Supplementary information

Oligomer formation

Probing the protein surface with a 1.4 Å probe reveals the octameric Wza buries a total of approximately 72000 $Å^2$ accessible surface, consistent with its stability in SDS. The Wza oligomer is formed by a network of H-bonds, salt bridges and hydrophobic contacts. The subunit association between two Wza monomers is as follows: the N-terminal loop (residues 22-35) is wrapped around ring 3 via two salt bridges from Arg 33 and Asp 44, and Asp 47 and Lys 34. Further stabilization is provided by a H-bond (Tyr 341 and Gln 27) and a van der Waal's contact (Arg 349 and Gln 27).

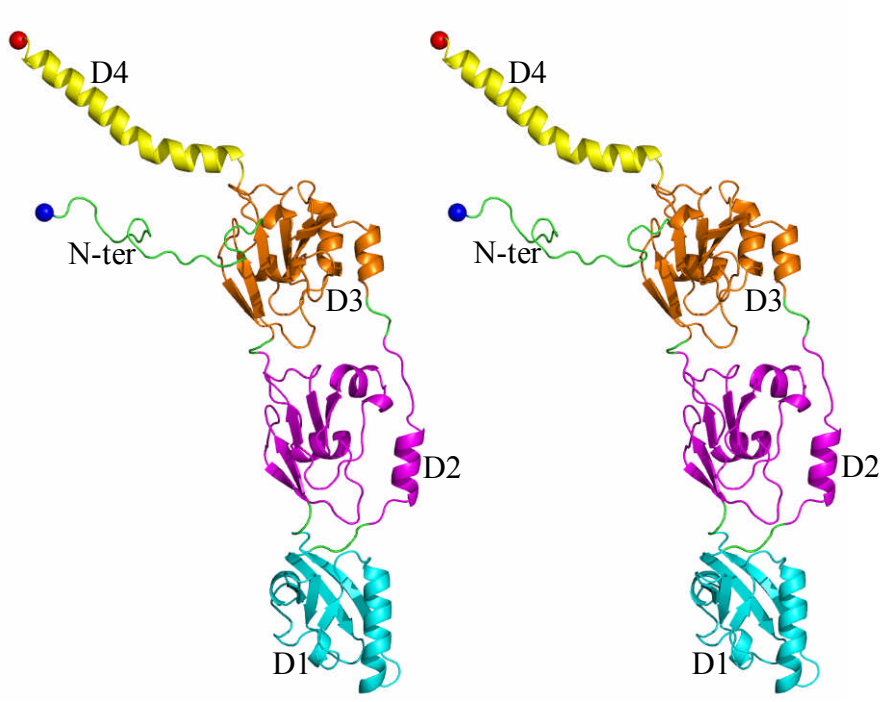
The first ring is stabilized by inter-subunit hydrogen bonds between Asp 99 and Tyr 130, Thr 116 and Arg 111, Arg 111 and Arg 111 on an adjacent monomer. Van der Waals interactions between Ile 190 and Lys 172, MSE 230 and Thr 176, MSE 230 and Phe 248, and two hydrogen bonds (Asp196 and Tyr174, Asn 199 and Thr 176) stabilize the second ring. The third ring is stabilized by salt-bridges (Asp 272 and Lys 256, Glu 280 and Lys 256), hydrogen bonds (Ile 288 and Gln 266, Arg 273 and Asp 253), and a van der Waal's contact between Ala 323 and Tyr 341. There is a potential hydrogen bond between His 365 and Glu 369 in adjacent transmembrane α-helices.

Table S1

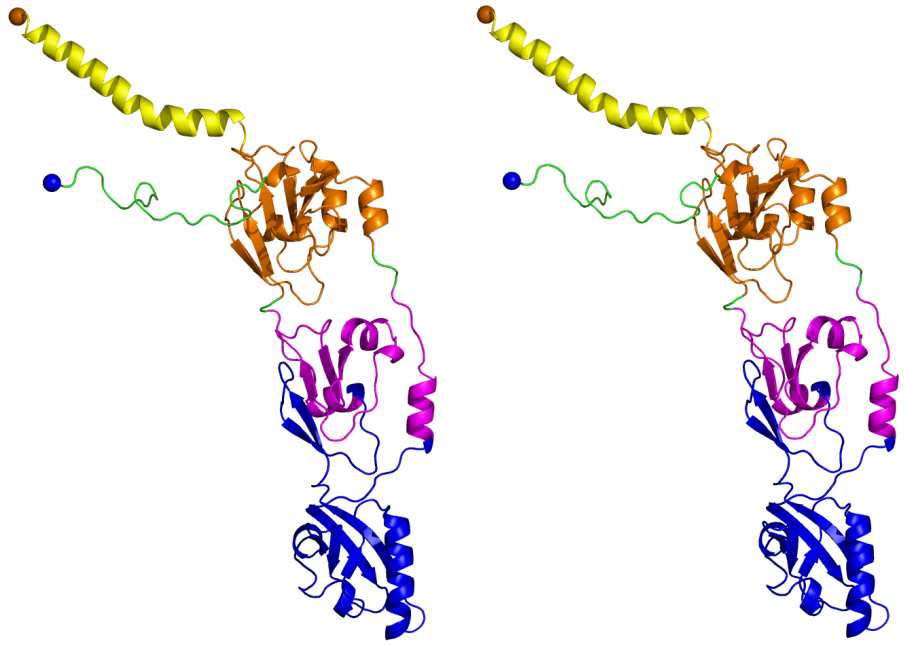
Data collection	**SAD (Se Met)**
Wavelength (Å)	(Peak) 0.9791
Resolution (Highest Shell, Å)	107.83-2.26 (2.32-2.26)
Spacegroup	$P2_12_12_1$
Cell constants (Å) α=β=γ=90°	a=95.0 b=215.6 c=221.4
Unique reflections	200359 (14046)
Average redundancy	5.0 (2.6)
Mean I/σ	18.4 (1.8)
Completeness (%)	99.2 (94.9)
Anom complete (%)[a]	95.4 (71.5)
R_{merge}[b]	0.081(0.350)
Refinement	
R	0.189 (0.243)
R_{free}	0.227 (0.313)
rmsd bonds (Å) / angles (º)	0.020/1.68
B-factor deviation bonds / angles (Å^2): main chain side chains	1.5/1.6 2.9/3.9
Residues in Ramachandran Core (%)	90.5
Protein atoms	22080
Water atoms	1264
Sulfate ions	16
Lipid atoms	112
Average B-factor (Å^2)	38.0
PDB accession code	

[a]Anomalous completeness corresponds to the fraction of possible acentric reflections for which an anomalous difference has been measured. [b] $R_{merge} = \Sigma_{hkl} \Sigma_l |I_i - \langle I \rangle| / \Sigma_{hkl} \Sigma_l \langle I \rangle$, where I_i is an intensity for the *i*th measurement of a reflection with indices *hkl* and $\langle I \rangle$ is the weighted mean of the reflection intensity.

a



b



c

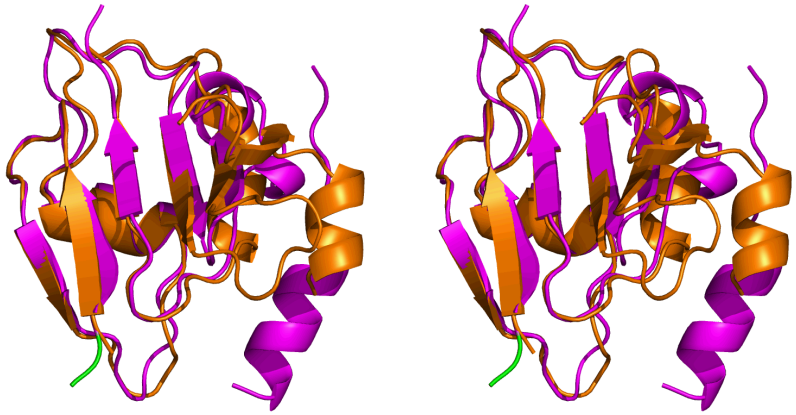
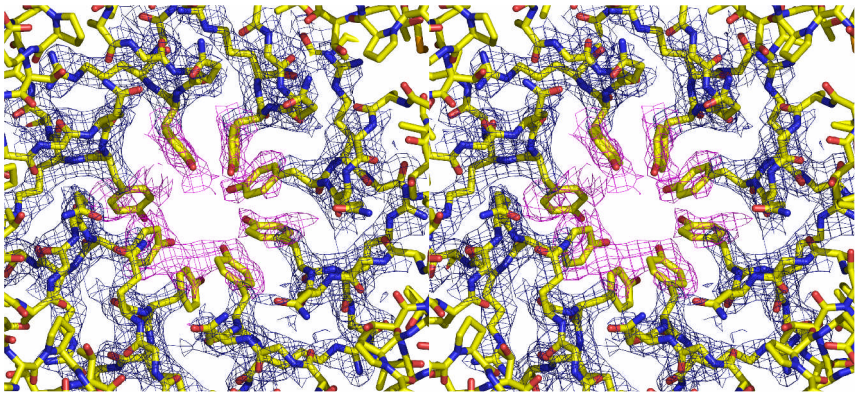


Figure S1 Monomer of Wza. **a,** The four domains are labeled and identified by color. Domain 1 (residues 89-169) is shown in cyan, domain 2 (residues 68-84 and 172-252) magenta, domain 3 (residues 46-64 and 255-344) orange and domain 4 (residues 345-379) yellow. The long loop at the N-terminus is colored green, as are loops which connect domains. The N-terminal residue (Cys 21) is shown as a blue ball and the C-terminal residue (Arg 376) as a red ball. **b,** The so called PES (polysaccharide export sequence) motif (colored as blue), Pfam02563, is found between residues 82-195 encompasses domain 1. However, the sequence motif also includes part of domain 2, in light of the structure we suggest the PES domain is redefined to be residues 89-169 (domain 1). **c,** Domain 3 (orange) has been superimposed upon domain 2 (magenta). This shows clearly the domain duplication which has occurred. The root mean square deviation of 69 Cα atoms is 1.2 Å. Domain 3 is slight larger due to a longer loop and an extra β strand.

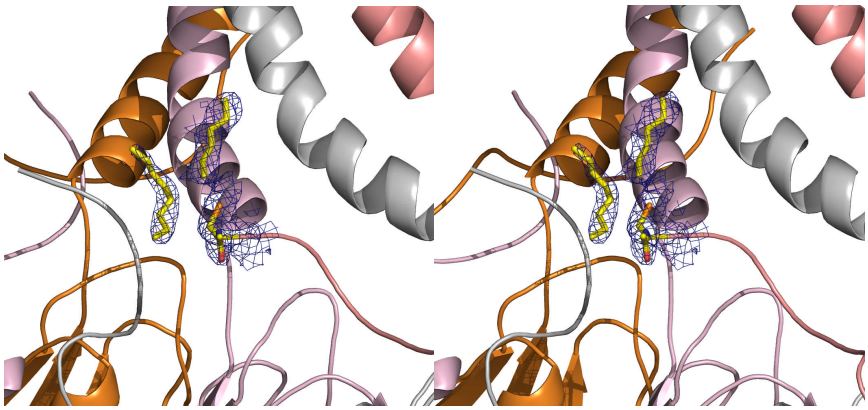
```
     domain 3 >|  |<- domain 2 ->| |<----- domain 1 -----
Wza    -PGLIDQLRPEPVIARSNPQLDNLLKS-YEYRIGVGDVLMVTVWDHPELTTPAGQYRSAS
BexD   -ESLVQQIYAAQQSQRFSGFADVRGNGGYAGAVNVGDVLEISIWEAPPAVLFGTTFSSEG
CtrA   -HTVAQLLYKAQINQSFTQFGDGYAS---AGTLNIGDVLDIMIWEAPPAVLFGGGLSSMG
KpsD   PPVVMSRMFGAQLFNGTSADSGATVGFNPDYILNPGDSIQVRLWG---AFTFDGAL----
                                      |<++++++++++++++++++++++++++

       ------------------------------------------------------------
Wza    DTG-----N---WVNSDGTIFYPYIGKVQVAGKTVSQVRQDITSRLTTYIES-PQVDVSI
BexD   QGSGHVTQLPSQIVNKNGTVTVPFVGNISVAGKTPEAIQAQIVASLSRKANQ-PQAVVKI
CtrA   SGSAHQTKLPEQLVTARGTVSVPFVGDISVVGKTPGQVQEIIKGRLKKMANQ-PQVMVRL
KpsD   Q------------VDPKGNIFLPNVGPVKIAGVSNSQLNALVTSKVKEVYQSNVNVYASL
       ++++++++++++++++++++++++ PES motif ++++++++++++++++++++++++

       ->|      |<------------------------- domain 2 -----
Wza    AAFRSQKVYVTGEVANSGKQAITNIPLTVMDAINAAGGLAADADWRNVVLTHNGKDTK-I
BexD   ANNNSSDVTVIRQ-GSAVRMPLTANDERVLDAVAAIGGSTGNIEDVTVQLTRGNQVKT-L
CtrA   VQNNAANVSVIRA-GNSVRMPLTAAGERVLDAVAAVGGSTANVQDTNVQLTRGNVVRT-V
KpsD   LQAQPVKVYVTGFVRNPGLYGGVTSDSLLNYLIKAGGVDPERGSYVDIVVKRGNRVRSNV
       ++++++++++++++++++++++++++>

       -------------------------->| |<----- domain 3 -------
Wza    SLYALMQKGDLTQNHLLYHGDILFIPSNDDLKVFVMGEVGKQSTLKMDRSGMTLAEALG-
BexD   AFETLIADP--KQNIVLRAGDVVSLLNTP-YKFTGLGAVGNNQQLRFSSSGITLAEAIG-
CtrA   ALEDLVANP--RQNILLRRGDVVTMITNP-YTFTSMGAVGRTQEIGFSARGLSLSEAIG-
KpsD   NLYDFLLNG-KLGLSQFADGDTIIVGPRQ-HTFSVQGDVFNSYDFEFRESSIPVTEALSW
```

Figure S2 A sequence alignment of outer membrane auxiliary (OMA) proteins with Wza (residues 57 – 283). Conserved residues are shown in red, similar residues in blue. BexD is a capsule export protein from *H. influenzae*[18], CtrA is a capsule export protein from *N. meningitidis*[17] and KpsD is an *E. coli* group 2 capsule assembly protein[16]. The domains found in the structure of Wza and the conserved Pfam02563 polysaccharide biosynthesis/export motif (PES motif) is marked with + below the residues. We suggest the PES signature sequence should be modified to include only domain 1 of the structure.

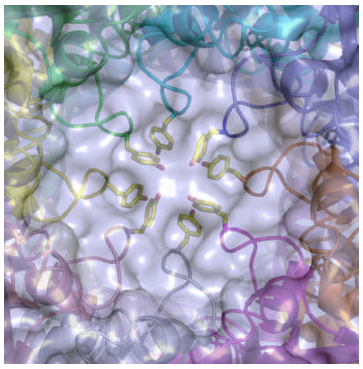
a



b



c



d

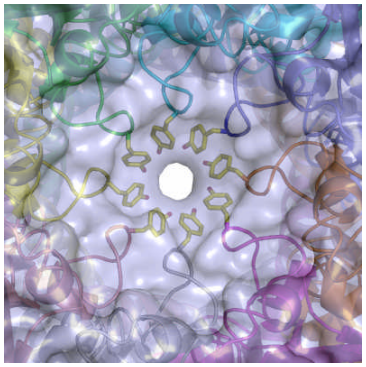
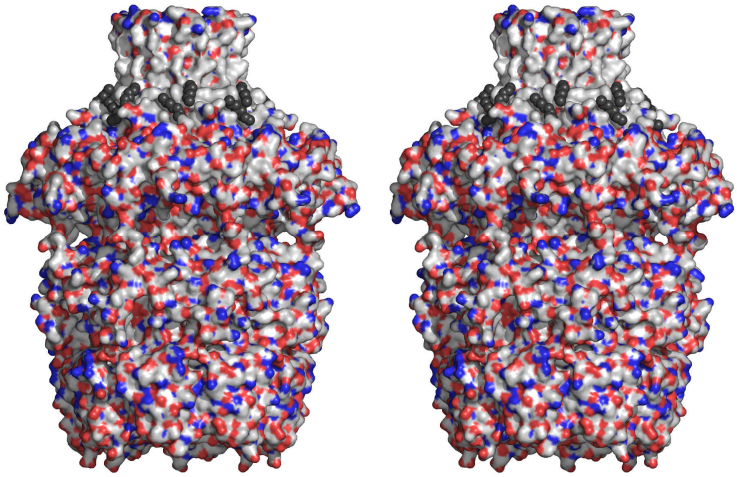


Figure S3 Tyr 110 is at the entrance to the cavity.
a, The side chain of Tyr 110 was set to low occupancy (0.1) throughout the refinement in order to exclude it from phase calculation. Shown in blue chicken wire is the 2Fo-Fc map contoured at 1.2 σ and magenta chicken wire is the Fo-Fc map at 2.6 σ. Six tyrosines are modeled in the second most common conformation of this residue, two in the most common. The side chain is clearly disordered but there is some weak density in each of the two possible conformations. **b,** Density can be seen for part of the diacylglycerol group which connects to the S atom of Cys 21. However, after refinement the 2Fo-Fc density is not continuous, so we conservatively modeled it as a simple aliphatic molecule. Whether the other chain is from detergent (also modeled a simple aliphatic) or from the diacyglycerol is unclear. **c,** A surface and cartoon representation of Wza, looking from the periplasm into the central cavity. Tyr 110 adopts one set of conformations which completely seal the cavity. **d**, The same view as in Fig. S3b but with an alternate conformation of Tyr 110 that creates a hole of diameter of 8 Å. Tyr 110 is disordered in the structure, but there is weak density for both conformations (Fig. S3a).

Figure S4

a



b

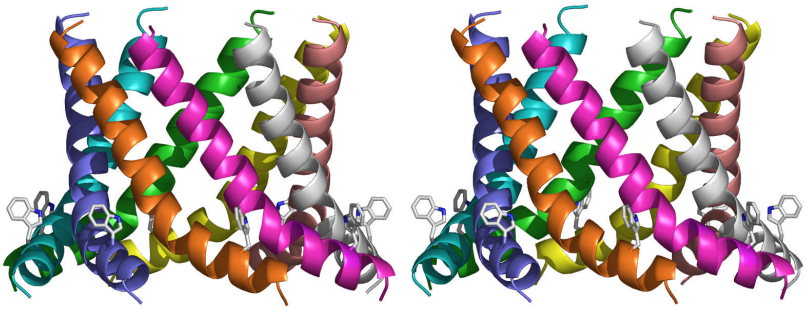
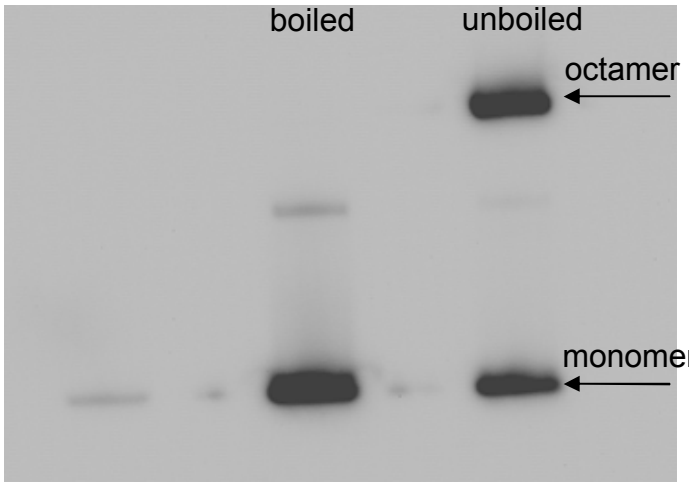


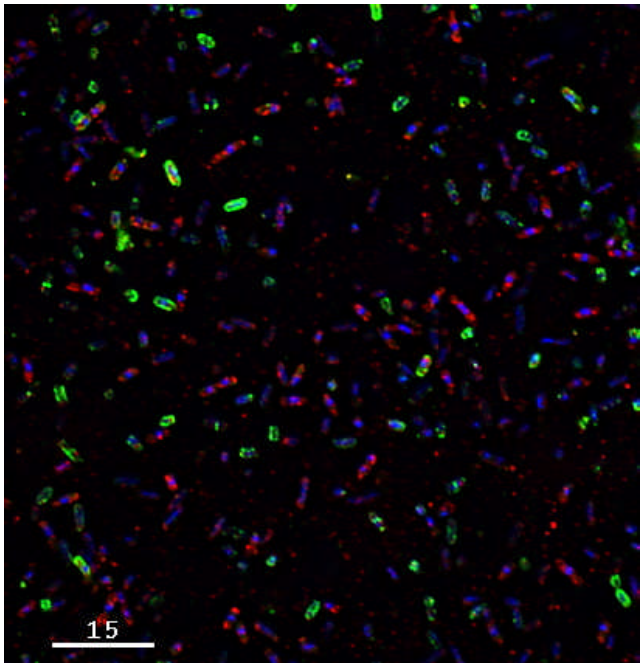
Figure S4 The helical barrel. a. Wza is shown as a surface with polar atoms colored (colors are Fig. 2c.). Wza is oriented as the Fig. 2b (first view). The lipid molecules are shown as black spheres and are located near the top of the structure. The helical barrel is clearly non-polar. **b,** There is a band of tryptophan residues (W350) exposed on the surface of helical barrel. Such bands of tryptophans are a common feature transmembrane helices and are thought to act as "lock" keeping the protein embedded in the membrane.

Figure S5.

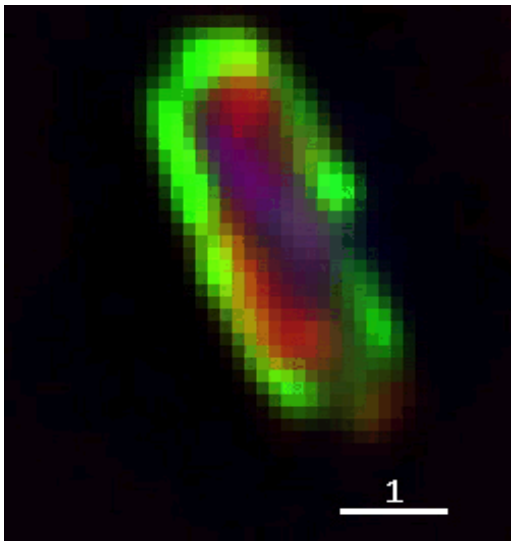
a



b



c



d

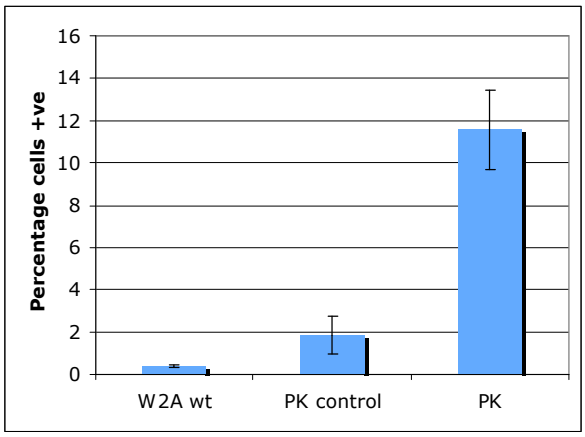
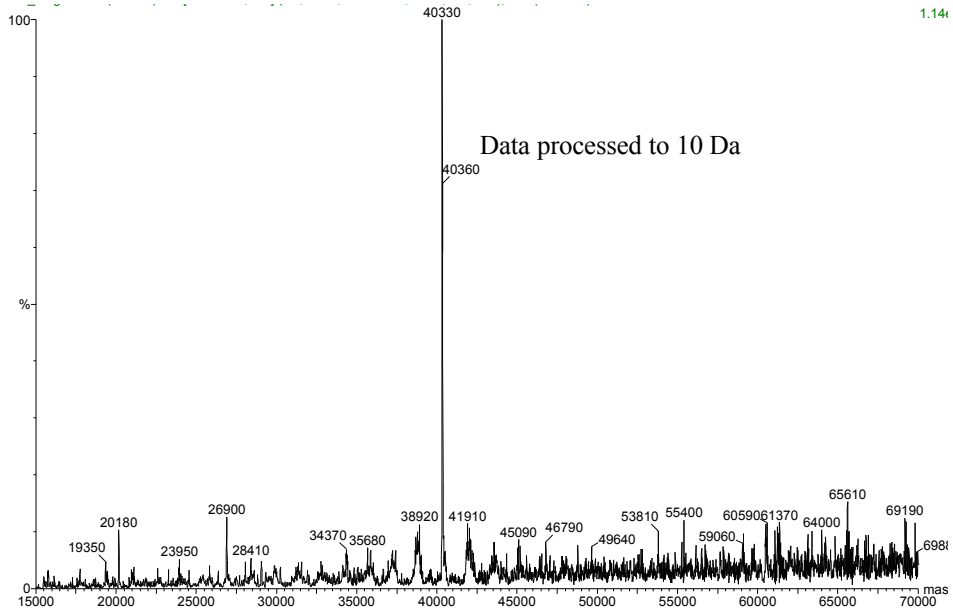


Figure S5 The orientation of Wza in the outer membrane.
a, Western blot of purified Wza-PK shows the octamer in the unboiled samples and monomer in the boiled samples. **b**. *E. coli* expressing Wza modified by addition of PK tag to the C-terminus (Wza-PK) binds the green fluorescent anti-PK antibody. Under our expression conditions we use only about 10% of cells seem to bind the antibody strongly. The red anti-alkaline phosphatase antibody locates the periplasm and the DNA is stained blue by DAPI. The scale bar is 15µm. **c,** A single cell is shown, the scale bar is 1µm. **d,** Flow cytometry analysis of *E. coli* expressing the modified Wza-PK construct (column "PK") after induction with arabinose. We used the same greenfluorescent anti-PK antibody used for direct visualization. Our controls of uninduced *E. coli* cells with the Wza-PK construct (column "PK control") and *E. coli* with the native Wza (column "Wza wt") induced by addition of arabinose are shown.

a



b

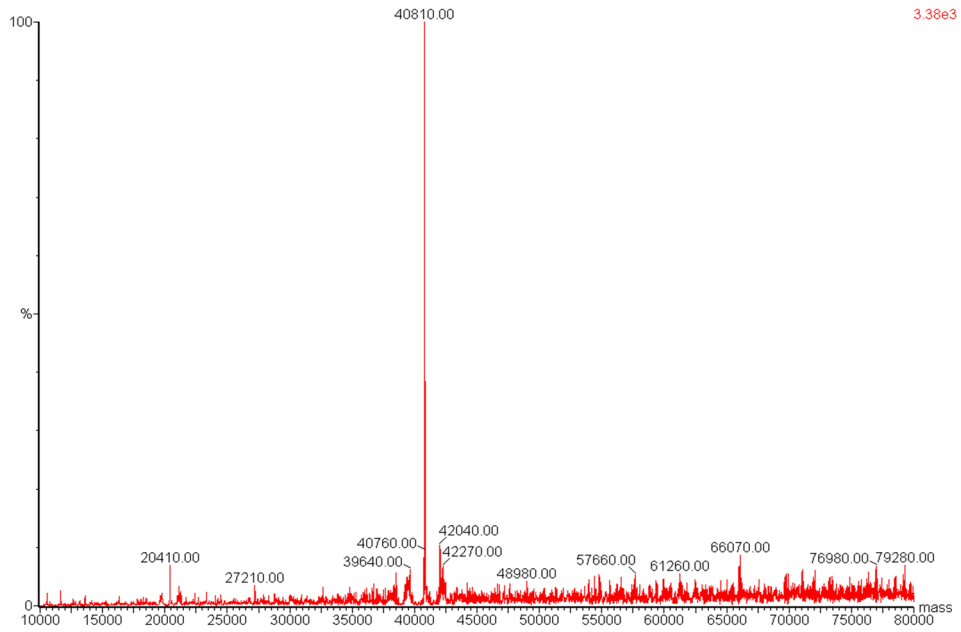


Figure S6 Protein mass spectrometry

a, The peak at 40330 ± 10 Da corresponds to the predicted weight of 40340 Da. This is derived from mature protein (39550) plus the three predicted palmitoyl groups (717) and a glycerol moiety (73), the classical lipoprotein signature. **b** The selenomethionine protein peak at 40810 is increased from native by 470, which corresponds to 10 selenium atoms per monomer (^{79}Se - ^{29}S).