

# Structure of the formate transporter FocA reveals a pentameric aquaporin-like channel

Yi Wang<sup>1\*</sup>, Yongjian Huang<sup>2\*</sup>, Jiawei Wang<sup>2\*</sup>, Chao Cheng<sup>2</sup>, Weijiao Huang<sup>2</sup>, Peilong Lu<sup>1</sup>, Ya-Nan Xu<sup>3</sup>, Pengye Wang<sup>3</sup>, Nieng Yan<sup>2</sup> & Yigong Shi<sup>1</sup>

**FocA is a representative member of the formate–nitrite transporter family, which transports short-chain acids in bacteria, archaea, fungi, algae and parasites. The structure and transport mechanism of the formate–nitrite transporter family remain unknown. Here we report the crystal structure of *Escherichia coli* FocA at 2.25 Å resolution. FocA forms a symmetric pentamer, with each protomer consisting of six transmembrane segments. Despite a lack of sequence homology, the overall structure of the FocA protomer closely resembles that of aquaporin and strongly argues that FocA is a channel, rather than a transporter. Structural analysis identifies potentially important channel residues, defines the channel path and reveals two constriction sites. Unlike aquaporin, FocA is impermeable to water but allows the passage of formate. A structural and biochemical investigation provides mechanistic insights into the channel activity of FocA.**

Formate is a major carbon source for methanogenic archaea such as *Methanobacterium formicicum*<sup>1,2</sup>. It is also a signature metabolite of enteric bacteria under anaerobic conditions, during which pyruvate is cleaved by pyruvate–formate lyase (PFL) to yield acetyl CoA and formate<sup>3,4</sup>. As much as one-third of the carbon in the sugar was thought to be converted to formate during fermentative growth<sup>3,4</sup>, reaching a concentration of up to 20 mM in the cytoplasm<sup>5</sup>. Intracellular accumulation of formate may lead to a substantial decrease in cytoplasmic pH. In *Escherichia coli*, formate is metabolized by three formate dehydrogenases (FDHs)<sup>4–7</sup>, of which FDH-N and FDH-O have their active sites located in the periplasm and FDH-H in the cytoplasm as part of a multiprotein complex named formate hydrogenlyase (FHL). Formate must therefore be able to pass through the cell membrane. However, with a pK<sub>a</sub> of 3.75, formate exists predominantly in the deprotonated anionic form at physiological pH, necessitating a transport system. The integral membrane protein FocA was identified as a putative formate transporter in *E. coli*<sup>7</sup>.

FocA is a representative member of the formate–nitrite transporter (FNT) family (transporter classification 2.A.44), which was thought to transport structurally similar short-chain acids such as formate and nitrite<sup>8</sup>. Other known members of the FNT family include NirC of *E. coli* for nitrite uptake and export<sup>9–12</sup>, and FdhC of *M. thermoformicum* for formate uptake<sup>8,13</sup>. FocA homologues have been identified in bacteria, archaea, fungi, algae and parasites (Supplementary Fig. 1). Although members of the FNT family share considerable sequence homology, they have no apparent sequence similarity to other proteins. The structure and transport mechanism of the FNT family remain largely unknown.

We purified the full-length, recombinant FocA protein from *E. coli* and generated crystals under several conditions. However, these crystals persistently diffracted X-rays to low resolutions. We subjected the full-length FocA to V8 protease digestion, which removed 21 amino-acid residues from the amino terminus. The truncated FocA was

crystallized in two space groups,  $P2_12_12_1$  and  $P3_2$ , each with an improved diffraction limit. The structure in  $P2_12_12_1$  was determined by platinum-based single-wavelength anomalous dispersion (SAD) (Supplementary Table 1). The experimental electron density was of adequate quality (Supplementary Fig. 2a), and the final atomic model was refined to a free *R*-factor of 0.22 at 2.25 Å resolution (Supplementary Table 1 and Supplementary Fig. 2b–d). We also solved the structure of FocA in the  $P3_2$  space group at 3.2 Å resolution (Supplementary Table 1 and Supplementary Fig. 3a).

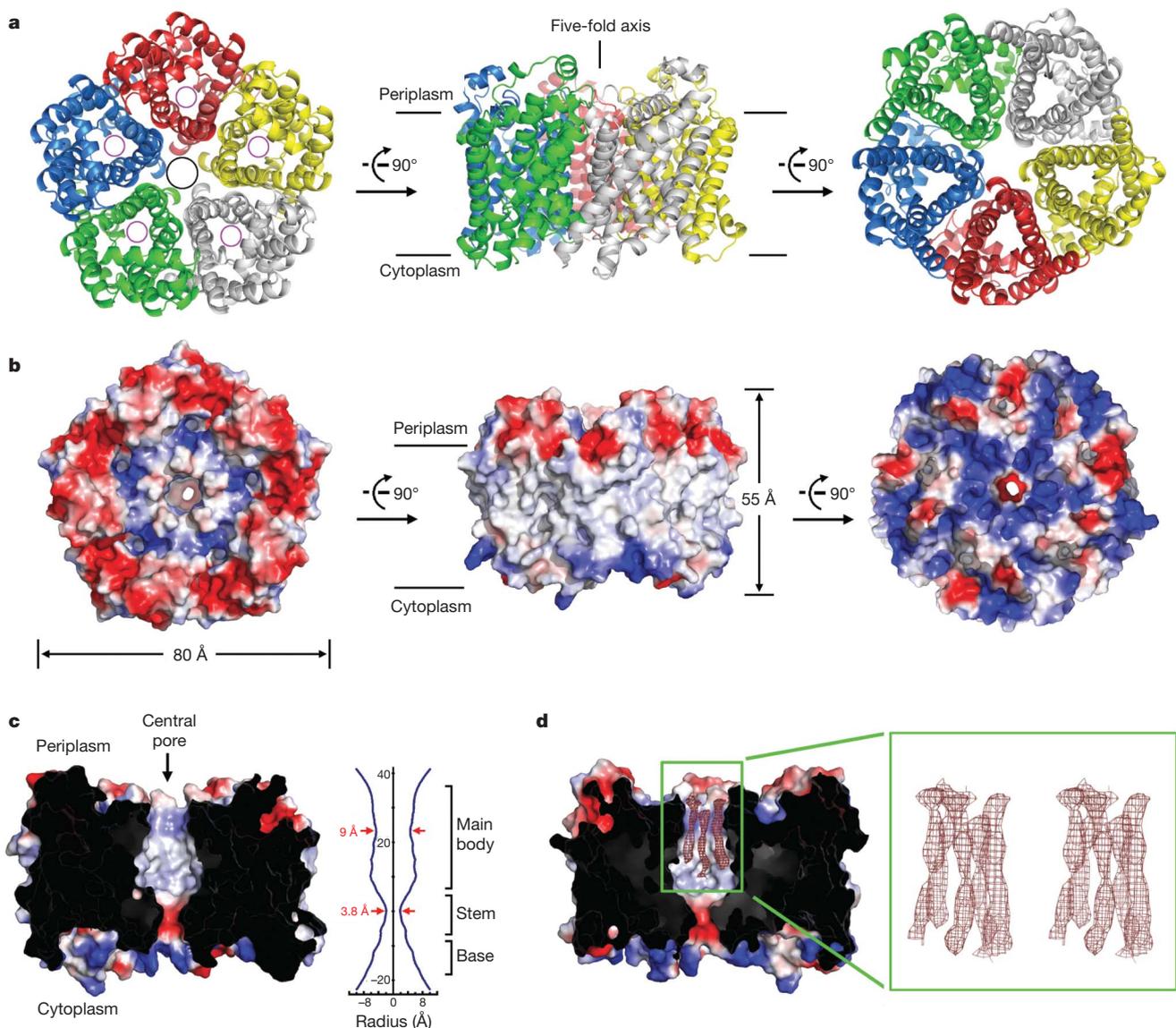
## Overall structure

Each asymmetric unit in  $P2_12_12_1$  contains five molecules of FocA, arranged as a symmetric homopentamer through a five-fold axis perpendicular to the plane of the lipid membrane (Fig. 1a). The five FocA protomers associate with each other through extensive interactions, resulting in the burial of 15,800 Å<sup>2</sup> of surface area. This analysis suggests that FocA may exist as a pentamer in lipid membrane. Supporting this notion is the observation that there are ten molecules of FocA in an asymmetric unit of the  $P3_2$  space group, organized into two homopentamers (Supplementary Fig. 3a). These homopentamers are nearly identical to each other and to that in the  $P2_12_12_1$  space group (Supplementary Fig. 3b), with a pairwise root mean squared deviation (r.m.s.d.) of about 0.7 Å. For simplicity, we limit our discussion to the FocA pentamer in the  $P2_12_12_1$  space group.

The five protomers of FocA form a short cylinder, with a diameter of about 80 Å and a height of about 55 Å (Fig. 1b). The cylindrical outer surface is hydrophobic. By contrast, the periplasmic and cytoplasmic faces of the FocA cylinder have negative and positive electrostatic potentials, respectively (Fig. 1b). The charged amino acids are located mainly at the periphery of the FocA cylinder, which seems to correlate with the lipid composition of the membrane<sup>14</sup>. Each FocA protomer contains an axial passage that is roughly perpendicular to the plane of the lipid membrane (Fig. 1a).

<sup>1</sup>Ministry of Education Protein Science Laboratory, <sup>2</sup>State Key Laboratory of Biomembrane and Membrane Biotechnology, Center for Structural Biology, School of Life Sciences and School of Medicine, Tsinghua University, Beijing 100084, China. <sup>3</sup>Beijing National Laboratory for Condensed Matter Physics, Institute of Physics, Chinese Academy of Sciences, Beijing 100190, China.

\*These authors contributed equally to this work.



**Figure 1 | The overall structure of FocA.** **a**, Ribbon representation of the FocA pentamer, shown in three perpendicular views. Note the pore in the centre of the pentameric assembly (black circle) and the axial passage in each of the five protomers (magenta circles). **b**, Surface electrostatic potential of the FocA pentamer. The three views shown here correspond to those in **a**. The periplasmic and cytoplasmic sides of FocA are negatively and positively charged, respectively. **c**, The FocA pentamer contains a central

intriguingly, the pentameric assembly of FocA also contains a central pore, whose shape resembles that of a wineglass (Fig. 1c). The main body of this pore is exclusively hydrophobic, whereas the stem is acidic and the base on the cytoplasmic side is positively charged (Fig. 1c). The narrowest point in the stem has a diameter of 3.8 Å as calculated by HOLE<sup>15</sup>. The hydrophobic main body is filled with five strings of linear electron density that is characteristic of the hydrophobic tails of detergent molecules (Fig. 1d). This structural feature suggests that the central pore is likely to be occupied by lipid molecules within the plasma membrane of *E. coli*. However, as suggested by others<sup>16</sup>, we cannot rule out the possibility that the central pore may serve as a channel of some defined function.

#### Structure of the FocA protomer

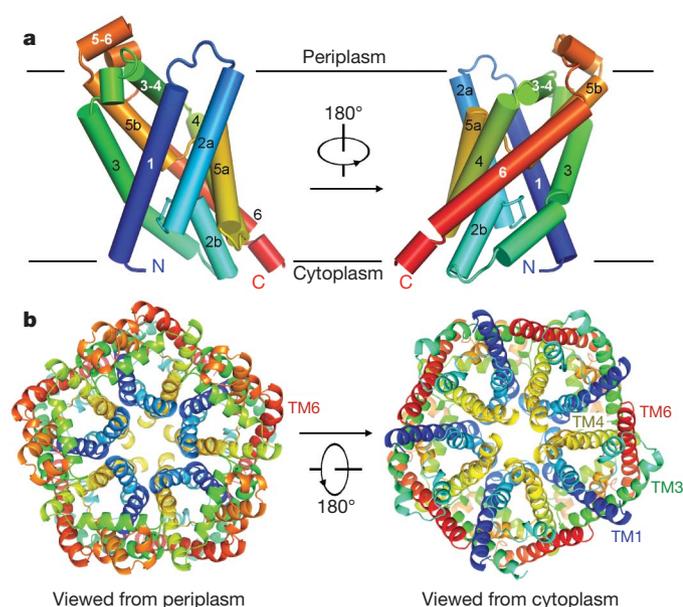
The five FocA protomers have identical structure, with a pairwise r.m.s.d. of 0.19–0.34 Å. As previously predicted<sup>7</sup>, each FocA protomer consists of six transmembrane segments (TMs) (Fig. 2a). The amino and carboxy termini of FocA are both placed on the cytoplasmic side,

pore. A cut-through section of the FocA pentamer is shown to indicate the shape and surface features of the pore. The calculated pore diameters are shown at the right. **d**, The central pore is filled with five strings of linear electron density that are characteristic of the hydrophobic tails of detergent molecules. The  $2F_o - F_c$  electron density is contoured at  $2\sigma$  and shown in stereo. All structural figures were prepared with PyMol<sup>16</sup>.

which is consistent with the predicted topology of FocA<sup>7</sup>. TM1, TM3, TM4 and TM6 each consist of a single  $\alpha$ -helix (Supplementary Fig. 4). TM2 and TM5, in contrast, consist of two  $\alpha$ -helices connected by an extended loop, which is highly conserved among the FNT family members. These two signature loops, placed roughly parallel to the plane of lipid membrane, are located in the axial passage of the FocA protomer. Two additional  $\alpha$ -helices, both on the periplasmic side, connect TM3 and TM4 ( $\alpha$ 3–4) and TM5 and TM6 ( $\alpha$ 5–6) (Supplementary Fig. 4).

Each FocA protomer contains an internal structural repeat. The N-terminal half of the protomer, TM1–TM3 (residues 29–136), are structurally related to the C-terminal half, TM4–TM6 (residues 160–276), with a quasi-two-fold axis in the plane of the lipid membrane (Supplementary Figs 4 and 5). Despite a low sequence identity of about 8%, these two halves can be superimposed with an r.m.s.d. of 3.3 Å.

In the pentameric FocA cylinder, TM3 and TM6 of each protomer constitute the bulk of the outer layer (Fig. 2b). The central elements



**Figure 2 | Structural features of the FocA protomer.** **a**, Ribbon representation of the FocA protomer, shown in two views related by a 180° rotation. FocA is rainbow-coloured, with its N terminus in blue and its C terminus in red. **b**, Role of individual transmembrane segments in the FocA protomer. The FocA pentamer is shown in two views, with the transmembrane segments colour-coded.

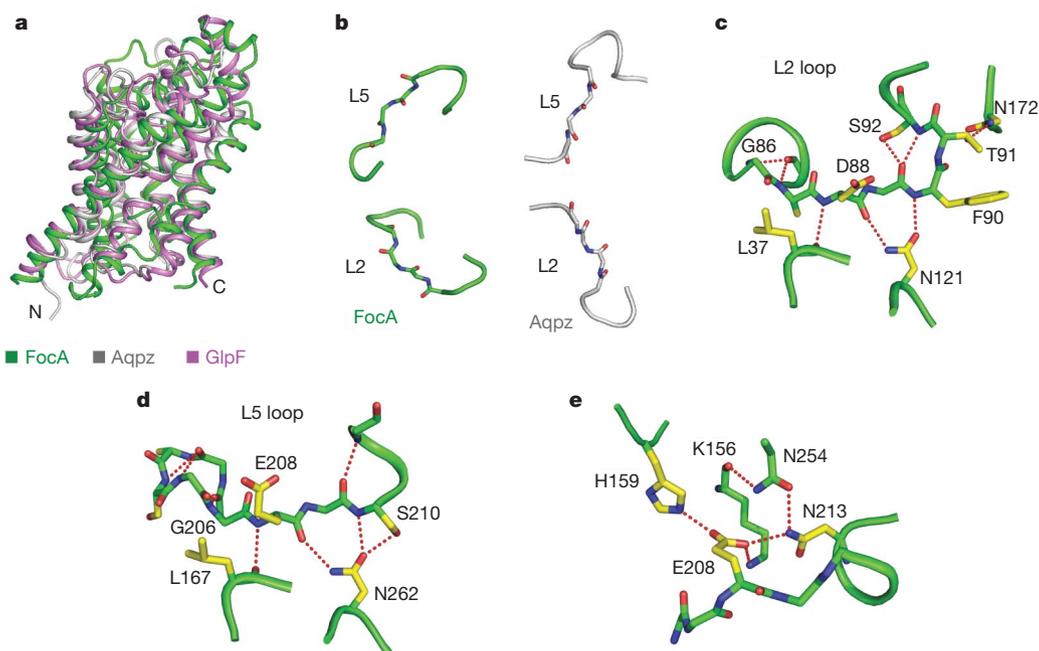
of the interface between adjacent protomers are TM1 and TM4, which traverse the full span of the lipid membrane bilayer and radiate from the centre to the outer layer of the FocA cylinder (Fig. 2b). Additional elements of this interface include TM2a and TM5a, as well as portions of TM3 and TM6 (Supplementary Fig. 6). Formation of the inter-protomer interface involves a large number of hydrophobic amino acids that are moderately conserved among the FocA homologues (Supplementary Figs 1 and 6). For example,

residues Phe 44/Ile 47/Phe 51 on TM1 and Met 174/Leu 177/Met 181 on TM4, which reside at the centre of the interface, are preserved between 17% and 67% in each of the FocA homologues. The central pore of the FocA pentamer is formed exclusively by amino acids from TM2a and TM5a of the FocA protomers (Supplementary Fig. 7).

### Structural mimicry with aquaporins

The pentameric architecture and the extensive packing interactions between adjacent FocA protomers are more indicative of a channel than a transporter. Supporting this notion, structural features of FocA are reminiscent of those of the aquaporin family of proteins (AQPs), which are a family of structurally conserved channels permeable to water or other small organic molecules such as glycerol<sup>17–20</sup>. To examine this situation systematically, we searched for structural homologues of FocA in the Protein Data Bank with DALI<sup>21</sup>. The result was unequivocal: all hits with a Z score (similarity score) of 10 or higher are AQPs. The FocA protomer can be superimposed on that of the *E. coli* water channel AqpZ (PDB code 1RC2)<sup>22</sup> and the glycerol channel GlpF (PDB code 1FX8)<sup>16</sup> with r.m.s.d.s of 3.2 and 3.3 Å, respectively (Fig. 3a and Supplementary Fig. 8). In both FocA and AQPs, the six transmembrane segments of a protomer comprise two structural repeats: TM1–TM3 and TM4–TM6, with the second transmembrane segment in each repeat (TM2 or TM5) disrupted by a highly conserved loop. FocA is an integral membrane protein with no sequence homology with AQPs but sharing the same fold.

Despite the overall structural similarity, FocA has prominent features that distinguish it from AQPs. First, in contrast to the homotetrameric AQPs, FocA is organized as a homopentamer (Supplementary Fig. 9). In comparison with AQPs, the transmembrane segments within each FocA protomer are arranged at slightly different orientations to accommodate the extra protomer in the assembly. The size of the central pore of the FocA pentamer is larger than that of the AQP tetramer. Second, unlike AQPs, the two internal structural repeats of FocA share little sequence similarity, and neither contains an Asn-Pro-Ala (NPA) motif, which is characteristic of AQPs (Supplementary Fig. 1). Third, and importantly, the two loops that disrupt TM2 and TM5, named L2



**Figure 3 | The FocA protomer is structurally similar to aquaporin.** **a**, Structural overlay of FocA (green), AqpZ<sup>22</sup> (grey) and GlpF<sup>16</sup> (magenta). **b**, The L2 and L5 loops of FocA (green) show different structural features from those in the corresponding loops of AqpZ (grey). Highlighted here are the overall conformation and main-chain carbonyl groups from these loops. **c**, A close-up view of the L2 loop in FocA. Hydrogen bonds are indicated by

dashed red lines. **d**, A close-up view of the L5 loop in FocA. Similarly to the L2 loop, the L5 loop is hydrogen-bonded by a conserved asparagine residue, Asn 262. **e**, A network of hydrogen bonds around Glu 208 in the L5 loop. Glu 208 in the L5 loop and Asp 88 in the L2 loop are both highly conserved in FocA homologues.

and L5, respectively, have different configurations from those of AQPs. For example, in AQPs but not in FocA, the backbone carbonyls in these loops are positioned along the same side of the channel and point into the channel passage (Fig. 3b). Consequently, these carbonyl oxygen atoms directly hydrogen-bond to water or glycerol molecules in the channel and presumably facilitate their movement<sup>16,22–25</sup>.

In FocA, the central portion of the L2 loop is constrained by three hydrogen bonds (Fig. 3c). The amide nitrogen and carbonyl oxygen of Asp 88 are hydrogen-bonded to the carbonyl oxygen of Leu 37 and the side-chain nitrogen of Asn 121, respectively. The side-chain oxygen atom of Asn 121 accepts an additional hydrogen bond from the amide nitrogen of Phe 90. In addition, Thr 91 at the C-terminal end of the L2 loop donates a hydrogen bond to the side-chain nitrogen atom of Asn 172. The central segment of the L5 loop is similarly confined by hydrogen bonds (Fig. 3d). In this case, the amide nitrogen and carbonyl oxygen of Glu 208 interact with the carbonyl oxygen of Leu 167 and the side-chain nitrogen of Asn 262, respectively. The side-chain oxygen atom of Asn 262 accepts two additional hydrogen bonds from the amide nitrogen and side-chain oxygen atoms of Ser 210. Furthermore, the side chain of Glu 208 organizes a network of hydrogen bonds, with its carboxylate group accepting three bonds from the side chains of Lys 156, Asn 213 and His 159 (Fig. 3e).

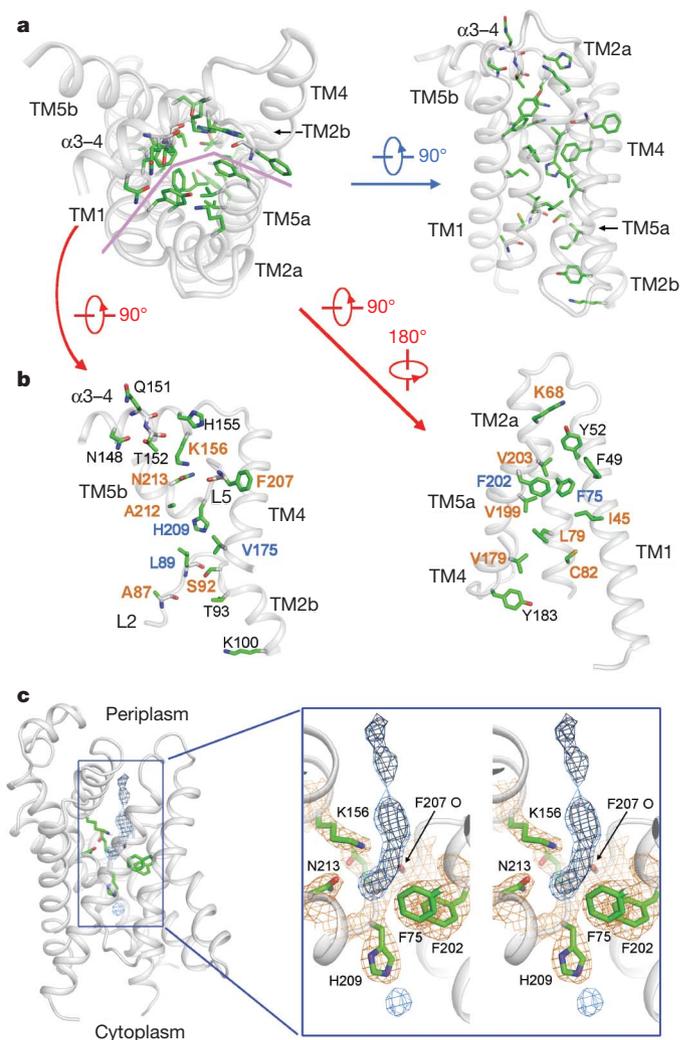
Unlike the AQPs, the L2 and L5 loops of FocA are unrelated by primary sequences, yet they share a number of notable structural features. Both loops are constrained by conserved hydrogen bonds, of which two are mediated by a similarly positioned Asn residue: Asn 121 in TM3 and Asn 262 in TM6 (Fig. 3c, d). These two Asn residues are highly conserved among members of the FNT family, suggesting functional significance. A highly conserved acidic residue is positioned in the middle of the loops: Asp 88 in L2 and Glu 208 in L5.

### Channel-lining residues

The axial channel of the FocA protomer contains a mixture of polar, charged and hydrophobic amino acids, which render the passage amphipathic (Fig. 4a). The hydrophilic side is formed by the L2 and L5 loops, the connecting helix  $\alpha$ 3–4, TM2b, TM4 and TM5b (Fig. 4b). This side comprises nine polar or charged residues and several main-chain groups. The other side, formed by TM1, TM2a, TM4 and TM5a, is considerably more hydrophobic, especially in the central portion of the channel. The amphipathic nature of the FocA channel is consistent with the chemical property of its putative substrate: formate or other short-chain acids.

Amino acids that line the channel are highly conserved in FocA homologues across several species, with five invariant residues in the central portion of the channel. The hydrophilic side of the channel contains three invariant residues, Leu 89, Val 175 and His 209, and six conserved amino acids, of which three are polar or charged (Ser 92, Lys 156 and Asn 213). The hydrophobic side comprises two invariant residues, Phe 75 and Phe 202, and seven conserved amino acids, one of which is charged (Lys 68). The conservation of channel-lining residues suggests a shared mechanism for the channel activity of these FocA homologues.

In the periplasmic side of each FocA channel there is an elongated stretch of well-defined electron density (Fig. 4c). The bottom portion of the electron density is surrounded by conserved amino acids Phe 75, Lys 156, Phe 202, His 209 and Asn 213 (Supplementary Fig. 1). Although we could have modelled water or formate molecules into the electron density, we chose not to because we could not differentiate unambiguously between formate and water at the present resolution. In addition, the shape of this electron density remains little changed with or without formate in the crystallization buffer. Nonetheless, a formate or water molecule placed into the bottom portion of the electron density would be well within hydrogen-bond distances of Lys 156, Asn 213 and the carbonyl oxygen of Phe 207.

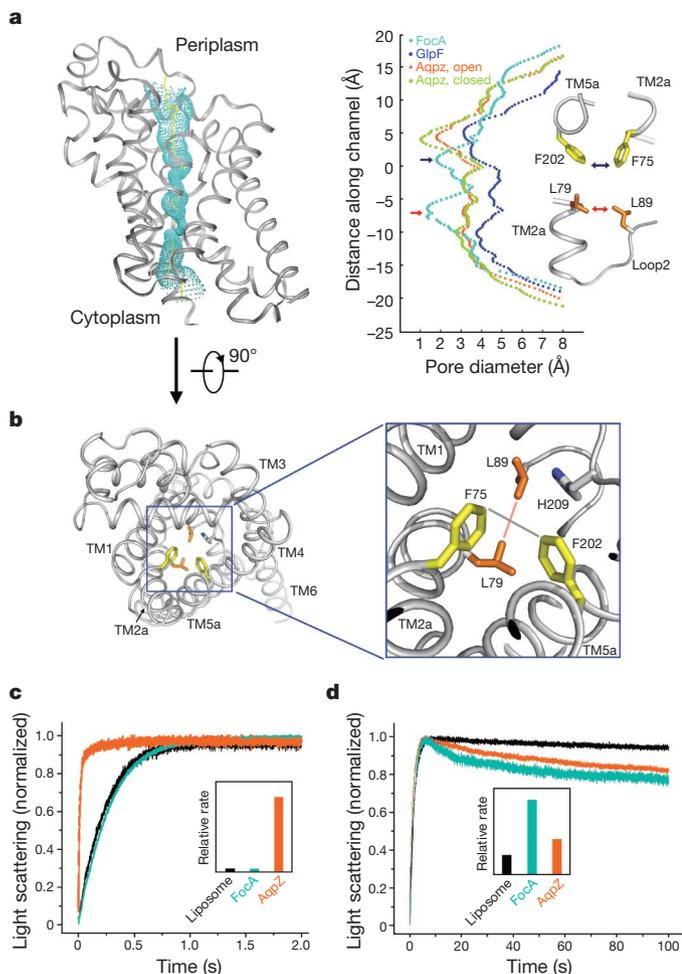


**Figure 4 | Features of the axial channel in the FocA protomer.** **a**, The FocA channel is amphipathic. Amino acids along the channel are shown in two perpendicular views. **b**, The FocA channel opened up into two halves to display the amphipathic feature. Amino acids identical in all 12 FocA homologues are shown with blue labels; those conserved in at least 6 FocA homologues are labelled in orange. **c**, The channel on the periplasmic side is occupied by an elongated stretch of electron density. The  $2F_o - F_c$  electron density, contoured at  $1.5\sigma$ , is coloured brown in FocA and black in the middle. The  $F_o - F_c$  density, contoured at  $3\sigma$ , is coloured cyan.

### Constriction sites

We calculated the channel diameter along the axial passage with HOLE<sup>15</sup>. Similarly to AQPs, the FocA channel consists of two vestibules, located on the periplasmic and cytoplasmic ends, and a narrow pore in between (Fig. 5a). A central portion of the pore, about 15 Å in length, has a diameter of 3.5 Å or less. There are two constriction sites. One site was formed mainly by the side chains of two invariant, aromatic residues: Phe 75 from TM2a and Phe 202 from TM5a. At this site, the pore diameter is narrowed to about 1.8 Å. The other constriction site is located 7.5 Å to the cytoplasmic side of the first constriction, with an even narrower diameter of 1.35 Å. This constriction was created mainly by the side chains of two highly conserved residues: Leu 79 from TM2a and Leu 89 from the L2 loop. The two pairs of constriction residues, Phe 75/Phe 202 and Leu 79/Leu 89, are positioned diagonally from each other (Fig. 5b). Neither constriction would allow the passage of water molecules, let alone formate or other solutes. Thus we conclude that the observed structure may represent that of FocA in a closed-pore state.

The observed pore diameters of the FocA protomer are generally within the range calculated for AQPs (Fig. 5a). The diameters of the



**Figure 5 | FocA contains two constriction sites and may exist in a closed-pore state.** **a**, The central channel in the FocA protomer contains two constriction sites. The channel passage (left panel), calculated by HOLE<sup>15</sup>, is indicated by cyan dots along a central yellow line. The diameters of the channel are tabulated in the right panel and compared with those from GlpF and AqpZ. **b**, The two pairs of amino acid residues that contribute to the two constriction sites, Phe 75/Phe 202 and Leu 79/Leu 89, are roughly diagonal to each other when viewed along the channel axis. **c**, FocA may be impermeable to water. Protein-free liposomes or FocA-loaded proteoliposomes were mixed with 500 mM sucrose in a stopped-flow apparatus. In response to high osmotic pressure, water molecules diffused through the lipid, causing the vesicles to deflate. The rapid changes in vesicle size are reflected by changes of light scattering. **d**, FocA is permeable to formate. The experiments conducted here are similar to those in **c** except that 20 mM sodium formate was used instead of 500 mM sucrose. FocA allowed the passage of formate, as demonstrated by the swelling of the vesicles again. AqpZ also allowed the passage of formate, though more slowly than FocA did. The relative rates of formate conduction were calculated on the basis of a published protocol<sup>16</sup>.

two constriction sites of FocA are slightly larger than that of the closed-pore AqpZ<sup>22</sup> but smaller than that of the open-pore AqpZ or GlpF<sup>16,25</sup>. Opening of the pore may require the putative gating residues, Phe 75/Phe 202 and Leu 79/Leu 89, to adopt other rotamer conformations. Gating by a hydrophobic amino acid has previously been observed for the water-selective channel AQP0 (ref. 26), in which the side chain of Met 176 regulates the opening and closure of the water pore.

To examine whether FocA is permeable to water, we generated FocA-loaded proteoliposomes and reconstituted a water permeability assay with the use of a stopped-flow apparatus (Fig. 5c). On mixing with 500 mM sucrose, water molecules diffused through the lipid, causing the lipid vesicles to deflate, as demonstrated by the increasing signal of light scattering. FocA-loaded proteoliposomes allowed a

similar rate of water passage to that of the empty liposomes, indicating that FocA may be impermeable to water. By contrast, proteoliposomes loaded with AqpZ allowed the passage of water molecules at a rate about 20-fold faster than that of the empty liposomes or FocA-loaded proteoliposomes.

Next, using a similar experimental setup, we examined whether FocA allowed the passage of formate molecules (Fig. 5d). On mixing with 20 mM formate, both empty liposomes and the FocA-loaded proteoliposomes deflated as a result of a rapid efflux of water. If FocA allows formate to pass through, the proteoliposomes will swell again over time, and the formate conductivity can be calculated by an established protocol<sup>16</sup>. As expected, the empty liposomes allowed rapid water efflux but little formate uptake. By contrast, the proteoliposomes loaded with FocA swelled again after the initial deflation phase, suggesting that FocA is permeable to formate. AqpZ also allowed the passage of formate, although more slowly than FocA.

### Perspective

How does FocA mediate the passage of formate? Although a conclusive answer remains to be found, the current structure and analysis provide tantalizing clues. To allow the passage of formate or another solute, the constriction sites of FocA must open. Movement of the aromatic side chains of Phe 75/Phe 202 or Leu 79/Leu 89 may widen the constriction site on the periplasmic or cytoplasmic side. Mutation of these bulky, hydrophobic residues to amino acids with smaller side chains is predicted to enlarge the constriction sites. Consistent with this prediction, the FocA mutants L79V/L89V and F202A both showed a markedly increased capacity for formate passage (Supplementary Fig. 10).

Substrate specificity is determined largely by the selectivity filter in GlpF<sup>16,25</sup> and the ar/R constriction site in AQPs<sup>22,23</sup>, both of which comprise hydrophobic and positively charged amino acids (Supplementary Fig. 11). Phe 43, His 174 and Arg 189 constitute the ar/R constriction site in AqpZ. Trp 48, Phe 200 and Arg 206 contribute to the selectivity filter in GlpF, where Arg 206 forms hydrogen bonds to the hydroxyl groups of glycerol and Trp 48/Phe 200 provide a hydrophobic wedge to accommodate the carbon atoms of glycerol. The selectivity filter of GlpF corresponds to the constriction site on the periplasmic side of FocA, with Trp 48/Phe 200 of GlpF replaced by Phe 75/Phe 202 of FocA. Arg 206 is replaced by Ala 212, but its functionality might be substituted for by Lys 156 or Asn 213 (Supplementary Fig. 11). This analysis suggests that these residues in FocA may constitute the selectivity filter. Another notable residue in this region is His 209, which is invariant among all FocA homologues. The imidazole side chain of His 209 is aligned with the channel passage and points to the cytoplasmic side (Fig. 4c). The side-chain rotamer conformation of His 209 is probably important in the channel activity of FocA and its homologues.

We provide strong evidence that the previously classified FNTs may constitute a previously unrecognized class of solute channels. FocA, a representative member of the FNT family, is structurally similar to AQPs, with the same membrane topology and some detailed structural features. FocA and its sequence homologues probably attained their current structure through convergent evolution. In fact, previous biochemical characterization suggested that FocA might function as a bidirectional formate channel<sup>7</sup>. Taken together, these observations suggest that FocA and its homologues may be reclassified, perhaps as a family of formate–nitrite channels (FNCs).

The FNC family members have been found in bacteria, archaea, fungi, algae and parasites, but not in higher eukaryotes<sup>8</sup>. This observation is consistent with our current understanding that, as with several other small molecules such as carbon dioxide and ammonia, formate is likely to have been a key molecule in the early evolution of life on Earth. FocA may therefore represent a member of an ancient and important family of channels that were responsible for the transfer of small molecules across cell membranes. The evolution of channel proteins that facilitate organic acid transfer across the cell membrane without an ATP requirement would be particularly beneficial to

organisms that live under fermentative conditions, in which energy conservation is at a premium. Understanding the functional mechanisms of FocA may reveal important insights into the mechanism of these channel proteins.

## METHODS SUMMARY

The recombinant FocA protein was overexpressed in *E. coli* and purified to homogeneity. Proteolysis by V8 protease removed the N-terminal 21 amino acids. The N-terminally truncated FocA was crystallized by the hanging-drop vapour-diffusion method. All data sets were collected at the Spring-8 beamline BL41XU and processed with HKL2000 (ref. 27) and the CCP4 suite<sup>28</sup>. The experimental phase in the  $P2_12_1$  space group was generated by Pt-SAD with SHELXD<sup>29</sup> and improved by DM<sup>30</sup> and DMMulti<sup>30</sup>. The model was built with BUCCANEER<sup>31</sup> and Coot<sup>32</sup>. The final model was refined with PHENIX<sup>33</sup>. The structure of FocA in the  $P3_2$  space group was solved by molecular replacement with the program PHASER<sup>34</sup> and refined with PHENIX<sup>33</sup>. The preparation of liposomes and the liposome-based assays were performed as described previously<sup>16,35</sup>.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Received 16 August; accepted 26 October 2009.**

1. Stams, A. J. & Plugge, C. M. Electron transfer in syntrophic communities of anaerobic bacteria and archaea. *Nature Rev. Microbiol.* **7**, 568–577 (2009).
2. White, W. B. & Ferry, J. G. Identification of formate dehydrogenase-specific mRNA species and nucleotide sequence of the *fdhC* gene of *Methanobacterium formicicum*. *J. Bacteriol.* **174**, 4997–5004 (1992).
3. Leonhartsberger, S., Korsal, I. & Bock, A. The molecular biology of formate metabolism in enterobacteria. *J. Mol. Microbiol. Biotechnol.* **4**, 269–276 (2002).
4. Sawers, R. G. Formate and its role in hydrogen production in *Escherichia coli*. *Biochem. Soc. Trans.* **33**, 42–46 (2005).
5. Sawers, G. The hydrogenases and formate dehydrogenases of *Escherichia coli*. *Antonie Van Leeuwenhoek* **66**, 57–88 (1994).
6. Stephenson, M. & Stickland, L. H. Hydrogenases: bacterial enzymes liberating molecular hydrogen. *Biochem. J.* **26**, 712–724 (1932).
7. Suppmann, B. & Sawers, G. Isolation and characterization of hypophosphite-resistant mutants of *Escherichia coli*: identification of the FocA protein, encoded by the *pfl* operon, as a putative formate transporter. *Mol. Microbiol.* **11**, 965–982 (1994).
8. Saier, M. H. Jr et al. Phylogenetic characterization of novel transport protein families revealed by genome analyses. *Biochim. Biophys. Acta* **1422**, 1–56 (1999).
9. Jia, W. & Cole, J. A. Nitrate and nitrite transport in *Escherichia coli*. *Biochem. Soc. Trans.* **33**, 159–161 (2005).
10. Jia, W., Tovell, N., Clegg, S., Trimmer, M. & Cole, J. A single channel for nitrate uptake, nitrite export and nitrite uptake by *Escherichia coli* NarU and a role for NirC in nitrite export and uptake. *Biochem. J.* **417**, 297–304 (2009).
11. Clegg, S., Yu, F., Griffiths, L. & Cole, J. A. The roles of the polytopic membrane proteins NarK, NarU and NirC in *Escherichia coli* K-12: two nitrate and three nitrite transporters. *Mol. Microbiol.* **44**, 143–155 (2002).
12. Clegg, S. J., Jia, W. & Cole, J. A. Role of the *Escherichia coli* nitrate transport protein, NarU, in survival during severe nutrient starvation and slow growth. *Microbiology* **152**, 2091–2100 (2006).
13. Nolling, J. & Reeve, J. N. Growth- and substrate-dependent transcription of the formate dehydrogenase (*fdhCAB*) operon in *Methanobacterium thermoformicicum* Z-245. *J. Bacteriol.* **179**, 899–908 (1997).
14. von Heijne, G. & Gavel, Y. Topogenic signals in integral membrane proteins. *Eur. J. Biochem.* **174**, 671–678 (1988).
15. Smart, O. S., Goodfellow, J. M. & Wallace, B. A. The pore dimensions of gramicidin A. *Biophys. J.* **65**, 2455–2460 (1993).
16. Fu, D. et al. Structure of a glycerol-conducting channel and the basis for its selectivity. *Science* **290**, 481–486 (2000).
17. Agre, P. The aquaporin water channels. *Proc. Am. Thorac. Soc.* **3**, 5–13 (2006).
18. Gonen, T. & Walz, T. The structure of aquaporins. *Q. Rev. Biophys.* **39**, 361–396 (2006).
19. Stroud, R. M., Nollert, P. & Miercke, L. The glycerol facilitator GlpF its aquaporin family of channels, and their selectivity. *Adv. Protein Chem.* **63**, 291–316 (2003).
20. Carbrey, J. M. & Agre, P. Discovery of the aquaporins and development of the field. *Handb. Exp. Pharmacol.* **190**, 3–28 (2009).
21. Holm, L. & Sander, C. Protein structure comparison by alignment of distance matrices. *J. Mol. Biol.* **233**, 123–138 (1993).
22. Savage, D. F., Egea, P. F., Robles-Colmenares, Y., O'Connell, J. D. III & Stroud, R. M. Architecture and selectivity in aquaporins: 2.5 Å X-ray structure of aquaporin Z. *PLoS Biol.* **1**, E72 (2003).
23. Sui, H., Han, B. G., Lee, J. K., Walian, P. & Jap, B. K. Structural basis of water-specific transport through the AQP1 water channel. *Nature* **414**, 872–878 (2001).
24. Murata, K. et al. Structural determinants of water permeation through aquaporin-1. *Nature* **407**, 599–605 (2000).
25. Tajkhorshid, E. et al. Control of the selectivity of the aquaporin water channel family by global orientational tuning. *Science* **296**, 525–530 (2002).
26. Gonen, T. et al. Lipid-protein interactions in double-layered two-dimensional AQP0 crystals. *Nature* **438**, 633–638 (2005).
27. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
28. Collaborative Computational Project Number 4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).
29. Schneider, T. R. & Sheldrick, G. M. Substructure solution with SHELXD. *Acta Crystallogr. D* **58**, 1772–1779 (2002).
30. Cowtan, K. dm: an automated procedure for phase improvement by density modification. *Joint CCP4 and ESF-EACBM News. Protein Crystallogr.* **31**, 34–38 (1994).
31. Cowtan, K. The Buccaneer software for automated model building. *Acta Crystallogr. D* **62**, 1002–1011 (2006).
32. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004).
33. Adams, P. D. et al. PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr. D* **58**, 1948–1954 (2002).
34. McCoy, A. J. et al. Phaser crystallographic software. *J. Appl. Cryst.* **40**, 658–674 (2007).
35. Borgnia, M. J., Kozono, D., Calamita, G., Maloney, P. C. & Agre, P. Functional reconstitution and characterization of AqpZ, the *E. coli* water channel protein. *J. Mol. Biol.* **291**, 1169–1179 (1999).
36. DeLano, W. L. PyMOL Molecular Viewer. <http://www.pymol.org> (2002).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank N. Shimizu, S. Baba and T. Kumasaka at the Spring-8 beamline BL41XU for assistance, and J. He and S. Huang for help at Shanghai Synchrotron Radiation Facility (SSRF). This work was supported by funds from the Ministry of Science and Technology of China (grants 2009CB918801 and 2009CB918802), Tsinghua University 985 Phase II funds, Project 30888001 supported by National Natural Science Foundation of China, and the Beijing Municipal Commissions of Education and Science and Technology. N.Y. acknowledges support from the Yuyuan Foundation and Li's Foundation.

**Author Contributions** Experiments were performed by Y.W., Y.H., J.W., C.C., W.H., P.L., Y.-N.X., P.W. and N.Y. Data were analysed by Y.W., Y.H., J.W., N.Y. and Y.S. The manuscript was prepared by N.Y. and Y.S.

**Author Information** The atomic coordinates of FocA in the  $P2_12_1$  and  $P3_2$  space groups have been deposited in the Protein Data Bank under accession codes 3KCU and 3KCV, respectively. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to N.Y. ([nyan@tsinghua.edu.cn](mailto:nyan@tsinghua.edu.cn)) or Y.S. ([shi-lab@tsinghua.edu.cn](mailto:shi-lab@tsinghua.edu.cn)).

## METHODS

**Protein preparation.** The cDNA of full-length FocA from *E. coli* strain O157:H7 was subcloned into pET21b (Novagen). Overexpression of FocA was induced in *E. coli* BL21(DE3) by 0.2 mM isopropyl  $\beta$ -D-thiogalactoside (IPTG) when the cell density reached an attenuation ( $D_{600}$ ) of 1.5. After growth for 16 h at 22 °C, the cells were harvested, homogenized in buffer containing 25 mM Tris-HCl pH 8.0 and 150 mM NaCl, and lysed by sonication. Cell debris was removed by low-speed centrifugation for 10 min. The supernatant was collected and applied to ultracentrifugation at 150,000g for 1 h. Membrane fraction was harvested and incubated with 1.5% (w/v) *n*-octyl- $\beta$ -D-glucopyranoside ( $\beta$ -OG; Anatrace) for 3 h at 4 °C. After another ultracentrifugation step at 150,000g for 30 min, the supernatant was collected and loaded on Ni<sup>2+</sup>-nitrilotriacetate affinity resin (Qiagen) and washed with 25 mM Tris-HCl pH 8.0, 150 mM NaCl, 10 mM imidazole, 1.06%  $\beta$ -OG. The protein was eluted from the affinity resin by 25 mM Tris-HCl pH 8.0, 300 mM imidazole, 1.06%  $\beta$ -OG, and concentrated to about 10 mg ml<sup>-1</sup> before further purification by gel filtration (Superdex-200; GE Healthcare). The buffer for gel filtration contained 25 mM Tris-HCl pH 8.0, 150 mM NaCl and detergents. The peak fraction was collected and concentrated to about 6 mg ml<sup>-1</sup> for crystallization.

The FocA mutants were generated with a standard PCR-based strategy and were subcloned, overexpressed and purified in the same way as the wild-type protein. Limited proteolysis was used to identify the structural core of FocA. Mass spectrometry revealed that the N-terminal 21 amino-acid residues were removed by digestion with V8. A new construct (residues 22–285) was made to express the N-terminally truncated protein. Truncation of the N-terminal 21 residues had no apparent effect on the channel activity of FocA. AqpZ was cloned from *E. coli* genome DNA to pET21b, overexpressed, and purified as described for FocA.

**Crystallization.** Crystals were grown at 18 °C by the hanging-drop vapour-diffusion method. Full-length FocA protein purified in 0.4% (w/v) decyl- $\beta$ -D-maltopyranoside (DM; Anatrace) gave rise to large diamond-shaped crystals in multiple poly(ethylene glycol) (PEG) conditions. However, the best data set collected at beamline BL41XU of Spring-8 for these crystals, at a nominal resolution of 4 Å, was unsuitable for structural determination. The truncated FocA purified in 0.8% (w/v)  $\beta$ -OG and 0.046% (w/v) *n*-dodecyl-*N,N*-dimethylamine-*N*-oxide (LDAO; Anatrace) gave rise to crystals of trigonal plates. The crystals appeared overnight in the well buffer containing 0.1 M MOPS pH 7.5, 36% (w/v) PEG400, 200 mM NaCl or sodium formate, and grew to full size in one week. The crystals from space group  $P3_2$  diffracted to 3.2 Å at BL41XU. To further improve the resolution, a third detergent was screened as an additive. Finally, 0.2% Cymal-2 was shown to be essential for improving the diffracting resolution from 3.5 Å to 2.2 Å. The crystals belong to space group  $P2_12_12_1$ , with unit cell dimensions of  $a = 102.31$  Å,  $b = 107.07$  Å,  $c = 164.24$  Å, and  $\alpha = \beta = \gamma = 90^\circ$ . Derivative crystals were obtained by soaking crystals for 24 h in mother liquor containing 2 mM K<sub>2</sub>PtCl<sub>4</sub> followed by back-soaking for 3 min in well buffer plus 1.0%  $\beta$ -OG. Both native and heavy-atom-derived crystals were directly flash-frozen in a cold nitrogen stream at 100 K.

**Data collection and structure determination.** All data sets were collected at the Spring-8 beamline BL41XU, and processed with HKL2000 (ref. 27). Additional processing was performed with programs from the CCP4 suite<sup>28</sup>. Data collection statistics are summarized in Supplementary Table 1. The platinum sites were located with SHELXD<sup>29</sup> from the Bijvoet differences in the Pt-SAD data. The identified positions were refined and the phases were calculated with SAD

experimental phasing module of PHASER<sup>34</sup>. The real-space constraints, including solvent flattening, histogram matching and non-crystallographic symmetry averaging, were applied to the electron density map in DM<sup>30</sup>. Cross-crystal averaging in DMMulti<sup>30</sup> gave rise to electron density maps of sufficient quality for model building, using the Pt-SAD and  $P2_12_12_1$  native data. The initial model was built with BUCCANEER<sup>31</sup>. Additional missing residues in the automatically built model were added manually in *Coot*<sup>32</sup>. The final model in the  $P2_12_12_1$  space group was refined with PHENIX<sup>33</sup>. Of the amino acids in the final atomic model, 94.1%, 5.8% and 0.1% are in the most favourable, additional allowed, and generously allowed regions of the Ramachandran plots, respectively. No amino acid is in the disallowed region. The refined model for one FocA protomer was used for molecular replacement with the program PHASER<sup>34</sup> into the hexagonal crystal form, and ten protomers per asymmetric unit were found. The  $P3_2$  structure was also refined with PHENIX<sup>33</sup>. Of the amino acids in the final atomic model, 87.5%, 11.5%, 0.8% and 0.2% in the  $P3_2$  space group are in the most favourable, additional allowed, generously allowed, and disallowed regions of the Ramachandran plots, respectively.

**Preparation of liposome and proteoliposome.** *E. coli* polar lipids (Avanti Polar Lipids) were dissolved in chloroform/methanol mixture (3:1, v/v) at 50 mg ml<sup>-1</sup> and dried under a nitrogen stream. The lipids were then dissolved at 20 mg ml<sup>-1</sup> in buffer containing 20 mM HEPES pH 7.0 and 0.4 mM dithiothreitol, and incubated at 22 °C for 1 h followed by sonication for 2 h.

The liposomes and proteoliposomes used for the water permeability assay were reconstituted as described in ref. 35; those used for the formate permeability assay were prepared as described in ref. 16.

**Assay for water permeability.** To measure the water permeability of different target proteins, the osmotic responses of proteoliposomes were monitored as described previously<sup>16,35,37,38</sup>. In brief, 75  $\mu$ l of liposome or proteoliposome solution was rapidly mixed with an equal volume of 500 mM sucrose in the same buffer (20 mM HEPES pH 7.0). The osmotic pressure then causes water efflux from the lipid vesicle, leading to a decrease in the vesicle volume and an increase in the light scattering signal. The experiment was performed on a stopped-flow device (Applied Photophysics PiStar180) at 5 °C. The light scattering signal was recorded at an emission wavelength of 440 nm. The data were processed with software Origin to fit the equation  $Y = ae^{-kt} + b$ , where  $t$  is time and  $Y$  is the signal of light scattering. The relative rate of water conduction ( $k$  value) was calculated to compare the water permeability of AqpZ and FocA.

**Assay for formate permeability.** To measure the formate permeability, the liposomes or proteoliposomes were rapidly mixed with 20 mM formate at equal volume. The change in vesicle size was detected by recording the light scattering signal at 440 nm. The above experiments were performed on a stopped-flow apparatus (Applied Photophysics PiStar180) at 10 °C. On the basis of a published protocol<sup>16</sup>, changes in light scattering could be fitted by two exponentials in the equation  $Y = [a(1 - e^{-kt}) - b]e^{-\mu t} + c$ , where  $t$  is time and  $Y$  is the signal of light scattering. The first time-constant ( $k$ ) corresponds to the rapid water efflux. The second time-constant ( $\mu$ ) corresponds to the relative rates of swelling again due to formate conduction<sup>16</sup>. The data were analysed with Origin.

37. Borgnia, M. J. & Agre, P. Reconstitution and functional comparison of purified GlpF and AqpZ, the glycerol and water channels from *Escherichia coli*. *Proc. Natl Acad. Sci. USA* **98**, 2888–2893 (2001).

38. Khademi, S. *et al.* Mechanism of ammonia transport by Amt/MEP/Rh: structure of AmtB at 1.35 Å. *Science* **305**, 1587–1594 (2004).