake, Natural competence, Secretin, Structural dynamics.

Highlights

- Determined the structure of the competence protein HofQ extra-membranous domains

- emHofQ comprises two secretin-like folds, the first formed by a domain swap

- emHofQ binds dsDNA and is structurally similar to nucleotide-binding KH domains

- emHofQ contains a GGSG motif which may contribute to DNA binding

Introduction

Aggregatibacter (*Actinobacillus*) *actinomycetemcomitans* belongs to the family of Gram-negative *Pasteurellaceae*, and is one of the major opportunistic pathogens in localized aggressive periodontitis,[1,2,3] an infection that attacks and destroys the supportive tissues of the teeth. In addition to causing damage to the oral cavity, DNA and viable cells of some periodontal pathogens, including *A. actinomycetemcomitans*, have been found in atherosclerotic lesions[4,5] indicating a potential link to systemic diseases.

A. actinomycetemcomitans is one of at least 44 species of bacteria that are naturally competent for transformation,[6,7] allowing cells to take up external DNA. In *A. actinomycetemcomitans*, as in *Haemophilus influenzaeae* and *Neisseria gonorrheaea*, effective uptake of DNA involves a species-specific DNA sequence, termed the Uptake Signal Sequence (USS).[7,8,9] The USS, which is approximately ten nucleotides long, is frequently repeated in the genomes of these species, resulting in a preference for homo-specific DNA in natural transformation. In the majority of cases the DNA that is taken up most likely serves as a source of nutrition and nucleotides but, if it contains regions homologous to the host genome, it may be incorporated into the host bacterium's genome. This mechanism could be of relevance in the acquisition of antibiotic resistance genes or virulence factors.[10,11] A recent study suggests that natural competence combined with transformation may stabilize the otherwise hyper-dynamic core genome in *Pasteurellaceae*.[12]

Gram-negative bacteria need a complex machinery to move DNA from outside of the cell into the bacterial cytoplasm. The DNA must be bound from the extracellular milieu, traverse the outer membrane, followed by the cell wall and the cytoplasmic membrane.[13] Uptake is facilitated by a macromolecular transport system composed of many different proteins which display similarities to those utilized in assembly of Type IV pili and Type II secretion systems.[14] Double-stranded DNA (dsDNA) is transported across the outer membrane through a channel into the periplasm, where it becomes DNase resistant prior to subsequent transport of single-stranded DNA (ssDNA) across the cytoplasmic membrane.[15] *pilABCD* were the first genes to be identified as being required for competence in *A. actinomycetemcomitans*, and belong to the type IV pilus-like machinery.[16] These genes most likely code for prepilin (PilA), two transport proteins (PilB and C), and prepilin peptidase (PilD).[16] No *A. actinomycetemcomitans* outer membrane

3(23)

proteins that could be responsible for DNA binding and translocation have yet been reported. The *H. influenzae* outer membrane contains a DNA uptake system composed of multiple copies of the secretin ComE, which has been shown to be essential for DNA uptake.[17,18] The *A. actinomycetemcomitans* protein HofQ (UniProtKB Accession code: C9R226), shows some 70% sequence identity to *H. influenzae* ComE (Supplementary Figure 1), and is thought to perform the same function.[17,19]

Secretins are large homo-oligomeric assemblies built up of 50-70 kDa subunits, with 12-14 subunits forming a ring structure of approximately 100-150 Å in diameter.[20,21,22,23,24,25,26,27,28] They comprise a superfamily[29,30] and form components of several distinct secretion systems in the outer membrane, including the type-two secretion system (TIISS),[31,32] type-three secretion system (TIIISS)[26] and Type-IV pilus biogenesis system.[30,33,34] All secretins consist of two major regions. The C-terminal region, or β domain, is believed to be the major determinant in oligomer formation and stability.[35] It is highly conserved and predicted to contain several β-strands[36] embedded in the outer membrane to form the actual pore through which transport occurs.[21] The N-terminal periplasmic region displays conservation only in secretins from related secretion pathways[29] and may be involved in substrate recognition,[28,37] gating of the proposed channel[35,38] and DNA binding[39] as well as contributing to subunit oligomerization.[35,40]

The atomic structures of periplasmic fragments from two secretins, EscC from enteropathogenic *E. coli* and GspD from enterotoxigenic *E. coli*, have previously been determined.[37,40] EscC and GspD belong to the type III and type II secretion systems respectively and share little sequence conservation in their N-terminal regions. In spite of this, the two structures display remarkable structural similarities, highlighted by the presence of conserved N-terminal secretin-like folds in both.[37,40]

In this study we cloned and expressed the extra-membranous domains of HofQ (emHofQ) and determined the structure at 2.3 Å. We also show that emHofQ is capable of binding dsDNA. The structure revealed two secretin-like folds, one of which displays extensive structural homology to KH-domains involved in DNA-binding, suggesting a possible structural basis for HofQ's role in natural competence. The structure also exhibits far-

reaching structural flexibility with one of the N-terminal secretin-like folds formed via a domain-swap.

Results

Overall structure

emHofQ crystallized as a dimer in the asymmetric unit (Figure 1A). Analysis of the molecular interactions using the PDBe-PISA server[41] suggests that this dimeric assembly of emHofQ is the stable form in solution (buried area = 4550 Å^2, ΔG_{int} = -15.3 kcal/mol, ΔG_{diss} = 18.1 kcal/mol). In addition, the protein migrated at an apparent molecular weight of 41 kDa on a calibrated size-exclusion column during purification, also suggesting a dimeric assembly (the molecular weight of the emHofQ construct is 19 kDa). Still, a conclusive assignment is not possible because elongated molecules are expected to migrate at an apparently larger molecular weight in size-exclusion.

Each emHofQ polypeptide chain consists of two globular domains, containing a total of 7 β-strands and 4 α-helices (Figure 1B, C). The two domains are linked by approximately 25 residues, which are disordered in the structure, indicated by dashed lines in Figure 1A and C. Because of the length of the disordered segment, there is a theoretical possibility for an alternative connection within the crystal lattice. However, this would produce a very extended structure with protein molecules weaved together throughout the crystal lattice. Though this may be theoretically possible we find this arrangement very unlikely. The two chains, referred to here as A and B, can be superimposed with an rms deviation of 0.98 Å between 133 Cα atoms (Figure 1D). The main deviations are focused on residues 47-51 and 59-63, and derive from differences in crystal packing. Residues 47-51, which form the C-terminal end of the first α-helix in chain B, adopt a loop conformation in chain A (Figure 1D, inset). A reorientation of the side chain residues allows histidine 51 to form a hydrogen bond via a water molecule to threonine 148 of a symmetry related chain A polypeptide (Figure 2A). Residues 59-63 are disordered in chain A whereas in chain B they form a hairpin loop between the second and third β-strands, stabilised via interactions with residues 49 and 54 of a symmetry related B chain. Furthermore, the third β-strand in monomer B packs against its symmetry related counterpart to form an extended anti-parallel β-sheet (Figure 2B).

Domain 1

Domain 1 comprises residues 27-100 of emHofQ, although residues 27 and 27-28 are disordered in chains A and B respectively. Residues 1-26 correspond to the signal peptide which is expected to be cleaved subsequent to translocation across the cytoplasmic membrane in *A. actinomycetemcomitans* and were therefore omitted from the emHofQ construct.

When viewed in the context of individual monomers, domain 1 adopts an extended conformation containing two α-helices and four β-strands (Fig. 1B,C) that displays no structural homology to any other known protein fold. However, the dimer exhibits quasi-domain swapping,[42] with a closed conformation formed via means of a domain swap between residues 39-88 of the two chains. The closed conformation observed in the domain swap has the two α-helices packed against a four-stranded, anti-parallel β-sheet (Figure 3A). A search of the Dali database[43] found high structural homology to several acyl phosphatases (for example to PDB ID: 1APS[44] with a Dali Z-score of 9.5 and rmsd of 2.5 Å over 72 aligned residues) and also to a putative PII-like signalling protein (PDB ID: 3DFE (JCSG, unpublished, http://www.jcsg.org/) with a Z-score of 9.0 and rmsd of 2.3 Å over 71 aligned residues).

Domain 2

Domain 2 of emHofQ comprises residues 126-191 (127-191 in chain B) and contains two α-helices flanked by a three-stranded anti-parallel β-sheet (Figure 1B,C). It belongs to the N-1 secretin-like protein family PF03958, in Pfam.[45] If the closed, quasi-domain swapped conformation is considered for domain 1, the two domains of emHofQ can be superimposed with a Z score of 5.2 and rmsd of 2.2 Å, with 15 % sequence identity (Figure 3B, C).

The N-1 secretin-like fold observed in domain 2 is also found in protein components from other types of secretion systems. The second domain of emHofQ can be superimposed onto the N-1 and N-2 sub-domains of the enterotoxigenic *E. coli* TIISS component GspD[37] (PDB ID: 3EZJ) (Figure 4A), with Z scores of 9.0 and 8.1 and rmsd's of 1.8 Å and 1.7 Å respectively. Domain 2 shares 30 % sequence identity over 43 residues with the N-2 subdomain of GspD compared to 15 % sequence identity over 62 residues with the N-1 subdomain.

The structure of the periplasmic region of the TIIISS component EscC (PDB ID: 3GR5) from enteropathogenic *E. coli* has also been determined.[40] The second domain of this structure overlays onto domain 2 of emHofQ with a Z-score of 8.4 and rmsd of 2.1 Å, and with a sequence identity of 18 % over 60 aligned residues (Figure 4A, cyan). EscC is the sole component of the outer membrane ring in the Type III secretion system apparatus.[46]

Spreter *et al*. observed that the second domain of EscC shares a common fold with EscJ and PrgH. These are also components of the Type III secretion system but are localized to, and form part of, the membrane ring in the inner membrane.[40,47] Domain 2 of emHofQ was not found to superimpose particularly well to either EscJ (PDB ID: 1YJ7)[47] or PrgH (PDB ID: 3GR0),[40] with Z scores of 3.3 and 2.1 respectively. However, superimposition with the closed, domain swapped conformation of domain 1 produced improved structural alignments, with Z scores of 4.3 and 3.5 and rmsds of 2.6 Å and 2.3 Å respectively.

The N-1 secretin-like fold found in the second domain of emHofQ also displays similarities to that of the eukaryotic type-I KH (hnRNP K homology) domains. These domains are often involved in DNA or RNA binding,[48,49] which is of particular interest given the proposed role of emHofQ in DNA uptake. The second domain of emHofQ can be superimposed onto the KH domain of the neuronal splicing factor Nova-1 RNA binding protein (PDB ID: 2ANN) (unpublished) with a Z score of 6.7 and rmsd of 2.2 Å over 61 aligned residues (Figure 3B). All typical KH domains contain a conserved GxxG motif,[50] which is involved in nucleotide binding.[49] A GxxG motif is also found in emHofQ, with residues 149-152 having the sequence GGSG, and the two motifs in emHofQ and the KH domain are situated in close proximity structurally (Figure 4B inset).

DNA binding of emHofQ

To test the hypothesis that emHofQ may be directly involved in DNA binding electrophoretic mobility shift assay (EMSA) experiments were performed. emHofQ was incubated with a linearised vector either containing or lacking the USS sequence. In both instances emHofQ was observed to bind to dsDNA (Figure 5). The T4 gene 32 protein (singe-stranded DNA-binding) and bovine serum albumin were used as negative controls.

Discussion

emHofQ crystallized as a dimer in the asymmetric unit (Figure 1A). As described above, we do not believe that the dimeric interaction is an artefact of crystallization as the construct appeared to be dimeric in solution. However, it is still possible that the observed dimer is a result of the particular construct used. The emHofQ construct lacks the C-terminal, β-domain. In the full-length protein this region anchors the secretin in the outer membrane in a fixed position relative to the other subunits. This anchoring reduces the degrees of freedom available to the extra-membranous domain and it is known that the β-domain is a determinant in oligomer formation and stability of the full-length protein.[35] The presence of the C-terminal β-domain may prevent the formation of the dimer observed in the crystal structure.

This hypothesis is supported by the relative orientation of the two C-termini in the emHofQ crystal structure. As discussed in the introduction, secretins form channels in the outer-membrane, with 12-14 subunits thought to form the ring. Such channels have been modelled for the structures of peri-GspD and EscC,[37,40] as well as observed in the crystal packing of the inner membrane ring structure EscJ.[47] A key pre-requisite for any such ring structure is that the C-termini of individual subunits adopt a common orientation. This would be required for the correct insertion of the β-domain into the outer-membrane. With the dimer present in our crystal structure, the C-termini can indeed be orientated so that they would both localize to the plane of the outer membrane (Figure 1A, top panel). However, in this orientation, it is impossible to form a ring structure unless the emHofQ dimer constitutes the basic repeat element of a double-walled ring, which would be inconsistent with other models.[37,40]

Nevertheless, a 12-membered, oligomeric ring with a different architecture is found in the emHofQ crystals (Figure 6). Similar to the ring structure observed for EscJ,[47] the subunits are arranged into a continuous spiral but, unlike for EscJ, the C-termini switch orientation between planes of the ring in adjacent subunits. Thus, even if the ring were modelled to a flat, planar arrangement it would be impossible for the ring to associate with the outer-membrane with all C-termini orientated to the membrane surface. For this reason we believe that the particular ring structure observed does not represent the *in vivo* architecture and is in part imposed by the dimerization of the construct. However, it does support the

hypothesis that the N-secretin like fold contributes to oligomerization and ring formation in secretins.[40]

The dimer interface in emHofQ almost certainly promotes crystallization by stabilization of flexible secondary structure elements within the individual monomers. Evidence for this flexibility can be readily observed in the first domain, which is quasi-domain swapped (Figure 3A) and could reasonably be capable of accessing both the open and closed conformations. This would presumably be facilitated by two hinge regions located between residues 37-40 and 87-90, allowing residues 39-88 to fold back onto the rest of the domain, forming the secretin-like fold. The flexible nature of the first domain is also very apparent in residues 47-51 which adopt completely different conformations in the two monomers (Figure 1D, inset). Such flexibility appears to be an inherent property of the N-terminal, extra-membranous region of secretins. Uncomplexed periGspD alone gave poorly diffracting crystals and it was only in the presence of nanobodies, used as crystallization chaperones, that the structure could be successfully determined.[37] Negative stain electron microscopy of the full length secretins InvJ from *S. typhimurium*[40] and PulD from *Klebsiella oxytoca*[23] could not fully resolve the N-terminal region implying a certain degree of disorder for this region. The N-terminal region can also become disordered during detergent extraction of full-length secretins,[51] which again supports the idea that these regions are inherently flexible. This type of extensive structural flexibility, including inter-domain movements, as indicated in emHofQ, may be coupled to the mechanism of transition between the open and closed states of the outer membrane complex, an important outstanding question in the field.[52]

In this study we also demonstrate that *A. actinomycetemcomitans* emHofQ is capable of direct binding of dsDNA. DNA binding to the secretin PilQ has previously been demonstrated in both *N. meningitidis* and *Thermus thermophilus* HB27.[39,53] In the case of *Neisseria meningitidis* PilQ, this binding has been mapped to the extra-membranous region of the secretin. The N-terminus of PilQ in *N. meningitidis* contains many small basic repeat elements that are thought to be critical for DNA binding.[39] Such elements are missing from emHofQ, indicating that DNA binding must be mediated via a different recognition motif. A possible motif was identified through our structural analysis of emHofQ. The second domain of emHofQ displays high structural similarity to the nucleotide-binding KH domain of Nova-1 (Figure 4B). KH domains contain a conserved GxxG motif involved in

nucleotide binding[49,50] and a GGSG motif is found in similar location structurally in emHofQ. Furthermore, the second domain of emHofQ is related to domain 1 in both in structure and sequence (Figure 3B, C). The key exception is the GGSG motif and succeeding three residues, which occur as an insertion between α-helix 3 and β-strand 6. The fact that this insertion has been preserved within the context of the overall secretin-like fold points to it performing a functional role, possibly in DNA binding.

Although Nova-1 and many other KH domains bind RNA, DNA binding has also been documented.[54,55,56] In KH domains, the GxxG motif forms a protruding loop between two α-helices. In emHofQ, the GGSG motif does not precede a second α-helix but is instead followed by an extended loop (Figure 4B). Given the ability of residues 47-51 to exist as both a coil and α-helix it is possible that this region could adopt an α-helical conformation in the presence of DNA. Furthermore, the GGSG motif protrudes from the axis of the preceding helix, which appears to be a key structural feature of the motif in KH domains. The GxxG loop of KH domains typically contains a lysine or arginine in either the second or third position, however in some instances both or either positions are replaced by serines or threonines,[57] and a serine is found in the third position in emHofQ. KH domains bind DNA and RNA with relatively low affinity,[54,55,58] but the clustering together of KH domains increases nucleic acid recognition and specificity levels.[59] A similar clustering of DNA-binding domains could be achieved for HofQ given its oligomeric assembly and inherent flexibility.

A previously reported phylogenetic study of *Pasteurella* species identified HofQ from *A. actinomycetemcomitans* (annotated then as ComE) as a competence gene involved in DNA uptake.[19] Eight other HofQ/ComE homologs were similarly identified, although only seven were believed to be functional. A sequence alignment of the extra-membranous region of these proteins (Figure 7) displays very high sequence similarity for the second domain compared to the first, with strict conservation in many regions. This suggests that domain 2 is the functionally important of the two domains for DNA uptake and competence and implies that the first domain may play more of a structural role and perhaps be involved in gating. This would be consistent with the apparent dynamic properties of this domain. It is also interesting to note the lack of sequence conservation between β-strands 4 and 5, which corresponds to the unstructured region in HofQ. The proteins all contain glycine rich motifs that align with the GGSG motif in HofQ, with some also containing flanking or

internal lysine residues. It is possible that these motifs serve to introduce structural flexibility in the C-terminal part of α3, perhaps as part of a DNA-recognition and binding mechanism. The fact that emHofQ binds both USS and non-USS-containing dsDNA is interesting considering that an USS has been shown to increase uptake efficiency in a number of naturally competent bacteria.[7,8,9] Further in-depth studies of the DNA binding properties of emHofQ, including various DNA and protein variants are ongoing.

In conclusion, the structural and DNA binding properties of emHofQ suggest a direct involvement in DNA acquisition and provides insight into the molecular basis for natural competence.

Materials and Methods

Cloning, expression and purification
hofQ bases 81-585, corresponding to residues 27-195, encoding the extra-membranous part of HofQ (emHofQ) from *A. actinomycetemcomitans* strain HK1651 were amplified using PCR and cloned into a pET15b-vector (Novagen). The primers used were 5´-ATACTCGAGTTATTGCTCAATGGGTTTATC-3´ (reverse) and 5´-ATACATATGCAAAATCCGGTGTTTTCCATT-3´ (forward) with the restriction sites for XhoI and NdeI underlined. The cloning was performed using the Phusion™ High Fidelity DNA Polymerase (Finnzymes), DNA modifying enzymes from Fermentas, and *A. actinomycetemcomitans* HK1651 genomic DNA from ATCC (700685D-5). The resulting construct (pHofQ) was verified by sequencing with sequence data obtained from the web site of the *Actinobacillus* Genome Sequencing Project (http://www.genome.ou.edu/act.html).

pHofQ was transformed into *Escherichia coli* strain BL21-CodonPlus(DE3)-RIL (Stratagene) for production of emHofQ protein. 1 L of TB medium containing 50 μg kanamycin was inoculated with 10 mL of overnight culture. Cells were grown at 37 °C to an OD_{660} of 1.2 at which time they were induced by addition of Isopropyl β-D-1-thiogalactopyranoside (IPTG) to a final concentration of 1 mM and protein production was allowed to continue for 4 hours. Cells were harvested by centrifugation (7,000 x *g*, 20 minutes), resuspended in 50 mM Na_2HPO_4, 400 mM NaCl, 20 mM imidazole pH 7.5

(buffer A) containing DNaseI (Sigma Aldrich), lysozyme (Sigma Aldrich) and a Complete protease inhibitor tablet (Roche) and disrupted by several passes through an Emulsiflex C3 (Avestin) at an operating pressure of 15,000 psi. Debris was removed by centrifugation (18,000 x g, 30 minutes) and the cleared lysate applied to a nickel charged, 1 mL HisTrap column (GE Healthcare) equilibrated in buffer A. The column was washed extensively with buffer A containing 30 mM imidazole and then incubated overnight at room temperature with 400 units of thrombin (GE Healthcare) dissolved in buffer A to cleave off the hexa-histidine tag. Cleaved emHofQ was eluted from the column with buffer A, concentrated using a 10 kDa MWCO Vivaspin-6 (Sartorius) and applied to a Superdex-200 16/60 size exclusion chromatography column (GE Healthcare) equilibrated in 20 mM HEPES pH 7.2, 100 mM NaCl (buffer B). Fractions containing emHofQ, as assessed by SDS-PAGE, were pooled and concentrated.

SeMet emHofQ purification

SeMet substituted emHofQ was expressed following the method of Doublie[60]. Briefly, 3 mL of overnight pre-culture was pelleted and resuspended in 3 mL of M9 minimal media and used to innoculate 1.5 L of pre-warmed M9 minimal media. Cells were grown at 37 °C to an OD_{600} of 0.4 at which time the following amino acids were added: lysine, phenylalanine and threonine to 100 mg/L; isoleucine, leucine and valine to 50 mg/L; L-selenomethionine to 60 mg/mL. Cells were grown for a further 15 minutes at 37 °C at which time they were induced by addition of IPTG to a final concentration of 1 mM. The temperature was lowered to 25 °C and protein production allowed to continue overnight. Seleno-methionine labelled emHofQ was purified as described for the native emHofQ but in the presence of 5 mM beta-mercaptoethanol throughout.

Crystallization

Crystals were grown by the vapour diffusion method in sitting drops at 21 °C. Screens were set up using a Mosquito crystallization robot (TTP Labtech) using SwissSci triple-drop crystallization plates (Molecular Dimensions). emHofQ at a protein concentration of 17 mg/mL in buffer B was mixed with mother liquor in a ratio of 3:1, 1:1 and 1:3 protein:mother liquor to give a final volume of 0.2 μL and equilibrated against 20 μL mother liquor. Crystals appeared as hexagonal rods after five days in 2 M sodium formate, 0.1 M sodium acetate pH 4.6. SeMet substituted protein crystallized under the same

conditions. Crystals were cryo-protected by streaking through mother liquor containing 25 % glycerol and flash-cooled in liquid nitrogen.

Structure Determination

Diffraction data were collected at the European Synchrotron Radiation Facility (ESRF) in Grenoble, France. Peak, inflection and remote wavelength datasets for the MAD phasing were collected from a single SeMet substituted crystal at beamline BM30. The presence of selenomethionine was verified by an X-ray fluorescence scan around the selenium absorption edge prior to data-collection. The native dataset was collected at beamline ID14-1. All X-ray data indexing and scaling were performed with the programs Mosflm and Scala[61,62] or XDS.[63]

The crystals belonged to space group $P6_2$ with two molecules in the asymmetric unit. The structure of emHofQ was solved by multi-wavelength anomalous diffraction phasing using the software package Phenix.[64] This identified 2 out of 8 possible selenium atom positions and allowed the automated building procedures implemented within Phenix to build the majority of the residues. Manual inspection of the maps in Coot[65] showed them to be of good quality and this initial structure was then used as a molecular replacement model in the program CCP4-Molrep[66] with the higher resolution native dataset. This gave an excellent solution and the final structure was refined to 2.3 Å resolution with R_{work} = 0.205 and R_{free} = 0.237 using multiple iterative rounds of refinement with REFMAC5[67] and model building in Coot. Non-crystallographic symmetry (NCS) restraints were applied in REFMAC5, with restraints lifted for individual residues as required. TLS groups were defined by the TLSMD server.[68] The stereochemical quality of the model was verified using Coot and MolProbity.[69] Full data collection and refinement statistics can be found in Table 1.

ACCESSION NUMBERS: Coordinates and structure factors have been deposited in the Protein Data Bank with accession number: **2Y3M**.

Electrophoretic mobility shift assay (EMSA)

To test the binding of emHofQ to dsDNA, agarose gel EMSA experiments were performed to investigate the binding to linearized plasmids. The plasmids with/without USS were constructed as follows. Chromosomal DNA, from *A. actinomycetemcomitans* strain D7S, was used as a template for PCR. A 760 nucleotides long region of the

aggregation locus (bases 13266 to 14026), GI 66384466, was amplified by PCR. One USS sequence (-5´-AAGTGCGGT) and one specific restriction enzyme site was engineered to both primers. The sequence of the forward primer was 5´-GCGA<u>CATATG</u>AAGTGCGGTTTTTTTGTTTAAA TCAGTTTCAGTGAGATA (NdeI site underlined) and the sequence of the reverse primer was 5´-AGAA<u>GGATCC</u>AAGTGCGGTGTGTTTTTGTGGCGTTTTA (BamHI site underlined). Primers were purchased from Eurofins MWG Synthesis GmbH. The PCR samples were initially denatured at 98 ˚C for 30 s, then temperature cycled for 30 cycles with the following parameters: denaturation for 30 s at 98˚, annealing for 30 s at 69°C and elongation for 15 s at 72 °C. A final elongation step was performed for 5 min at 72°C. The PCR product and pUC19 plasmid were incubated for 16 h at 37°C with BamHI and NdeI and purified with the Qiaquick GelExtraction kit. The PCR fragments were ligated (T4 ligase, Fermentas) with a BamHI/NdeI double digested pUC19 plasmid for 12 h at room temperature. The ligation mixture was used to transform competent *E. coli* XLI-blue strain cells by electroporation. Transformants were selected on LB agar containing 100 µg/ml ampicillin. The purification of plasmids from liquid cultures was performed using the QIAprep spin miniprep kit (Qiagen). Plasmid constructs were verified by sequencing. Transformants, which contained the 0.76 kba USS insert in their plasmids (pUSS), were purified using the QIAGEN Plasmid Midi Kit and stored at - 20 °C. The plasmid without USS was constructed in the same manner except different forward (5'-TTAACATATGCGGCGGTTTTTTGTTTAAATCAGTTTCA) and reverse (5'-TATTGGATCCGTGTGTTTTTGTGGCGTTTT) primers without added USS sequences were used. The EMSA was performed by incubating the BamHI linearized plasmids (300 ng) with emHofQ (100 µg), T4 Gene 32 Protein (10 µg) (New England Biolabs) or BSA (100 µg) for 30 min at RT. Samples were loaded on a 0.8% agarose gel and run at 70 V for 2 h and bands were visualized using GelRed DNA stain (Biotium) under UV light.

Acknowledgements

Figure captions

Figure 1. Structure of emHofQ

A. Two views of the dimeric emHofQ. Chains A and B are coloured blue and light green

respectively and the location of the 25-residue disordered fragment is indicated by a

dashed line.

B. Chain B of emHofQ with domains 1 and 2 coloured light green and green respectively.

C. Topology diagram of emHofQ with the domains coloured as in B. Disordered regions of

the structured are indicated by a broken black line

D. Superposition of chains A and B, coloured as in A. Inset: magnified view of the overlay

between residues 47-51.

Figure 2. Differences in structure between chains A and B of emHofQ

A. Stereo view displaying the crystal packing between residues 47-51 of chain A (blue)

with a symmetrically related chain A (cyan). Side chain residues are shown as sticks and

labelled accordingly. The water mediating a hydrogen bond between residues His51 and

Thr148 is drawn as red sphere.

B. Stereo view of the extended β-sheet formed by the crystal packing of the second β-

strand from chain B (light green) with its symmetrically related counterpart (green).

Residues 59-63 and the symmetry related residues 49 and 54 are shown as sticks. Strand

β1 from chain A and a symmetrically related molecule are shown in blue and light blue respectively.

Figure 3. Domain swap of domain 1 and comparison with domain 2

A. Two views of a putative closed conformation domain 1 formed by a domain swap of residues 39-88 from chain B to chain A. Chain A and chain B residues are coloured blue and light green respectively.

B. Superposition of the closed conformation domain 1 shown in A with domain 2. Domains 1 and 2 are coloured light green and green, respectively.

C. Structure-based sequence alignment of the closed conformation domain 1 with domain 2. Secondary structure elements correspond to the structure of the closed conformation domain 1 (top) and domain 2 (bottom) and are coloured as in B. Sequences were aligned with ClustalW[70] and coloured with WebESPript[71] Key: White character inside red box – strict identity; red character – similarity between sequences.

Figure 4. comparison of domain 2 with structural homologs

A. Domain 2 (green) superimposed with the N-1 (yellow) and N-2 (magenta) domains of peri-GspD from enterotoxigenic *E. coli* (PDB ID: 3EZJ)[37] and the first domain (cyan) of the TIIISS protein EscC from enteropathogenic *E. coli* (PDB ID: 3GR5).[40]

B. Domain 2 (green) superimposed with the KH domain (pink) of neuronal splicing factor Nova-1 bound to RNA (orange) (PDB ID: 2ANN). The GGSG and GKGG motifs of domain 1 and Nova-1 respectively are shown as sticks. Inset: magnified view of the overlay between the two structures in the region of the GxxG motif.

Figure 5. binding of emHofQ to dsDNA

emHofQ binds both USS- and nonUSS-dsDNA in EMSA. The tested proteins, emHofQ (HQ) and control proteins T4 gene 32 protein (T4) and bovine serum albumin (BA) were incubated with linearized USS/nonUSS-plasmid after which the samples were run on an agarose gel and visualized under UV-light.

Figure 6.

Oligomeric ring assembly observed in the crystal-packing of emHofQ. The helical structure together with the architecture of the assembly suggests that this does not represent the *in vivo* structure of the transmembrane pore (see text).

Figure 7.

Alignment of emHofQ sequences from selected sequences as identified by Redfield *et al.* (2006). Secondary structure elements corresponding to the crystal structure of *A. actinomycetemcomitans* emHofQ are indicated and the position of the signal peptide shown. Sequences were aligned with ClustalW [70] and coloured with WebESPript [71] Key: White character inside red box − strict identity; red character − similarity between sequences. The position of the glycine-rich motif is marked with a green outline.

References

1.   Olsen, I., Shah, H. N. & Gharbia, S. E. (1999). Taxonomy and biochemical characteristics of Actinobacillus actinomycetemcomitans and Porphyromonas gingivalis. *Periodontol. 2000* 20, 14-52.

2.   Slots, J. & Ting, M. (1999). Actinobacillus actinomycetemcomitans and Porphyromonas gingivalis in human periodontal disease: occurrence and treatment. *Periodontol. 2000* 20, 82-121.

3.   Zambon, J. J. (1985). Actinobacillus actinomycetemcomitans in human periodontal disease. *J. Clin. Periodontol.* 12, 1-20.

4.   Kozarov, E. V., Dorn, B. R., Shelburne, C. E., Dunn, W. A., Jr. & Progulske-Fox, A. (2005). Human atherosclerotic plaque contains viable invasive Actinobacillus actinomycetemcomitans and Porphyromonas gingivalis. *Arterioscler. Thromb. Vasc. Biol.* 25, e17-18.

5.   Haraszthy, V. I., Zambon, J. J., Trevisan, M., Zeid, M. & Genco, R. J. (2000). Identification of periodontal pathogens in atheromatous plaques. *J. Periodontol.* 71, 1554-1560.

6.   Lorenz, M. G. & Wackernagel, W. (1994). Bacterial gene transfer by natural genetic transformation in the environment. *Microbiol. Rev.* 58, 563-602.

7.   Wang, Y., Goodman, S. D., Redfield, R. J. & Chen, C. (2002). Natural Transformation and DNA Uptake Signal Sequences in Actinobacillus actinomycetemcomitans. *J. Bacteriol.* 184, 3442-3449.

8.   Fitzmaurice, W. P., Benjamin, R. C., Huang, P. C. & Scocca, J. J. (1984). Characterization of recognition sites on bacteriophage HP1c1 DNA which interact with the DNA uptake system of Haemophilus influenzae Rd. *Gene* 31, 187-196.

9.   Goodman, S. D. & Scocca, J. J. (1988). Identification and arrangement of the DNA sequence recognized in specific transformation of Neisseria gonorrhoeae. *Proc. Natl. Acad. Sci. USA* 85, 6982-6986.

10.   Kroll, J. S., Wilks, K. E., Farrant, J. L. & Langford, P. R. (1998). Natural genetic exchange between Haemophilus and Neisseria: intergeneric transfer of chromosomal genes between major human pathogens. *Proc. Natl. Acad. Sci. USA* 95, 12381-12385.

11.   Maiden, M. C. (1998). Horizontal genetic exchange, evolution, and spread of antibiotic resistance in bacteria. *Clin. Infect. Dis.* 27 Suppl 1, S12-20.

12.   Treangen, T. J., Ambur, O. H., Tonjum, T. & Rocha, E. P. (2008). The impact of the neisserial DNA uptake sequences on genome evolution and stability. *Genome Biol.* 9, R60.

13. Chen, I. & Dubnau, D. (2004). DNA uptake during bacterial transformation. *Nat. Rev. Micro.* 2, 241-249.

14. Chen, I., Christie, P. J. & Dubnau, D. (2005). The ins and outs of DNA transfer in bacteria. *Science* 310, 1456-1460.

15. Burton, B. & Dubnau, D. (2010). Membrane-associated DNA transport machines. *Cold Spring Harb. Perspect. Biol.* 2, a000406.

16. Wang, Y., Shi, W., Chen, W. & Chen, C. (2003). Type IV pilus gene homologs pilABCD are required for natural transformation in Actinobacillus actinomycetemcomitans. *Gene* 312, 249-255.

17. Maughan, H., Sinha, S., Wilson, L. & Redfield, R. (2008). Competence, DNA uptake and transformation in Pasteurellaceae. (Kuhnert, P. & Christensen, H., eds.), pp. 79-98. Caister Academic Press, Norfolk, UK.

18. Tomb, J. F., el-Hajj, H. & Smith, H. O. (1991). Nucleotide sequence of a cluster of genes involved in the transformation of Haemophilus influenzae Rd. *Gene* 104, 1-10.

19. Redfield, R. J., Findlay, W. A., Bosse, J., Kroll, J. S., Cameron, A. D. & Nash, J. H. (2006). Evolution of competence and DNA uptake specificity in the Pasteurellaceae. *BMC Evol. Biol.* 6, 82.

20. Bitter, W. (2003). Secretins of Pseudomonas aeruginosa: large holes in the outer membrane. *Arch. Microbiol.* 179, 307-314.

21. Brok, R., Van Gelder, P., Winterhalter, M., Ziese, U., Koster, A. J., de Cock, H., Koster, M., Tommassen, J. & Bitter, W. (1999). The C-terminal domain of the Pseudomonas secretin XcpQ forms oligomeric rings with pore activity. *J. Mol. Biol.* 294, 1169-1179.

22. Burghout, P., van Boxtel, R., Van Gelder, P., Ringler, P., Muller, S. A., Tommassen, J. & Koster, M. (2004). Structure and Electrophysiological Properties of the YscC Secretin from the Type III Secretion System of Yersinia enterocolitica. *J. Bacteriol.* 186, 4645-4654.

23. Chami, M., Guilvout, I., Gregorini, M., Remigy, H. W., Muller, S. A., Valerio, M., Engel, A., Pugsley, A. P. & Bayan, N. (2005). Structural insights into the secretin PulD and its trypsin-resistant core. *J. Biol. Chem.* 280, 37732-37741.

24. Collins, R. F., Frye, S. A., Kitmitto, A., Ford, R. C., Tonjum, T. & Derrick, J. P. (2004). Structure of the Neisseria meningitidis outer membrane PilQ secretin complex at 12 A resolution. *J. Biol. Chem.* 279, 39750-39756.

25. Crago, A. M. & Koronakis, V. (1998). Salmonella InvG forms a ring-like multimer that requires the InvH lipoprotein for outer membrane localization. *Mol. Microbiol.* 30, 47-56.

26. Koster, M., Bitter, W., de Cock, H., Allaoui, A., Cornelis, G. R. & Tommassen, J. (1997). The outer membrane component, YscC, of the Yop secretion machinery of

Yersinia enterocolitica forms a ring-shaped multimeric complex. *Mol. Microbiol.* 26, 789-797.

27. Opalka, N., Beckmann, R., Boisset, N., Simon, M. N., Russel, M. & Darst, S. A. (2003). Structure of the Filamentous Phage pIV Multimer by Cryo-electron Microscopy. *J. Mol. Biol.* 325, 461-470.

28. Reichow, S. L., Korotkov, K. V., Hol, W. G. & Gonen, T. Structure of the cholera toxin secretion channel in its closed state. *Nat. Struct. Mol. Biol.* 17, 1226-1232.

29. Genin, S. & Boucher, C. A. (1994). A superfamily of proteins involved in different secretion pathways in gram-negative bacteria: modular structure and specificity of the N-terminal domain. *Mol. Gen. Genet.* 243, 112-118.

30. Martin, P. R., Hobbs, M., Free, P. D., Jeske, Y. & Mattick, J. S. (1993). Characterization of pilQ, a new gene required for the biogenesis of type 4 fimbriae in Pseudomonas aeruginosa. *Mol. Microbiol.* 9, 857-868.

31. Koster, M., Bitter, W. & Tommassen, J. (2000). Protein secretion mechanisms in Gram-negative bacteria. *Int. J. Med. Microbiol.* 290, 325-331.

32. Filloux, A. (2004). The underlying mechanisms of type II protein secretion. *Biochim. Biophys. Acta* 1694, 163-179.

33. Collins, R. F., Davidsen, L., Derrick, J. P., Ford, R. C. & Tonjum, T. (2001). Analysis of the PilQ secretin from Neisseria meningitidis by transmission electron microscopy reveals a dodecameric quaternary structure. *J. Bacteriol.* 183, 3825-3832.

34. Mattick, J. S., Whitchurch, C. B. & Alm, R. A. (1996). The molecular genetics of type-4 fimbriae in Pseudomonas aeruginosa--a review. *Gene* 179, 147-155.

35. Guilvout, I., Hardie, K. R., Sauvonnet, N. & Pugsley, A. P. (1999). Genetic dissection of the outer membrane secretin PulD: are there distinct domains for multimerization and secretion specificity? *J. Bacteriol.* 181, 7212-7220.

36. Pugsley, A. P. (1993). The complete general secretory pathway in gram-negative bacteria. *Microbiol. Rev.* 57, 50-108.

37. Korotkov, K. V., Pardon, E., Steyaert, J. & Hol, W. G. J. (2009). Crystal Structure of the N-Terminal Domain of the Secretin GspD from ETEC Determined with the Assistance of a Nanobody. *Structure* 17, 255-265.

38. Spagnuolo, J., Opalka, N., Wen, W. X., Gagic, D., Chabaud, E., Bellini, P., Bennett, M. D., Norris, G. E., Darst, S. A., Russel, M. & Rakonjac, J. (2010). Identification of the gate regions in the primary structure of the secretin pIV. *Mol. Microbiol.* 76, 133-150.

39. Assalkhou, R., Balasingham, S., Collins, R. F., Frye, S. A., Davidsen, T., Benam, A. V., Bjoras, M., Derrick, J. P. & Tonjum, T. (2007). The outer membrane secretin PilQ from Neisseria meningitidis binds DNA. *Microbiology* 153, 1593-1603.

40. Spreter, T., Yip, C. K., Sanowar, S., Andre, I., Kimbrough, T. G., Vuckovic, M., Pfuetzner, R. A., Deng, W., Yu, A. C., Finlay, B. B., Baker, D., Miller, S. I. & Strynadka, N. C. J. (2009). A conserved structural motif mediates formation of the periplasmic rings in the type III secretion system. *Nat. Struct. Mol. Biol.* 16, 468-476.

41. Krissinel, E. & Henrick, K. (2007). Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* 372, 774-797.

42. Liu, Y. & Eisenberg, D. (2002). 3D domain swapping: as domains continue to swap. *Protein. Sci.* 11, 1285-1299.

43. Holm, L. & Rosenstrom, P. (2010). Dali server: conservation mapping in 3D. *Nucleic Acids Res.* 38, W545-549.

44. Pastore, A., Saudek, V., Ramponi, G. & Williams, R. J. (1992). Three-dimensional structure of acylphosphatase. Refinement and structure analysis. *J. Mol. Biol.* 224, 427-440.

45. Finn, R. D., Mistry, J., Tate, J., Coggill, P., Heger, A., Pollington, J. E., Gavin, O. L., Gunasekaran, P., Ceric, G., Forslund, K., Holm, L., Sonnhammer, E. L., Eddy, S. R. & Bateman, A. (2010). The Pfam protein families database. *Nucleic Acids Res.* 38, D211-222.

46. Ogino, T., Ohno, R., Sekiya, K., Kuwae, A., Matsuzawa, T., Nonaka, T., Fukuda, H., Imajoh-Ohmi, S. & Abe, A. (2006). Assembly of the type III secretion apparatus of enteropathogenic Escherichia coli. *J. Bacteriol.* 188, 2801-2811.

47. Yip, C. K., Kimbrough, T. G., Felise, H. B., Vuckovic, M., Thomas, N. A., Pfuetzner, R. A., Frey, E. A., Finlay, B. B., Miller, S. I. & Strynadka, N. C. (2005). Structural characterization of the molecular platform for type III secretion system assembly. *Nature* 435, 702-707.

48. Siomi, H., Matunis, M. J., Michael, W. M. & Dreyfuss, G. (1993). The pre-mRNA binding K protein contains a novel evolutionarily conserved motif. *Nucleic Acids Res.* 21, 1193-1198.

49. Valverde, R., Edwards, L. & Regan, L. (2008). Structure and function of KH domains. *FEBS Journal* 275, 2712-2726.

50. Grishin, N. V. (2001). KH domain: one motif, two folds. *Nucleic Acids Res.* 29, 638-643.

51. Krehenbrink, M., Chami, M., Guilvout, I., Alzari, P. M., Pecorari, F. & Pugsley, A. P. (2008). Artificial binding proteins (Affitins) as probes for conformational changes in secretin PulD. *J. Mol. Biol.* 383, 1058-1068.

52. Karuppiah, V., Berry, J. L. & Derrick, J. P. (2011). Outer membrane translocons: structural insights into channel formation. *Trends Microbiol.* 19, 40-48.

53. Schwarzenlander, C., Haase, W. & Averhoff, B. (2009). The role of single subunits of the DNA transport machinery of Thermus thermophilus HB27 in DNA binding and transport. *Environ. Microbiol.* 11, 801-808.

54. Braddock, D. T., Baber, J. L., Levens, D. & Clore, G. M. (2002). Molecular basis of sequence-specific single-stranded DNA recognition by KH domains: solution structure of a complex between hnRNP K KH3 and single-stranded DNA. *EMBO J* 21, 3476-3485.

55. Braddock, D. T., Louis, J. M., Baber, J. L., Levens, D. & Clore, G. M. (2002). Structure and dynamics of KH domains from FBP bound to single-stranded DNA. *Nature* 415, 1051-1056.

56. Du, Z., Lee, J. K., Tjhen, R., Li, S., Pan, H., Stroud, R. M. & James, T. L. (2005). Crystal structure of the first KH domain of human poly(C)-binding protein-2 in complex with a C-rich strand of human telomeric DNA at 1.7 A. *J. Biol. Chem.* 280, 38823-38830.

57. Adinolfi, S., Bagni, C., Morelli, M. A. C., Fraternali, F., Musco, G. & Pastore, A. (1999). Novel RNA-binding motif: The KH module. *Biopolymers* 51, 153-164.

58. Liu, Z., Luyten, I., Bottomley, M. J., Messias, A. C., Houngninou-Molango, S., Sprangers, R., Zanier, K., Kramer, A. & Sattler, M. (2001). Structural basis for recognition of the intron branch site RNA by splicing factor 1. *Science* 294, 1098-1102.

59. Lunde, B. M., Moore, C. & Varani, G. (2007). RNA-binding proteins: modular design for efficient function. *Nat. Rev. Mol. Cell Biol.* 8, 479-490.

60. Doublie, S. (1997). Preparation of selenomethionyl proteins for phase determination. *Methods Enzymol.* 276, 523-530.

61. CCP4. (1994). The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D Biol. Crystallogr.* 50, 760-763.

62. Leslie, A. G. W. (1992). *Joint CCP4 + ESF-EAMCB Newsletter on Protein Crystallography* 12, 1-11.

63. Kabsch, W. (1993). Automatic Processing of Rotation Diffraction Data from Crystals of Initially Unknown Symmetry and Cell Constants. *J. Appl. Cryst.* 26, 795-800.

64. Adams, P. D., Gopal, K., Grosse-Kunstleve, R. W., Hung, L. W., Ioerger, T. R., McCoy, A. J., Moriarty, N. W., Pai, R. K., Read, R. J., Romo, T. D., Sacchettini, J. C., Sauter, N. K., Storoni, L. C. & Terwilliger, T. C. (2004). Recent developments in the PHENIX software for automated crystallographic structure determination. *J Synchrotron Radiat.* 11, 53-55.

65. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* 66, 486-501.

66. Vagin, A. & Teplyakov, A. (2010). Molecular replacement with MOLREP. *Acta Crystallogr. D Biol. Crystallogr.* 66, 22-25.

67. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D Biol. Crystallogr.* 53, 240-255.

68. Painter, J. & Merritt, E. A. (2006). TLSMD web server for the generation of multi-group TLS models. *J. Appl. Cryst.* 39, 109-111.

69. Chen, V. B., Arendall, W. B., 3rd, Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S. & Richardson, D. C. (2010). MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D Biol. Crystallogr.* 66, 12-21.

70. Chenna, R., Sugawara, H., Koike, T., Lopez, R., Gibson, T. J., Higgins, D. G. & Thompson, J. D. (2003). Multiple sequence alignment with the Clustal series of programs. *Nucleic Acids Res.* 31, 3497-3500.

71. Gouet, P., Courcelle, E., Stuart, D. I. & Metoz, F. (1999). ESPript: analysis of multiple sequence alignments in PostScript. *Bioinformatics* 15, 305-308.

Highlights

- Determined the structure of the competence protein HofQ extra-membranous domains

- emHofQ comprises two secretin-like folds, the first formed by a domain swap

- emHofQ binds dsDNA and is structurally similar to nucleotide-binding KH domains

- emHofQ contains a GGSG motif which may contribute to DNA binding

Table 1

Table 1. Crystallographic data collection and refinement statistics

Dataset	Native	Peak	Inflection	Remote
X-ray source	ESRF ID14-1	ESRF BM30	ESRF BM30	ESRF BM30
Wavelength (Å)	0.9334	0.9789	0.9791	0.9073
Space group	$P6_2$	$P6_2$	$P6_2$	$P6_2$
$a=b$ **(Å)**	119.6	120.2	120.1	120.4
c **(Å)**	51.3	51.3	51.4	51.5
Resolution (Å)	46-2.3 (2.42-2.3)	46-3.2 (3.4-3.2)	46-3.2 (3.4-3.2)	46-3.2 (3.4-3.2)
Completeness (%)	100 (99.9)	99.7 (99.4)	99.8 (99.8)	99.9 (100)
Measured reflections	137093 (13521)	39043 (5546)	38858 (5472)	39260 (5718)
Unique reflections	18861 (2736)	7151 (999)	7142 (987)	7187 (1026)
Multiplicity	7.3 (4.9)	5.5 (5.6)	5.4 (5.5)	5.5 (5.6)
R_{merge}	0.033 (0.610)	0.126 (0.487)	0.133 (0.500)	0.146 (0.602)
<I/sigma(I)>	16.6 (2.3)	10.8 (3.4)	10.2 (3.2)	9.9 (2.9)
Refinement				
Resolution range (Å)	46-2.3			
R_{work}/R_{free}	0.205/0.237			
rmsd bond lengths (Å)	0.007			
rmsd bond angles (°)	1.04			
Residues modelled (range)	Chain A: 135 (27-58, 64-100, 126-191) Chain B: 138 (28-100, 127-191)			
Water molecules modelled	51			
Other molecules modelled	Glycerol: 2 Formate: 1			
Average B (protein) (Å)	Chain A: 48.3 Chain B: 51.3			
Average B (water) (Å)	49.5			
Ramachandran				
favoured (%)	98.1			
forbidden (%)	0.0			
PDB identifier	2Y3M			

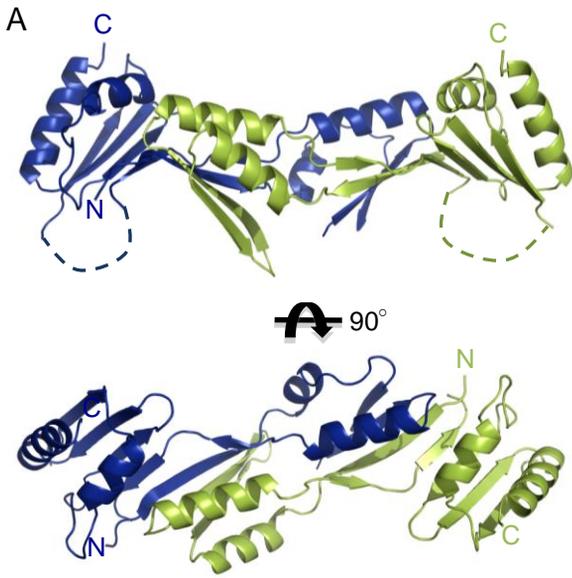Values for the highest-resolution shell are given in parentheses

Figure1
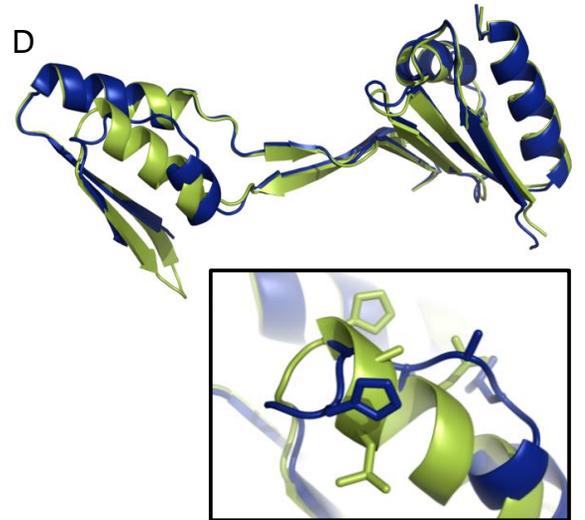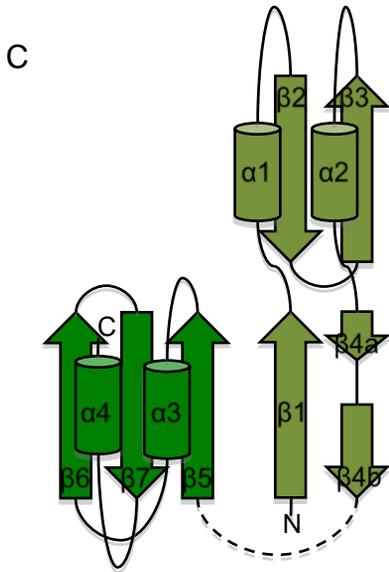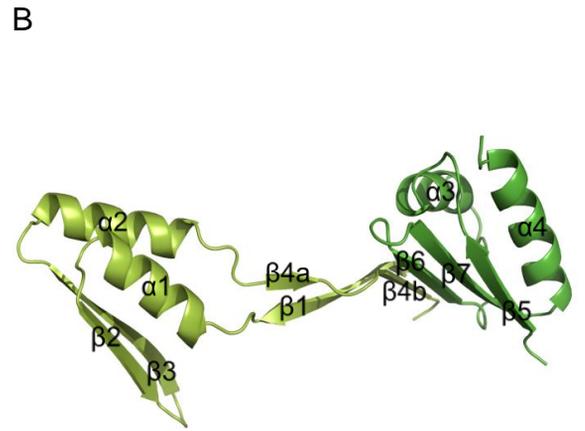
Figure2

A

B

Figure3

Figure4

Figure6

Figure7

A_actinomycetemcomitans
```
                                                              Signal peptide                           β1              α1
A_actinomycetemcomitans     1  ................MKSILSRRLKCGLFFGCFLSLAVQAEQ....NPVFSIRLKQAPLVPTLQQLALAHNTNL
M_succiniciproducens        1  MRAGRVDENKILKEYLPMESIISFGKKCGLFFGIFISSAFAGESGTFAERQFSIHLKKAPLVATLQQLALEQNANL
H_influenzae                1  ................MKKYF...LKCGYFLVCFCLPLIVFANPKTDNERFFIRLSQAPLAQTLEQLAFQQDVNL
P_multocida                 1  ................MWRAFR...KISLVY.FLCGVAYVGSSQAQDAEHFYLRLKQAPLVEMLQYLALQQHQDL
A_pleuropneumoniae          1  ................MRKMIWLLLFILTPVMAN......TFSLSLKNAPTAEILSYLAEEYGKNI
H_ducreyi                   1  ................MRILFSLFFFISLSLLAQ......PLSLSLKNAPTAEILSYLAEENMKNI
M_haemolytica               1  ................MWRIFIYFLCGFSLAYAN......TISIVLKNAPTSLIFTYLAEETGKNM
```

A_actinomycetemcomitans
```
          β2          β3                α2                β4a   β4b
A_actinomycetemcomitans    56  IIDDELQGTVSLQLENVDLDQLFRSVAKIKQLDLWQENGIYYFTKGDTN..TKFAGKMEEPF..PLSLPMAEEPAQ
M_succiniciproducens       77  VIDDELEGTLSLKLEKVNLERLFHSVAKLKNLSLHDKDIYYFTKNNLIEPSSIAGELKN....TENFTALSEP.N
H_influenzae               57  VIGDILENKISLKLNNIDMPRLLQIIAKSKHLTLNKDDGIYYLNGSQSGK.GQVAGNLTT...........NEP.H
P_multocida                56  LIDDHLEGTLSLQMKKTTFEKCLQSIARMKQLELHQEGKSYYLTS.....PSGVAANDTH..........HPTS
A_pleuropneumoniae         45  VLSDNIETNTTLRIENSDFDSVLKSITRANKLTSAYEKQIYFIGHKKDEKAATAIGV........NSDLL....KPK
H_ducreyi                  45  VLSDQIDKNTTLRIENSHFDEIVESIVRANQLSKRKEKQIYFIGHLKEDKAIEKID.........NGD......KPK
M_haemolytica             45  VLDDNIETKSTLRLENKSVDEIFKTLGKVNKLSLTQEDDIVYI.HKKEEKAADLTPLPIMNLQNNGQNTPLVTAPK
```

A_actinomycetemcomitans
```
          β5              α3              TT  β6          β7            α4
A_actinomycetemcomitans   128  LNTATIKLHFAKASEVMKSLTGG..SGSLLSPNGSITFDDRSNLLLIQDEPRSVRNIKKLIKELDKPIEQIA
M_succiniciproducens      148  LVSTTVKLHFAKASEVMKSLTSG..TGSLLSPVGSVSFDERSNQLLIQDERRSLQNIKNIIAQLDKPIEQIA
H_influenzae              120  LVSHTVKLHFAKASELMKSLTTG..SGSLLSPAGSITFDDRSNLLVIQDEPRSVQNIKKLIAEMDKPIEQIA
P_multocida               115  LMTSSIKLHFAKAAEVVKSLTSG..QGSLLSVGGSLSFDERTNLLLIQDEPQSIQRTKALVAEMDKPIEQIA
A_pleuropneumoniae        110  LITKTIKLDYAKAAEVIESLTKG..SGHFLSENGYLHFDDRSNLIIKDSPESMKNIVQLIRNLDKPTEQIA
H_ducreyi                 107  LITQTIKLDYAKASEVIESLTKGKGNGVLSEQGYLYFDERSNSLIIKDSPDSMKHILALIKNLDKPTEQIA
M_haemolytica             120  LITKTLKLHYAKASEVIESLTKG..SGTFLSENGYIHFDERSNSLIVKDSAKSLKNIENLVKQLDQPTEQIA
```