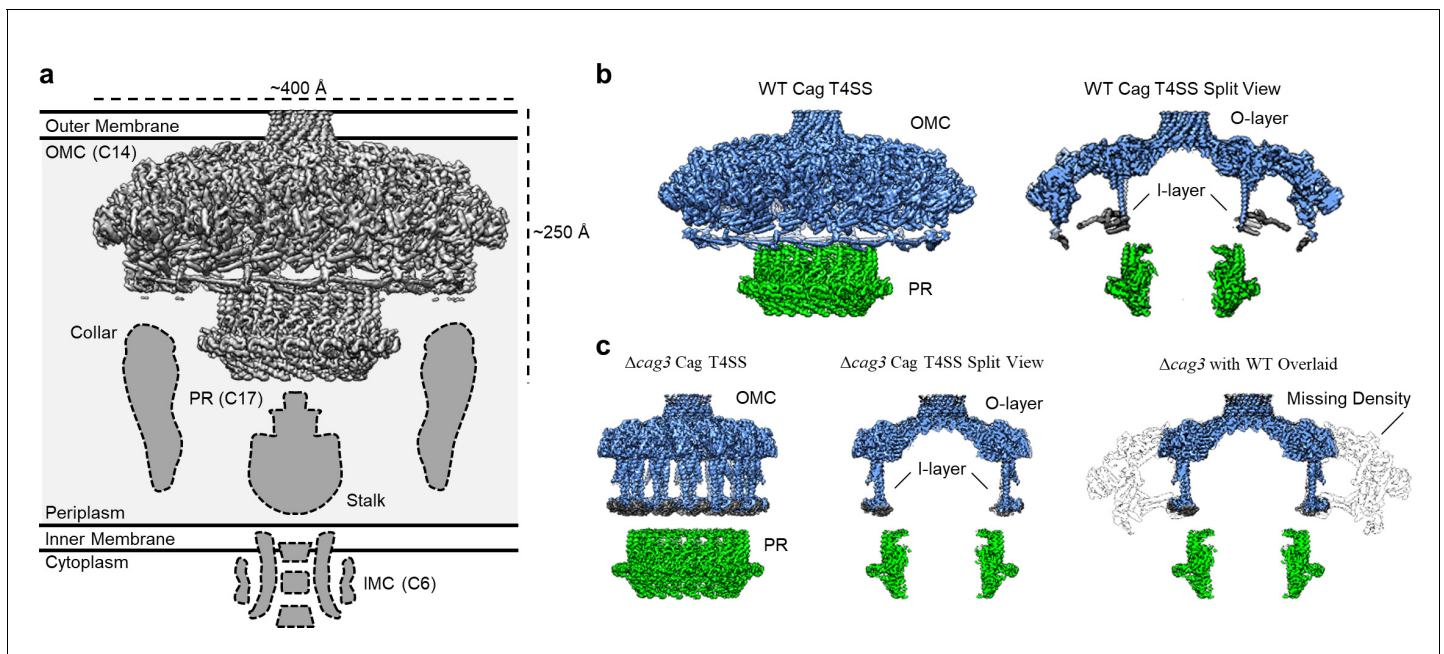


---

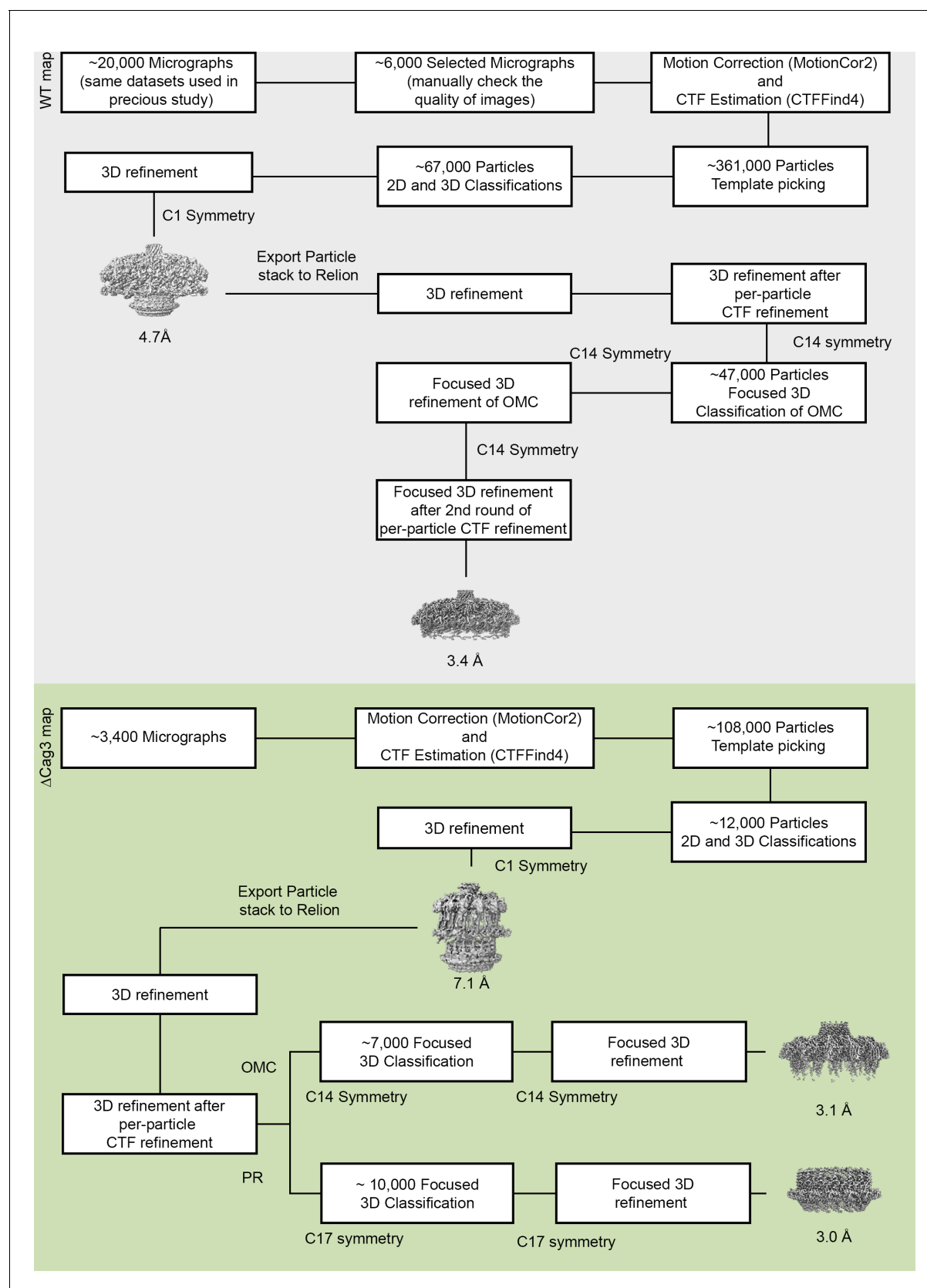
## Figures and figure supplements

Cryo-EM reveals species-specific components within the *Helicobacter pylori* Cag type IV secretion system core complex

**Michael J Sheedlo et al**



**Figure 1.** Comparison between maps of the wild-type (WT) and  $\Delta$ Cag3 Cag T4SS core complex. (a) The Cag T4SS spans the inner membrane and outer membrane and consists of distinct features with differing symmetry: the OMC (C14 symmetry), the PR (C17 symmetry), an inner-membrane complex (IMC, C6 symmetry), and the stalk and collar with unknown symmetries. (b) In the reconstruction of the WT Cag T4SS, we observe the O-layer of the OMC (shown in blue), the I-layer of the OMC (shown in gray), and the PR (shown in green). Left panel, WT Cag T4SS density map; Right panel, central slice of the WT density map. (c) All of these features are observed in the  $\Delta$ Cag3 Cag T4SS map (shown in the same colors as panel b), but peripheral regions of the OMC are missing in the  $\Delta$ Cag3 Cag T4SS map (shown in white). Left panel,  $\Delta$ Cag3 Cag T4SS density map; middle panel, central slice of  $\Delta$ Cag3 Cag T4SS density map; right panel, overlaid central slices of WT (grey) and  $\Delta$ Cag3 (blue and green) Cag T4SS density maps.

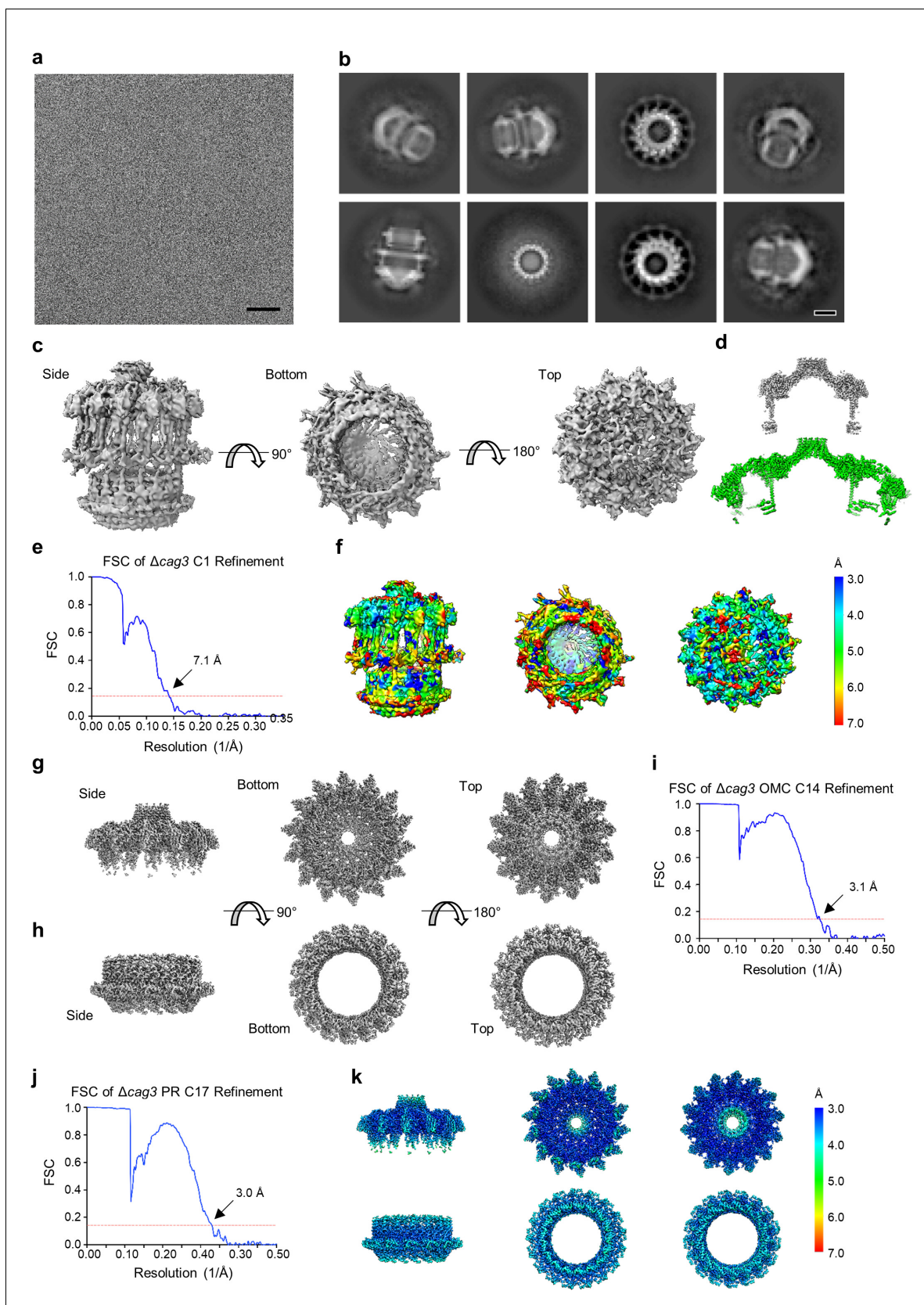


**Figure 1—figure supplement 1.** Flow chart of cryo-EM processing steps. The processing steps for WT Cag T4SS data are on a gray background and the processing steps for  $\Delta$ Cag3 Cag T4SS data are on a green background. All the raw micrographs used in the WT Cag T4SS 3D reconstructions were Figure 1—figure supplement 1 continued on next page

*Figure 1—figure supplement 1 continued*

identical to the datasets used in our previous study (**Chung et al., 2019**). Approximately 6000 high quality images were selected from ~20,000 images. ~361,000 particles were selected by template picking in CryoSPARC. After 2D and 3D classification, the best class (~67,000 particles) was chosen for further refinement, without symmetry (C1). All the particles used in the 3D refinement of CryoSPARC were exported into RELION for further steps and the 4.7 Å 3D model with C1 symmetry was filtered to 60 Å resolution and used as an initial model for 3D structure determination in RELION. Focused 3D classification and 3D refinement were used to determine higher resolution maps of the WT OMC (with 14-fold symmetry), resulting in a 3D map of the OMC (14-fold symmetry) at 3.4 Å. Similar steps were applied to  $\Delta$ Cag3 Cag T4SS data. Approximately 3400 micrographs were collected and ~108,000 particles were selected by template picking in CryoSPARC. After the 2D/3D filtration and further refinement (3D map at 7.1 Å with C1 symmetry), ~12,000 particles were exported into RELION for further steps. Focused 3D classification (without alignment) was used to determine higher resolution maps of the  $\Delta$ Cag3 OMC (with 14-fold symmetry) and the PR (with 17-fold symmetry). The maps of the OMC and PR were further refined using focused 3D refinement (with local refinement), resulting in 3D maps of the OMC (14-fold symmetry) and the PR (17-fold symmetry) at 3.1 Å and 3.0 Å, respectively.

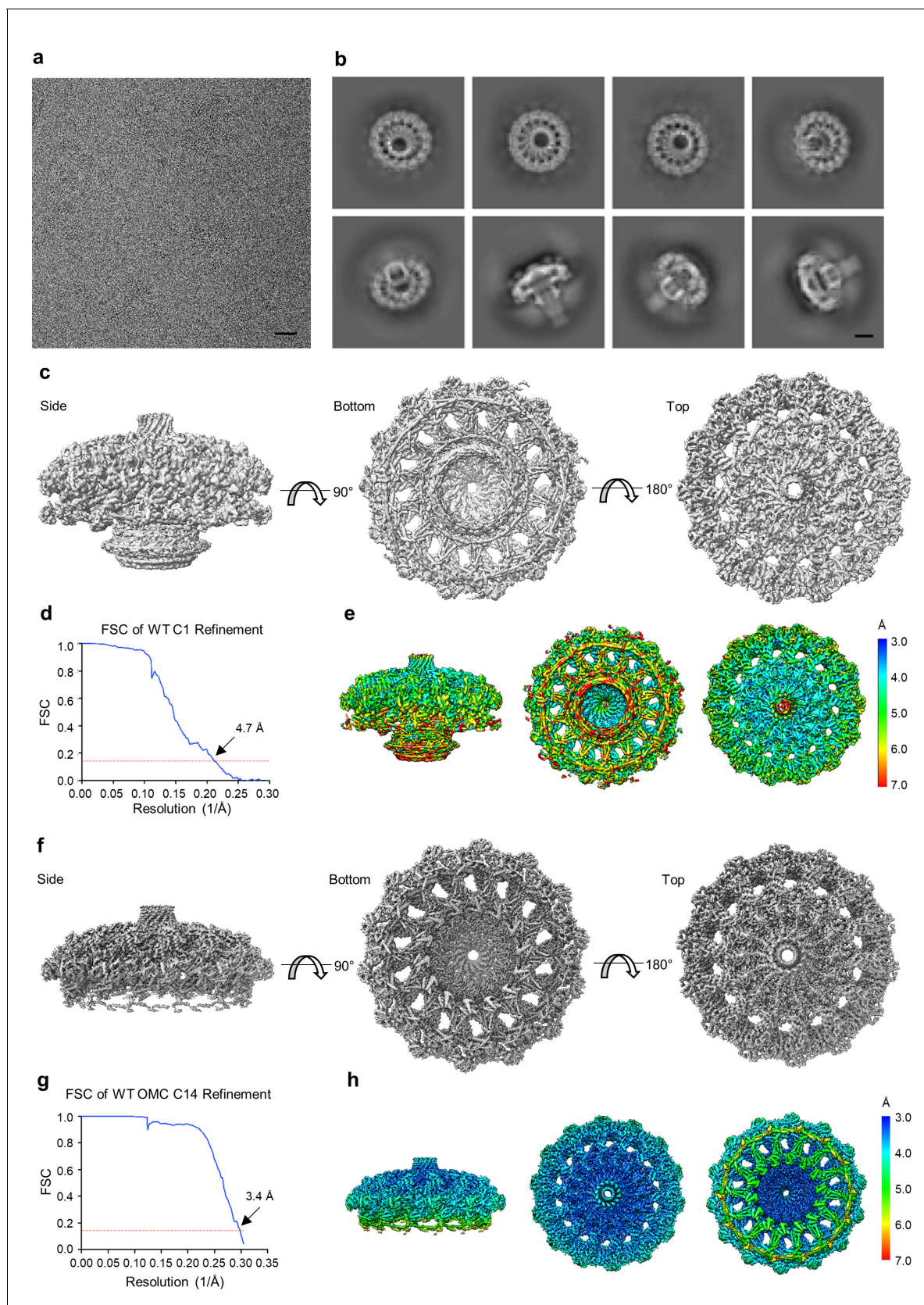




**Figure 1—figure supplement 2.** Cryo-EM processing of the  $\Delta$ Cag3 Cag T4SS core complex. (a) Sub-area of a motion-corrected image of vitrified  $\Delta$ Cag3 Cag T4SS particles. Scale bar, 50 nm. (b) Selected 2D class averages. Scale bar, 10 nm. (c) Rotated views of the 3D reconstructed model with no  
Figure 1—figure supplement 2 continued on next page

*Figure 1—figure supplement 2 continued*

applied symmetry (C1, 7.1 Å resolution). (d) A comparison between maps of the  $\Delta$ Cag3 (gray) and WT (green) OMC. (e) FSC of C1 refinement after applying a tight mask in cryoSPARC. Dotted red line, Fourier Shell Correlation (FSC) = 0.143. (f) Heat map showing local resolution of C1 3D reconstruction. The heat map scales are exponential. (g) Views of the focused refined 3D model of the  $\Delta$ Cag3 OMC with C14 symmetry applied (3.1 Å resolution) and (h)  $\Delta$ Cag3 PR with C17 symmetry applied (3.0 Å resolution). (i) FSC of focused  $\Delta$ Cag3 C14 OMC and (j)  $\Delta$ Cag3 C17 PR refinements. Dotted red line, FSC = 0.143. (k) Map of local resolution of focused  $\Delta$ Cag3 C14 OMC and  $\Delta$ Cag3 C17 PR refinement. The heat map scales are exponential.



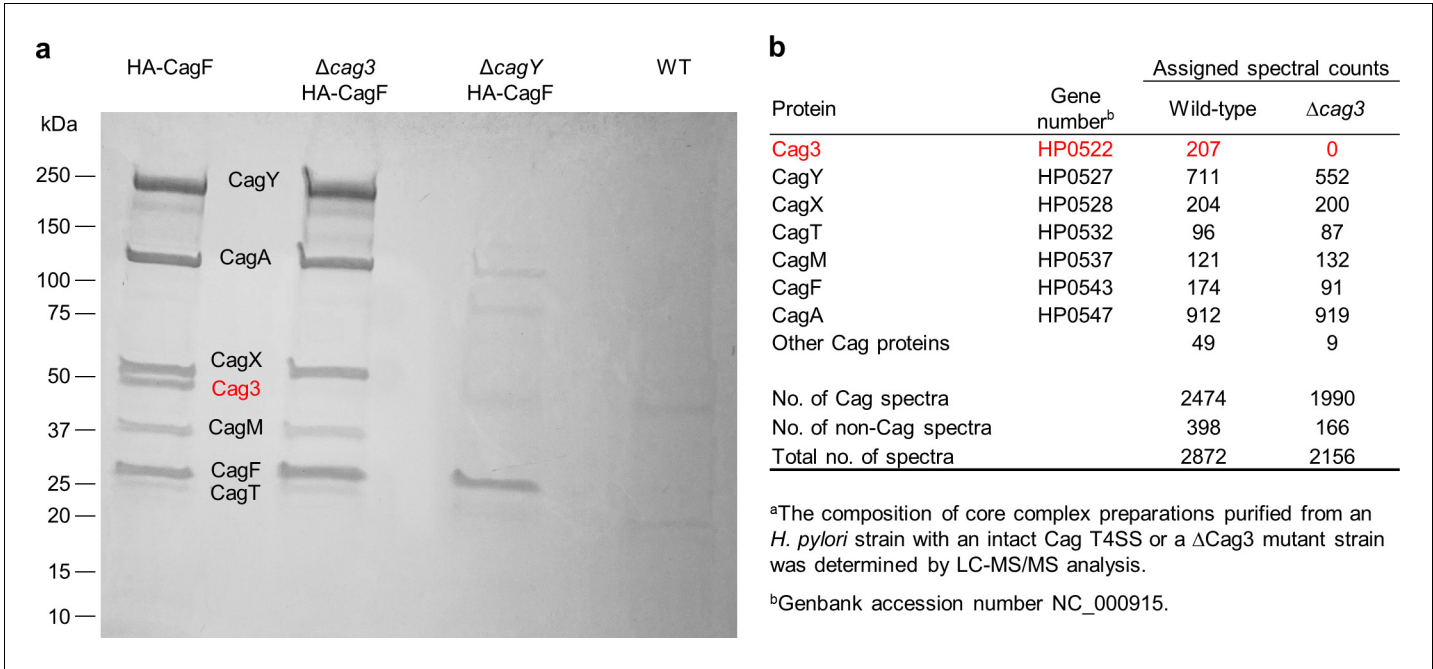
**Figure 1—figure supplement 3.** Cryo-EM processing of the WT Cag T4SS core complex. (a) Sub-area of a motion-corrected image of vitrified WT Cag T4SS particles. Scale bar, 50 nm. (b) Selected 2D class averages. Scale bar, 10 nm. (c) Rotated views of the 3D reconstructed model of the WT Cag T4SS

Figure 1—figure supplement 3 continued on next page

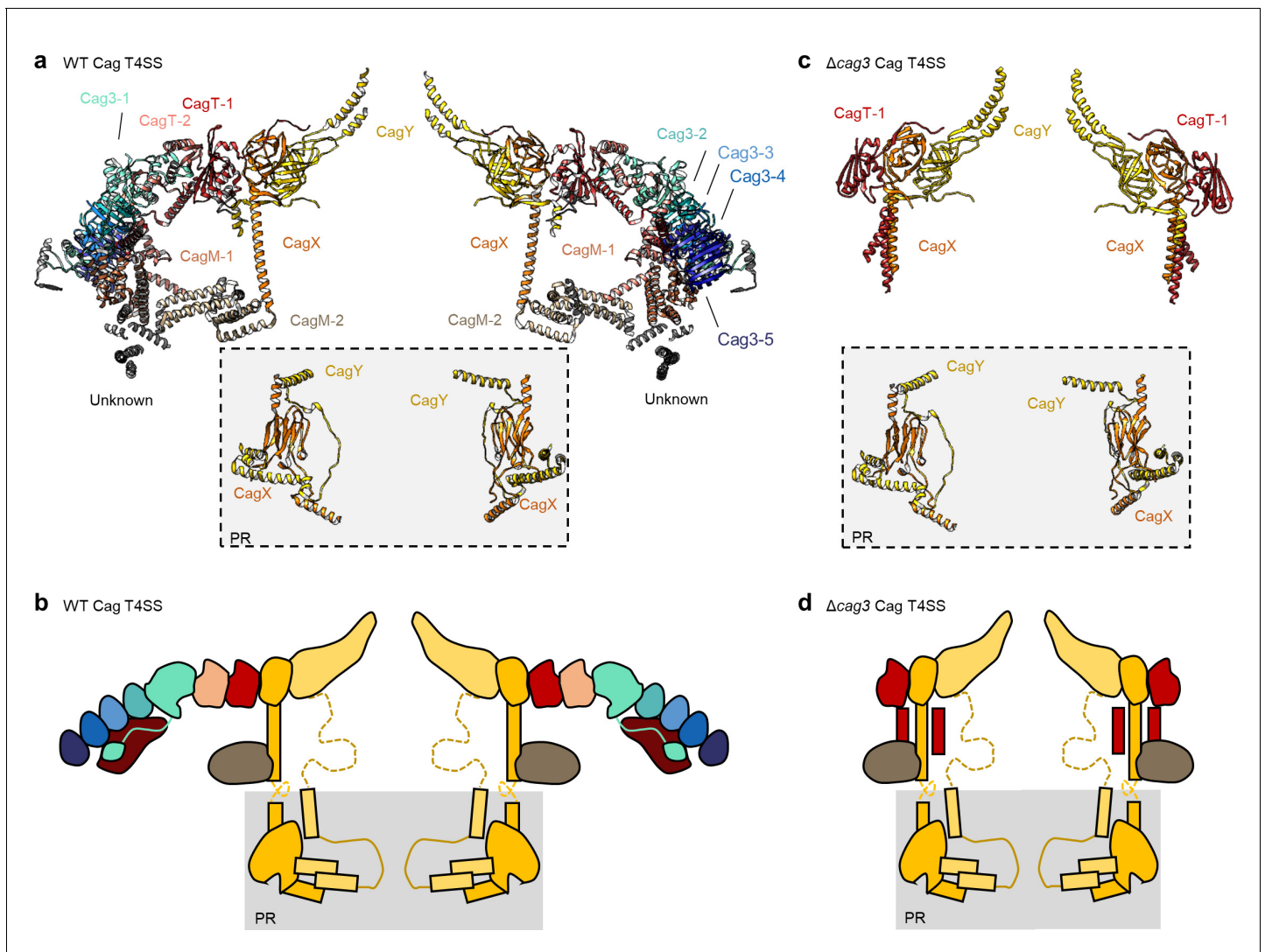
*Figure 1—figure supplement 3 continued*

with no applied symmetry (C1, 4.7 Å resolution). (d) FSC of C1 refinement after applying a tight mask in cryoSPARC. Dotted red line, Fourier Shell Correlation (FSC) = 0.143. (e) Heat map showing local resolution of C1 3D reconstruction. The heat map scales are exponential. (f) Views of the focused refined 3D model of the WT OMC with C14 symmetry applied (3.4 Å resolution). (g) FSC of focused C14 OMC refinement. Dotted red line, FSC = 0.143. (h) Map of local resolution of focused WT OMC refinement. The heat map scales are exponential.

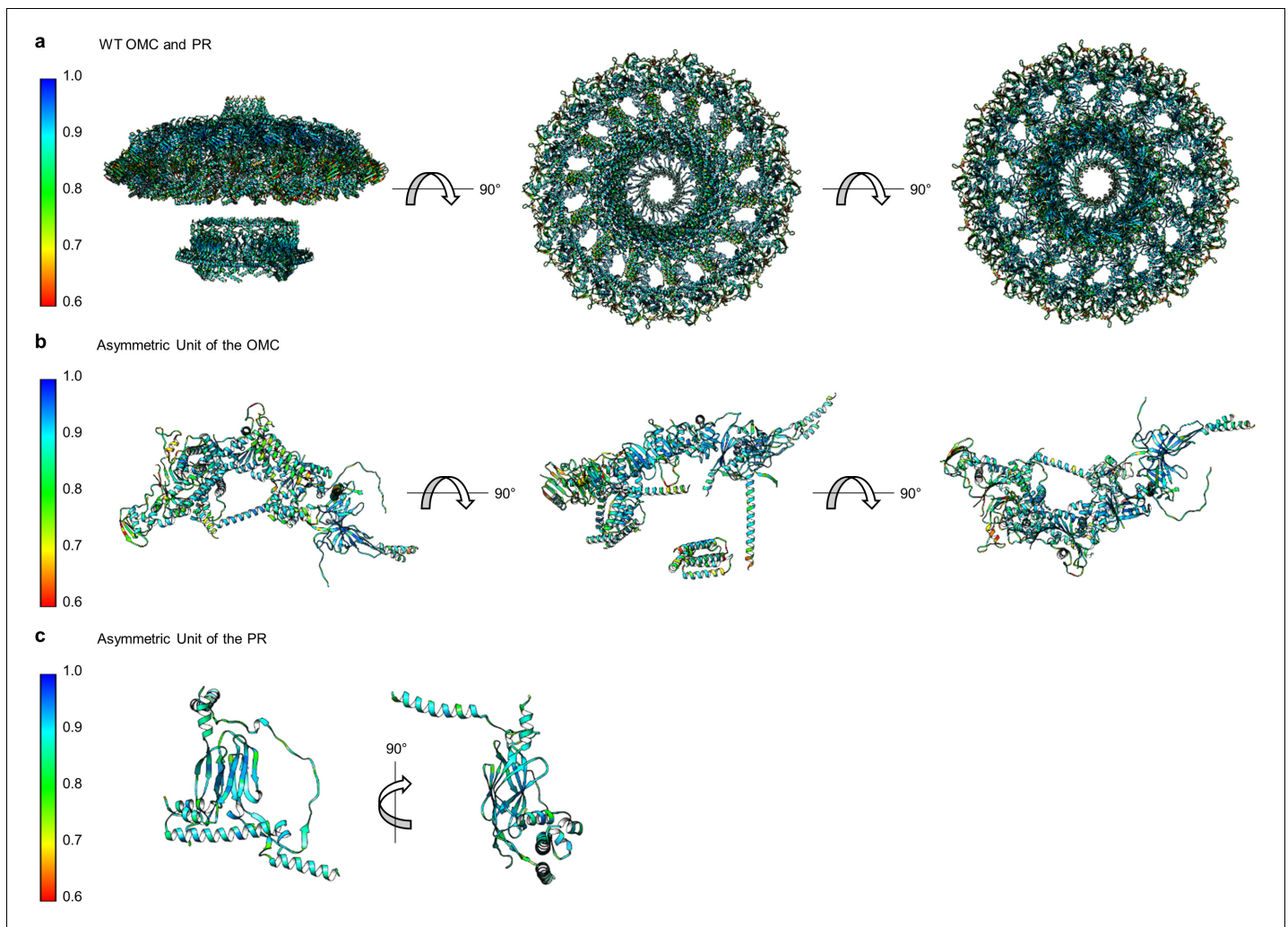




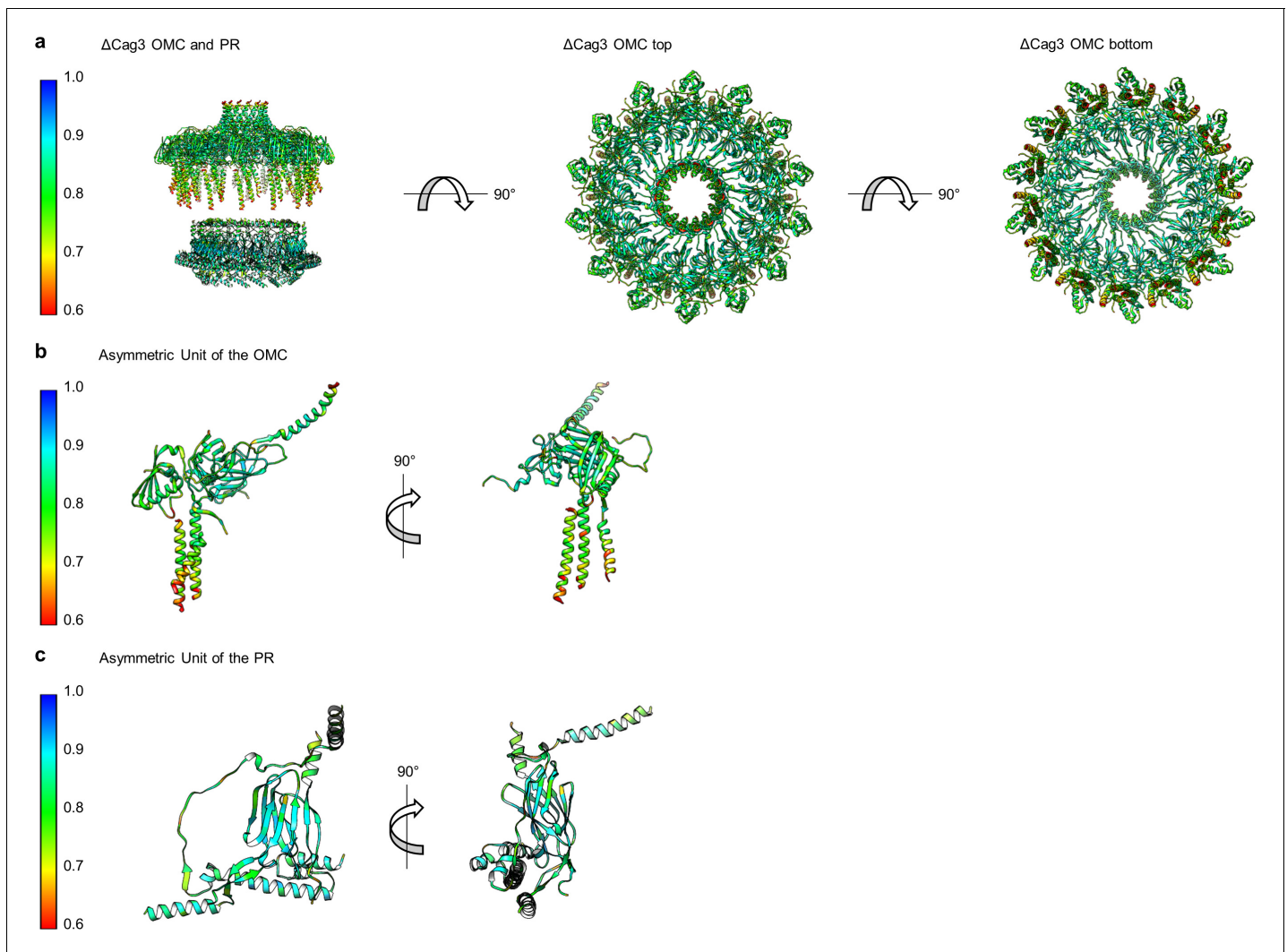
**Figure 1—figure supplement 4.** Analysis of Cag T4SS core complex preparations. (a) SDS-PAGE analysis of core complexes purified from wild-type and  $\Delta$ cag3 *H. pylori* strains, each producing HA-tagged CagF (first and second lanes). The same purification methods were applied to a  $\Delta$ cagY mutant strain producing HA-tagged CagF (defective in core complex assembly) or a control wild-type strain with no modification of CagF. Dominant bands visible in the first lane have molecular masses consistent with CagY (219 kDa), CagA (132 kDa), CagX (61 kDa), Cag3 (55 kDa), CagM (44 kDa), CagT (32 kDa), and CagF (32 kDa). LC-MS/MS analysis of excised bands confirmed these identifications. The 55 kDa band (indicated in red) is absent from the  $\Delta$ Cag3 mutant core complex preparation. (b) Mass spectrometric analysis of core complex samples purified from the WT and  $\Delta$ Cag3 *H. pylori* strains.



**Figure 2.** The asymmetric unit of the WT and  $\Delta Cag3$  Cag T4SS core complex. (a) A cross-section model of the mapped portions of CagY (yellow), CagX (orange), CagT (red and salmon), CagM (brown and tan), and Cag3 (various shades of green and blue) in the OMC from the WT Cag T4SS. Portions of CagX and CagY are present in both the OMC and the PR (denoted by the gray box). (b) A cartoon representation of the WT Cag T4SS core complex proteins, colored as in panel a. The dotted lines represent densities that were not clearly seen in the density maps. (c) A cross-section model of the mapped portions of CagY (yellow), CagX (orange), and CagT (red) from the  $\Delta Cag3$  Cag T4SS core complex. (d) A cartoon representation of the  $\Delta Cag3$  Cag T4SS core complex proteins with CagY (yellow), CagX (orange), CagT-1 (red). The red rectangles represent the alternate conformation of the C-terminal helices of CagT-1. CagT-2 and Cag-3 were not observed within this map. The dashed lines represent densities that were not clearly seen in the density map.



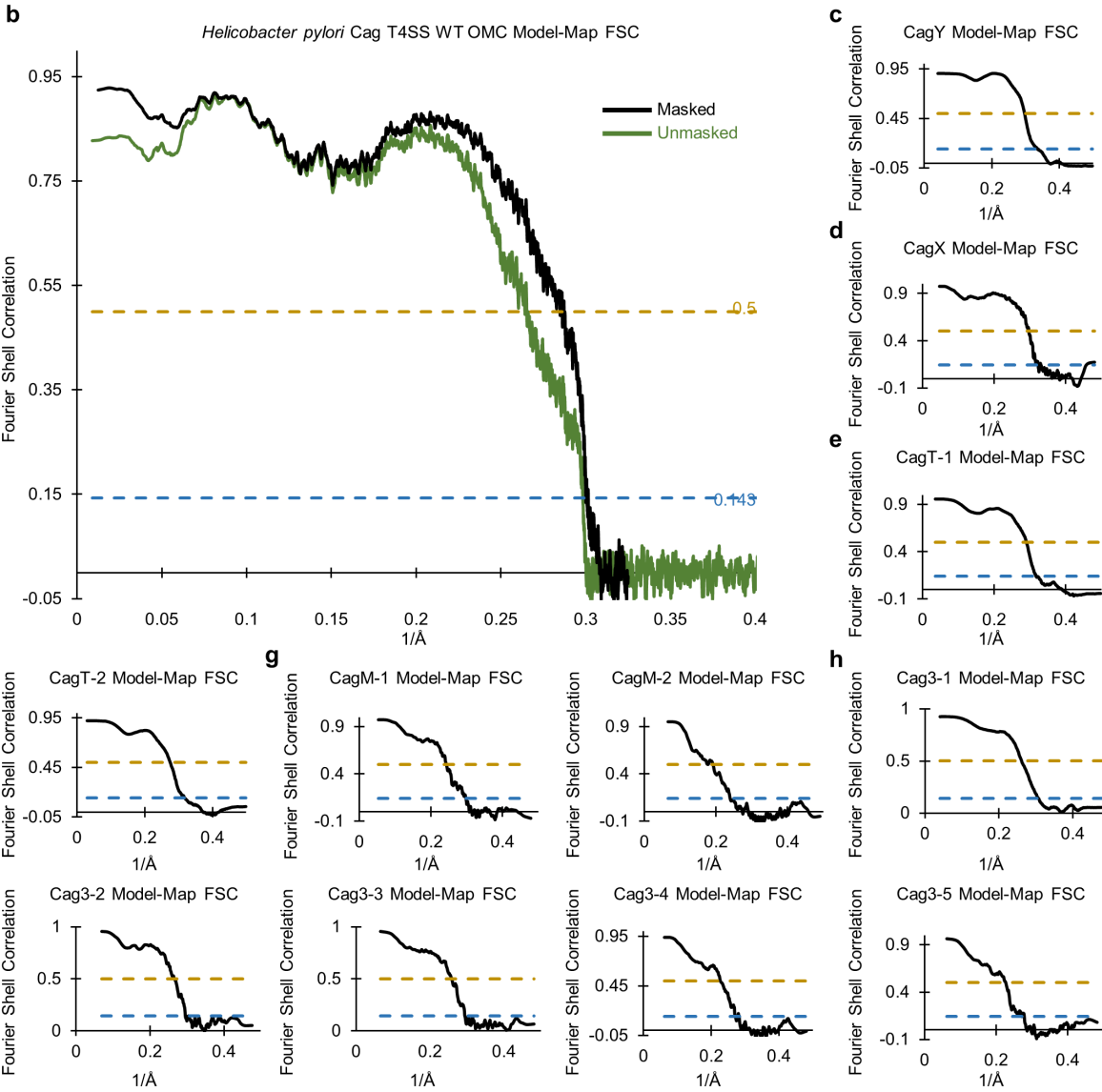
**Figure 2—figure supplement 1.** Correlation of each component of the WT Cag T4SS to experimental maps. (a) The correlation coefficient of each residue is plotted on the structure of the entire WT Cag T4SS as a heat map. The asymmetric units of the OMC (b) and the PR (c) depict the relative quality of the models in each portion of the structure.



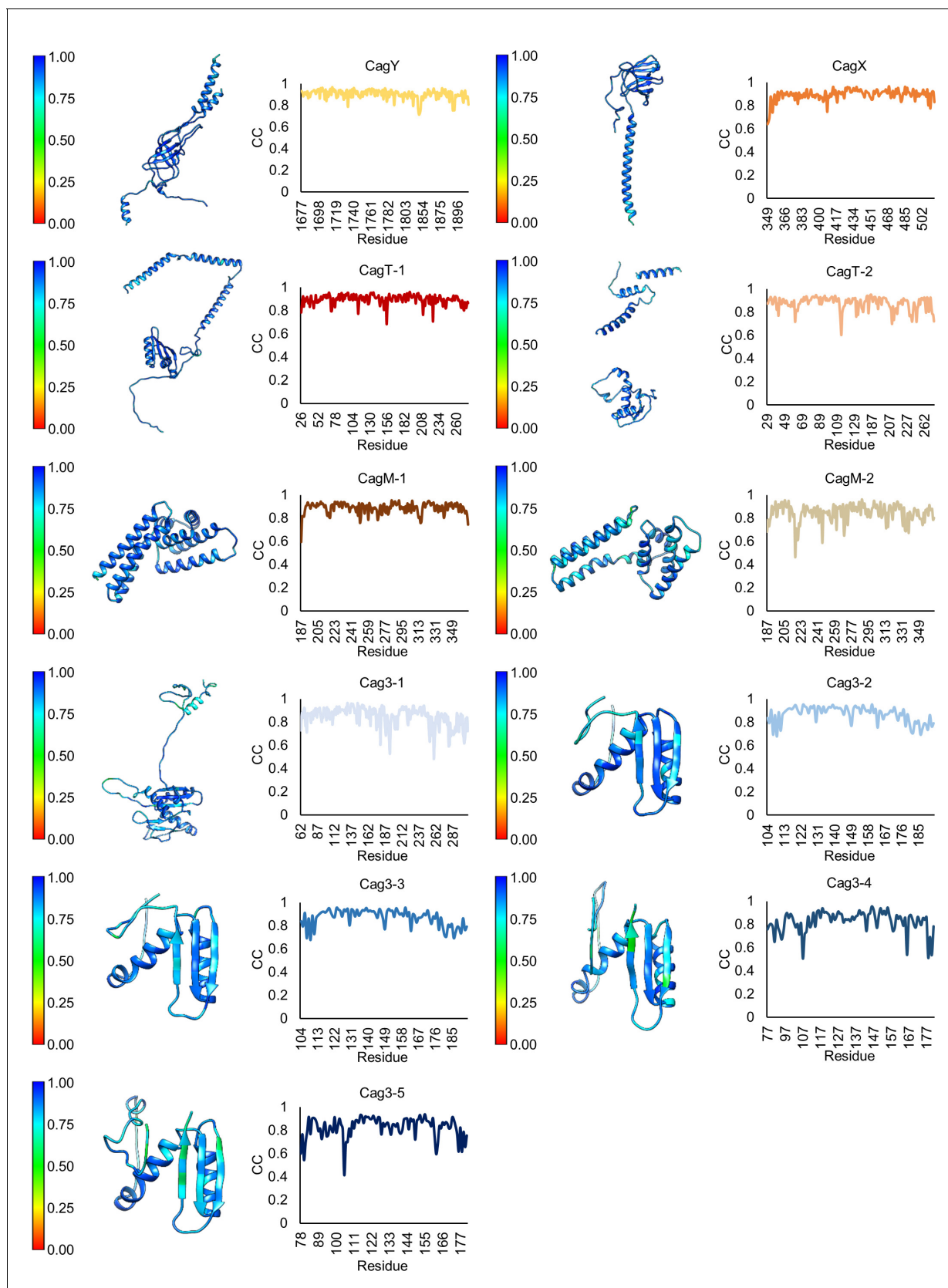
**Figure 2—figure supplement 2.** Correlation of each component of the  $\Delta$ Cag3 Cag T4SS to experimental maps. (a) The correlation coefficient of each residue is plotted on the structure of the entire  $\Delta$ Cag3 Cag T4SS as a heat map. The OMC is rotated to show 'top' and 'bottom' views. The asymmetric units of the OMC (b) and the PR (c) depict the relative quality of the models in each portion of the structure.



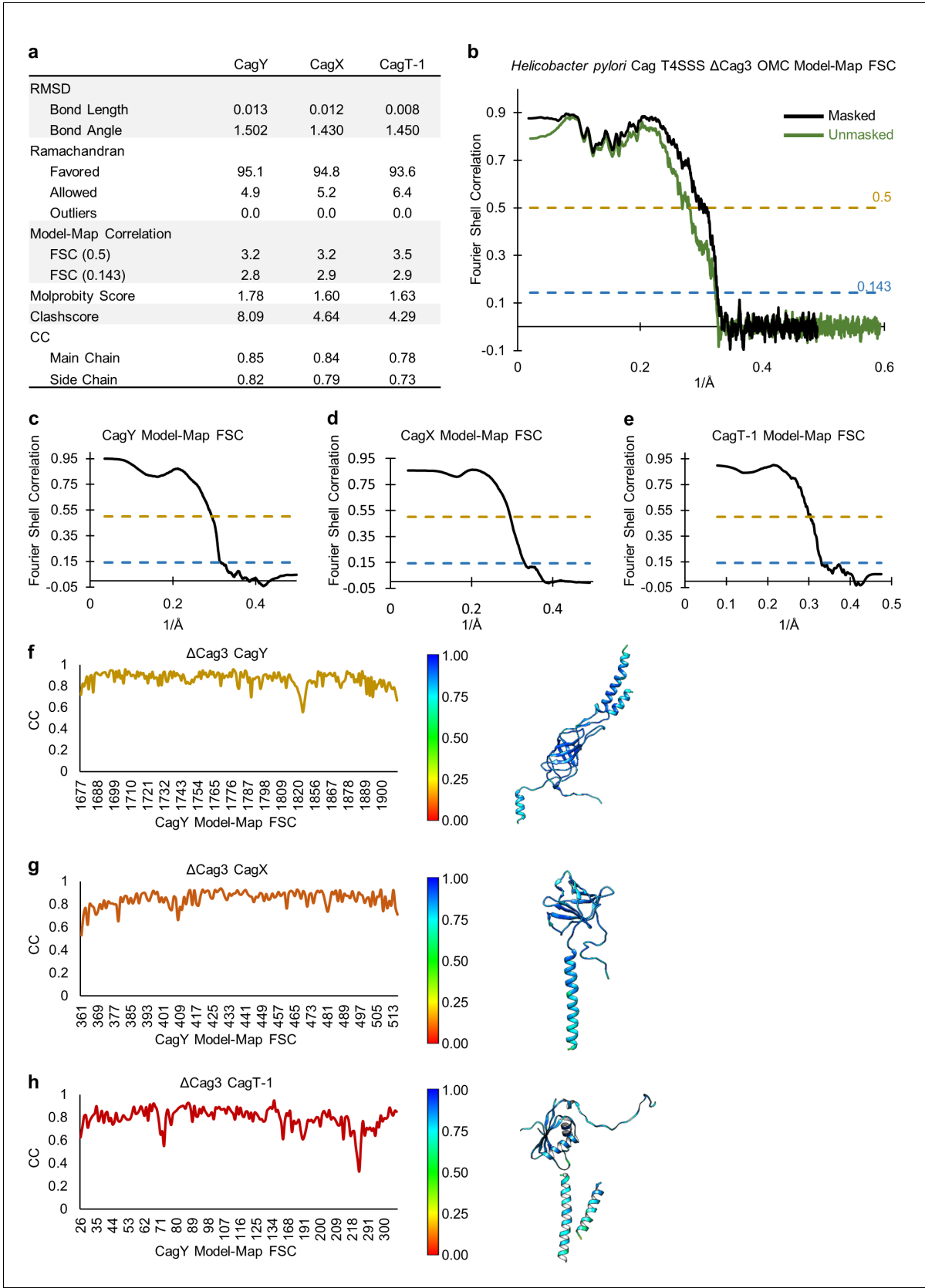
<b>a</b>	CagY	CagYX	CagT-1	CagT-2	CagM-1	CagM-2	Cag3-1	Cag3-2	Cag3-3	Cag3-4	Cag-5
RMSD											
Bond Length	0.012	0.011	0.010	0.009	0.009	0.009	0.010	0.016	0.014	0.017	0.015
Bond Angle	1.585	1.715	1.419	1.490	1.479	1.664	1.863	1.893	1.847	2.236	2.231
Ramachandran											
Favored	95.5	94.5	96.8	94.1	94.9	94.9	89.4	92.1	96.7	84.6	96.1
Allowed	4.5	5.5	3.2	5.9	5.1	5.1	10.6	7.9	3.3	15.4	3.9
Outliers	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Model-Map Corr.											
FSC (0.5)	3.4	3.4	3.4	3.6	4.1	5.3	3.8	3.7	3.9	4.3	4.4
FSC (0.143)	3.0	3.1	3.1	3.1	3.4	4.1	3.2	3.4	3.4	3.6	3.6
Molprob. Score	2.31	2.12	1.96	2.15	1.75	1.88	2.36	1.96	1.65	2.12	1.87
Clashscore	9.02	3.27	6.46	6.49	7.36	10.12	7.60	8.87	8.03	8.34	12.44
CC											
Main Chain	0.88	0.88	0.87	0.84	0.85	0.81	0.83	0.82	0.80	0.78	0.79
Side Chain	0.86	0.86	0.84	0.83	0.84	0.81	0.80	0.82	0.79	0.74	0.77



**Figure 2—figure supplement 3.** Correlation between the WT OMC cryo-EM map and models. (a) Statistics for each of the constructed models indicate their relative quality. (b) A model-map correlation curve for the entire WT OMC is shown, with the unmasked data in green and the masked data in black. (c–h). The model-map correlation was also determined for each of the individual protein structures within the OMC.

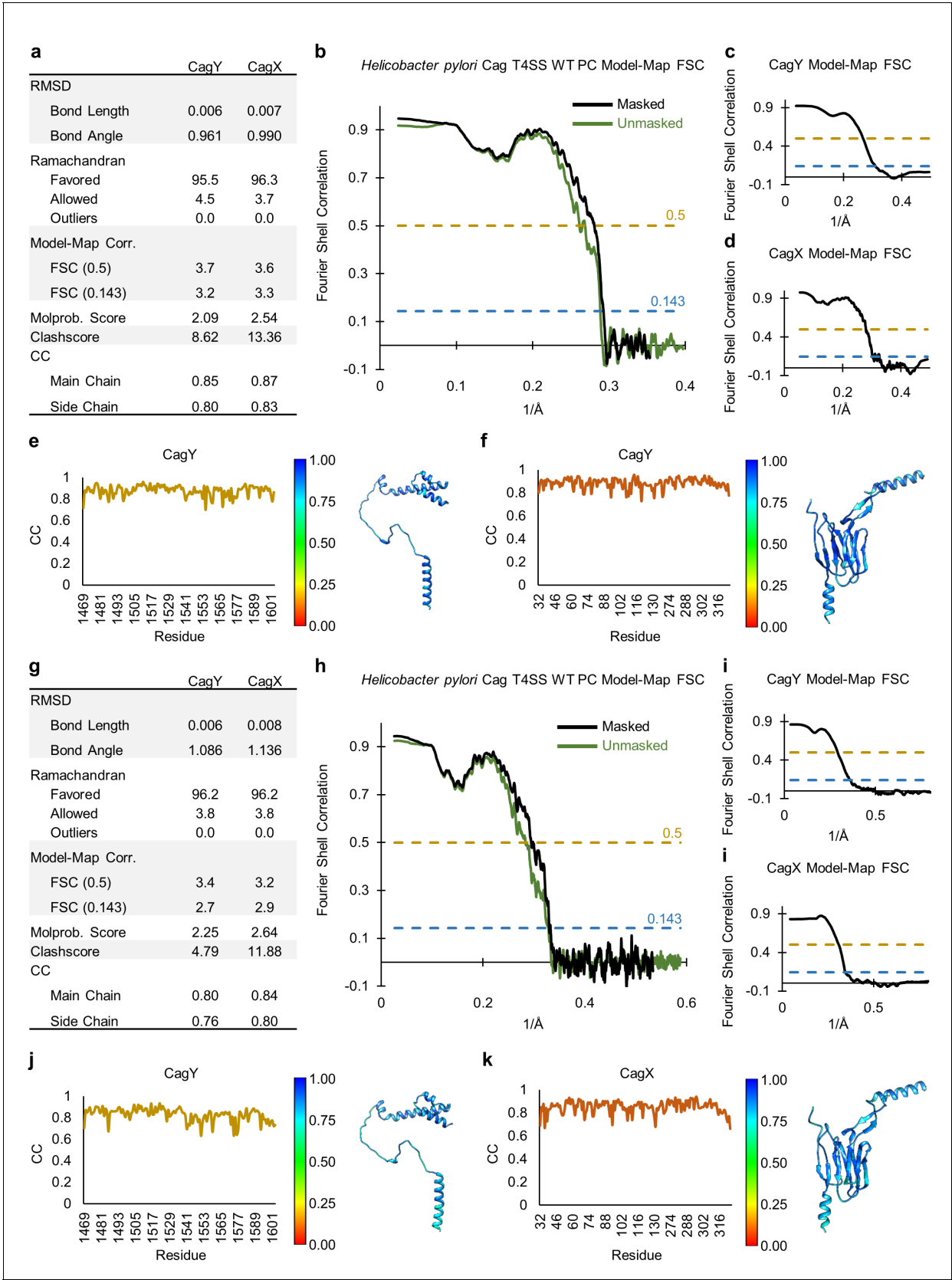


**Figure 2—figure supplement 4.** Model-map correlation for each protein within the OMC. (a) The correlation coefficient for each protein within the OMC is depicted as a heat map on the corresponding structure and is plotted against the residue number.



*Figure 2—figure supplement 5 continued*

unmasked map and the black line indicates data derived using a masked map. (c–e) A model-map FSC curve is shown for all models constructed in the  $\Delta$ Cag3 OMC map. (f–h) Per-residue correlation coefficients are plotted against residue number (left) and are depicted as a heat map on the corresponding structures (right).



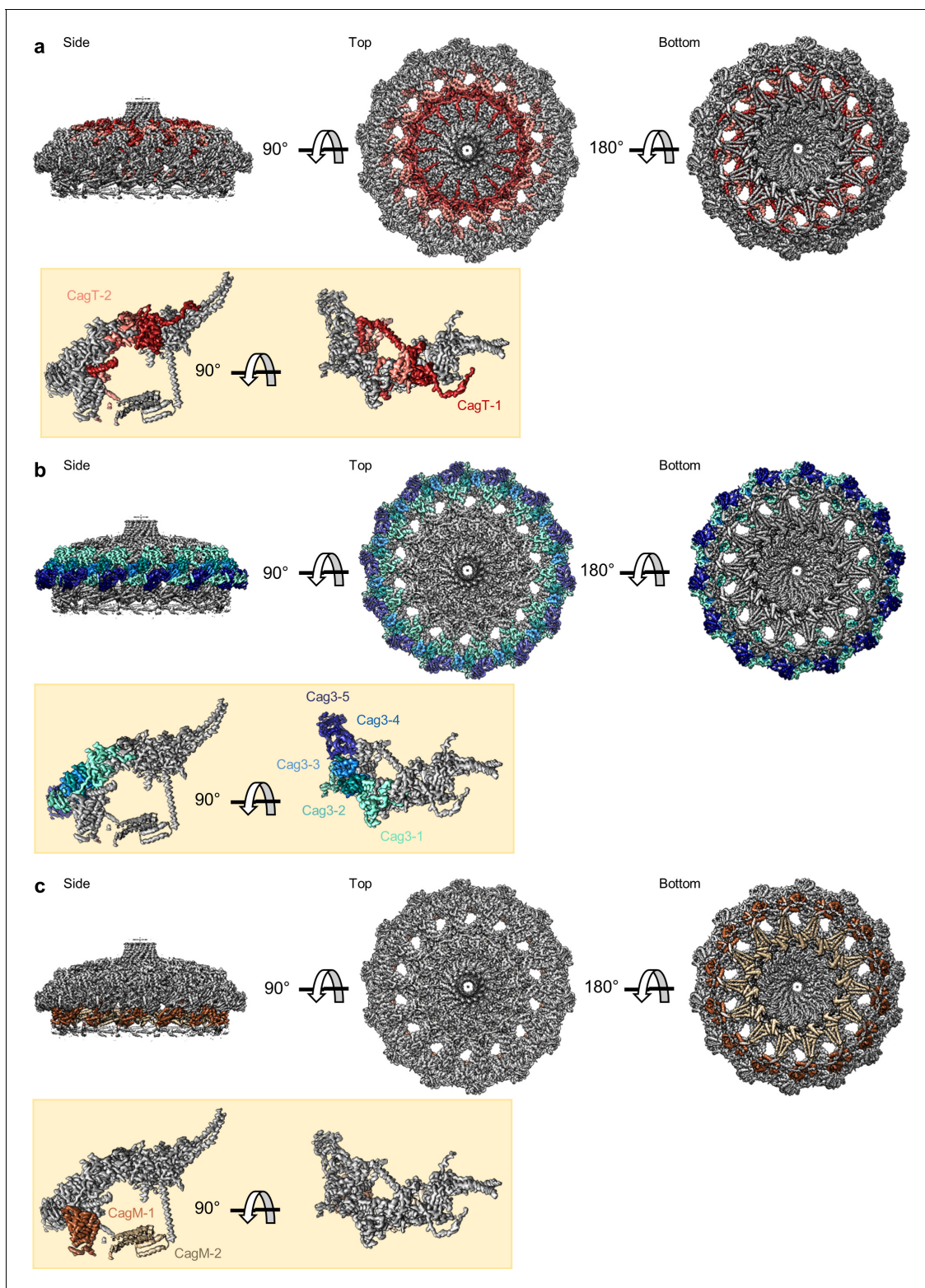
**Figure 2—figure supplement 6.** Quality of models within the PR of the WT and  $\Delta$ Cag3 maps. (a) Statistical analysis of all models constructed within the PR of the WT T4SS. (b) A model-map FSC curve is shown for the PR of the WT T4SS. The green line indicates data derived using an unmasked map

Figure 2—figure supplement 6 continued on next page

*Figure 2—figure supplement 6 continued*

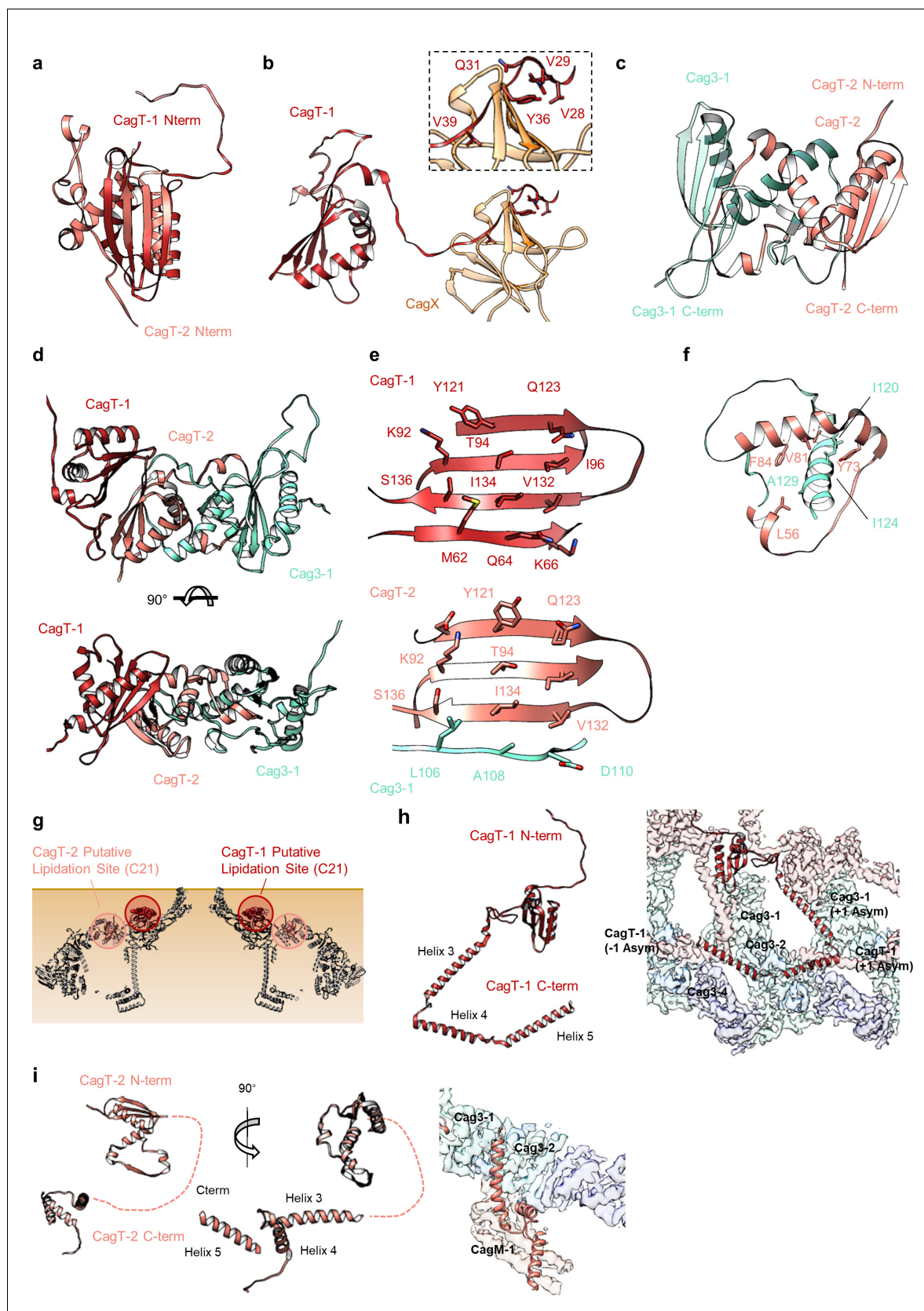
and the black line indicates data derived using a masked map. (c, d) A model-map FSC curve is shown for CagY and CagX. (e, f) Per-residue correlation coefficients are plotted against residue number (left) and are depicted as a heat map on the corresponding structure (right). (g) Statistical analysis of all models constructed within the PR of the  $\Delta$ Cag3 T4SS. (h) A model-map FSC curve is shown for the PR of the  $\Delta$ Cag3 T4SS. The green line indicates data derived using an unmasked map and the black line indicates data derived using a masked map. (i) A model-map FSC curve is shown for CagY and CagX. (j, k) Per-residue correlation coefficients are plotted against residue number (left) and are depicted as a heat map on the corresponding structures (right).





**Figure 3.** Locations of components within the OMC. (a) The O-layer of the OMC contains two copies of CagT (CagT-1 and CagT-2) (shown in red and salmon). The two copies are localized within the center of the asymmetric unit (inset) as shown in red and salmon. (b) Cag3 comprises a significant portion of the O-layer of the Cag T4SS, as shown in blue and green. Within the asymmetric unit (inset), we observe five copies of Cag3 (denoted Cag3-1 through Cag3-5, shown in various shades of blue and green). (c) Within the I-layer of the OMC, there are two similar folds, each defined as CagM. Within the asymmetric unit, we observe two copies (inset) that are colored in brown and tan.

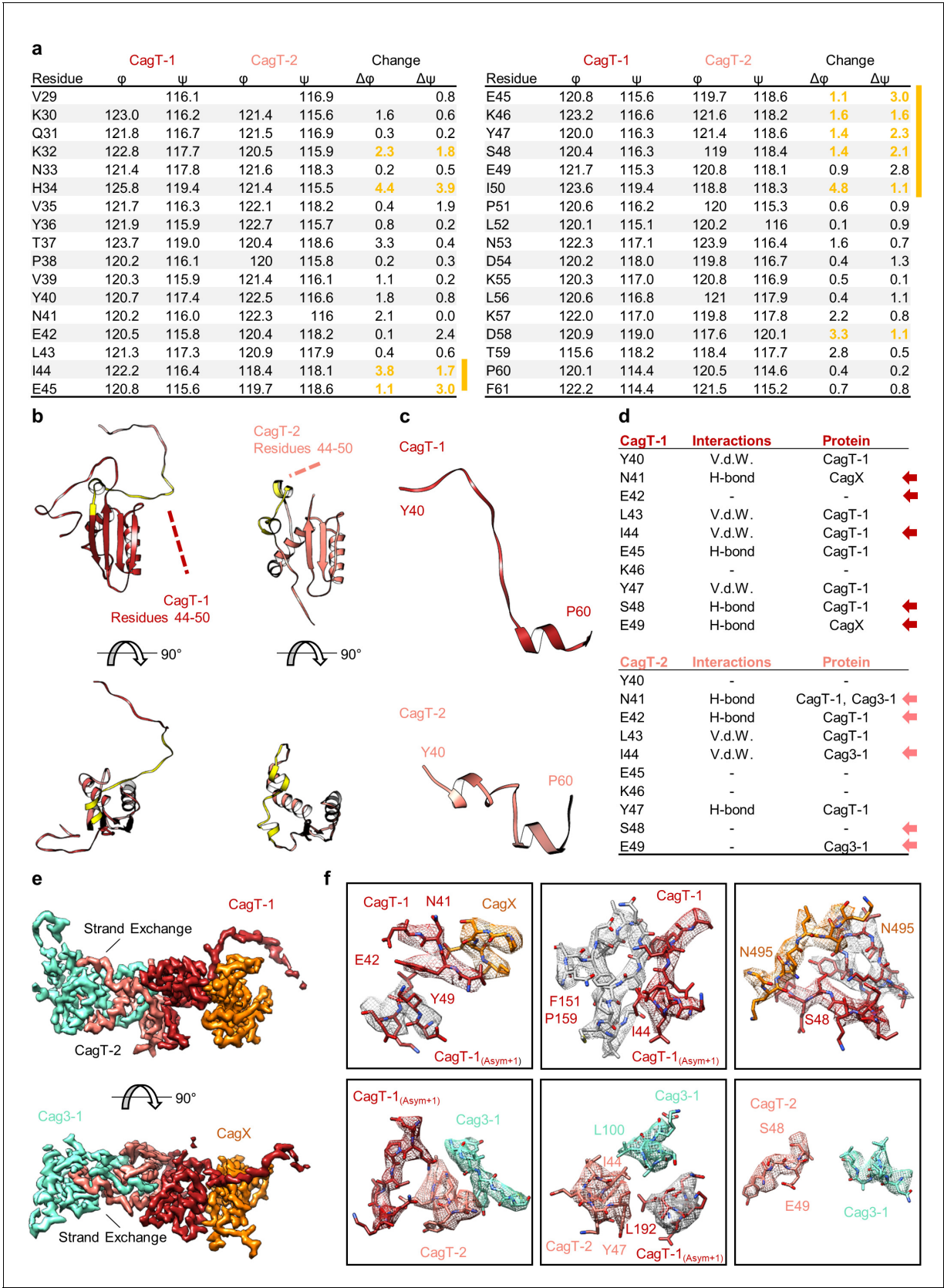




**Figure 4.** The N- and C-termini from CagT-1 and CagT-2 are positioned differently. (a–c) The N-terminus of CagT-1 extends inward toward the center of the map and interacts with CagX from the next asymmetric unit. The residues that contribute to this interaction are largely hydrophobic, as indicated Figure 4 continued on next page

*Figure 4 continued*

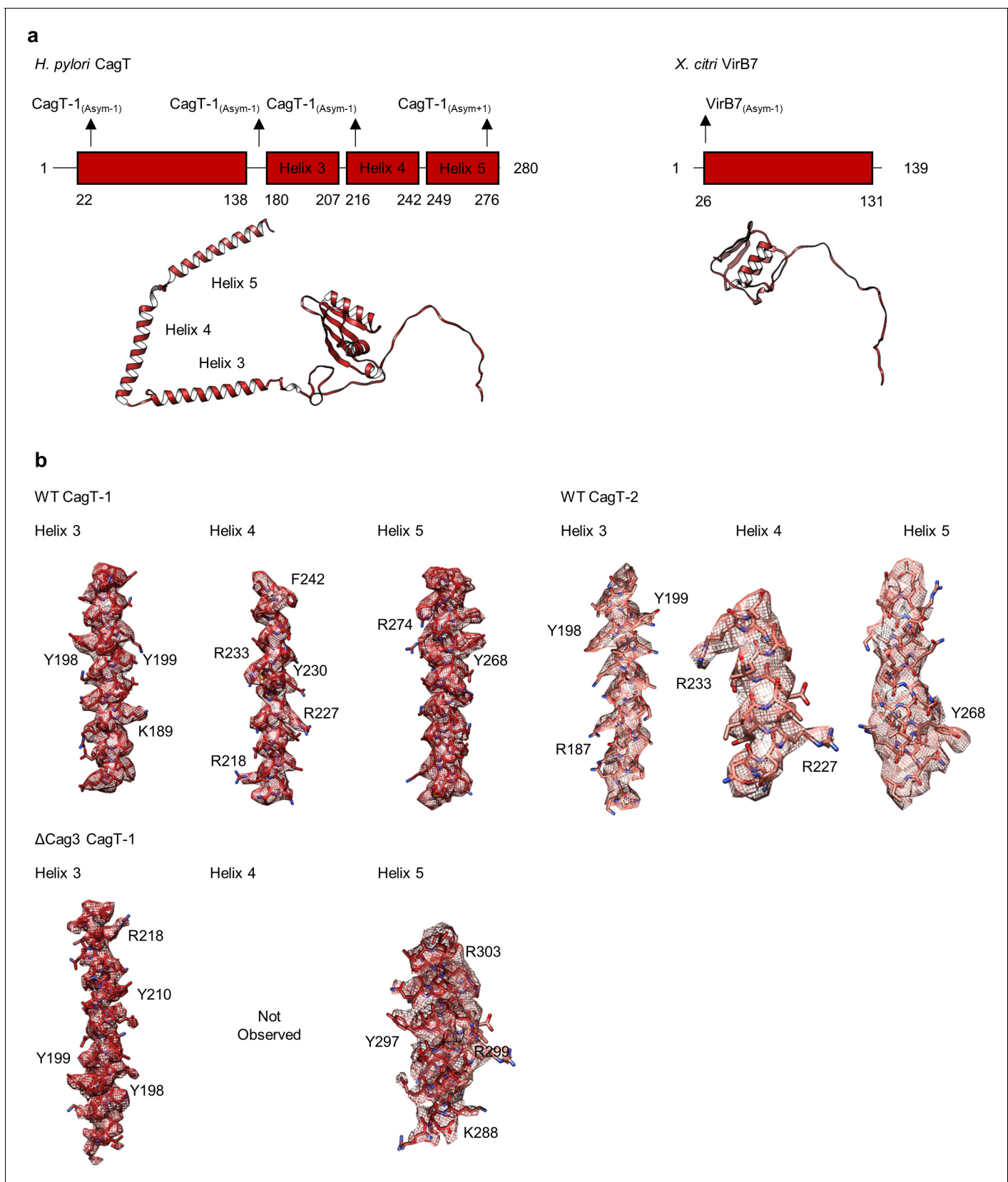
in the inset panel. CagT-2 differs from CagT-1 in that the N-terminus of the protein extends outward toward the periphery of the map, forming the last strand of a  $\beta$ -sheet with Cag3-1. (d) The three proteins (CagT-1, CagT-2 and Cag3-1) have an interwoven architecture. (e) The interface that is formed between the three proteins consists of two  $\beta$ -sheets that include strands from all three molecules. (f) The interface of CagT-2 and Cag3-1 is a pair of  $\alpha$ -helices that bury hydrophobic residues within the interface. (g) The position of the N-terminal loops of CagT-1 (red) and CagT-2 (salmon) are such that the putative lipidation sites are near the outer membrane. (h) The C-terminal  $\alpha$ -helices of CagT-1 adopt an extended conformation (left) and interact with Cag3-1, Cag3-2, and Cag3-4 within the same asymmetric unit and Cag3-1 and CagT-1 in neighboring asymmetric units (right). (i) The C-terminal  $\alpha$ -helices of CagT-2 are connected by an apparently flexible linker (left) and interact with Cag3-1, Cag3-2, and CagM-1 (right).



**Figure 4—figure supplement 1.** The N-terminal loop of CagT is observed in two different conformations. (a) The dihedral angles of residues within the N-terminal loop of CagT-1 and CagT-2 are plotted. The difference between the angles was used to determine where the largest change occurred. Figure 4—figure supplement 1 continued on next page

*Figure 4—figure supplement 1 continued*

Residues in which both the  $\phi$  and  $\psi$  angles changed by  $>1^\circ$  are highlighted in yellow. **(b)** The N-terminal loop of CagT-1 (red) and CagT-2 (salmon) is highlighted in yellow. **(c)** The residues of CagT-1 and CagT-2 that exhibited the most striking change are depicted. **(d)** Residues within these loops engage in protein-protein interactions with different subsets of proteins. The table lists relevant residues along with the type of interaction mediated and the interacting protein. Arrows indicate residues with different interacting partners. **(e)** The orientations of the N-terminal loops are shown. A loop of CagT-1 interacts with CagX and a loop of CagT-2 completes a  $\beta$ -sheet formed by Cag3. **(f)** The interactions mediated by the N-terminal loops are shown here. The top row depicts interactions mediated by the N-terminal loop of CagT-1, and the bottom row depicts interactions mediated by the N-terminal loop of CagT-2. Proteins are colored as they appear in panel **(e)**, with proteins from adjacent asymmetric units depicted with gray density.



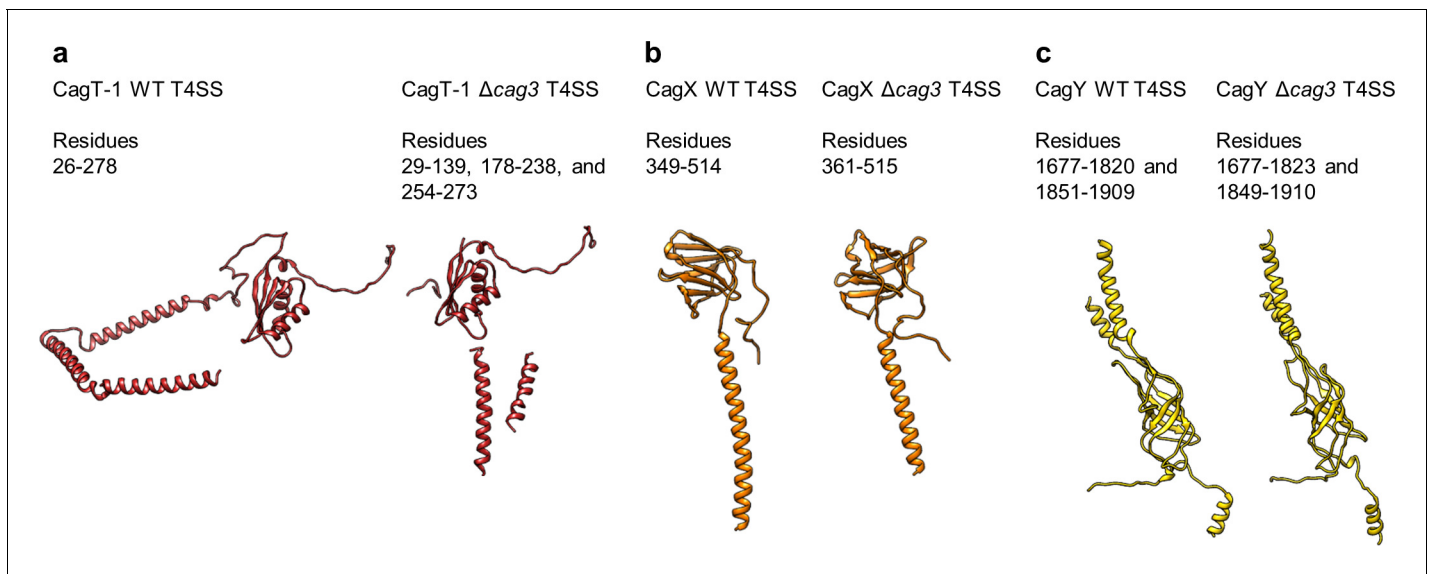
**Figure 4—figure supplement 2.** The C-terminal helices of CagT mediate interactions within the OMC. (a) Domain diagrams of *H. pylori* CagT and *X. citri* VirB7. The two proteins have similar N-terminal domains that mediate interactions with other components in the respective T4SSs. CagT contains

Figure 4—figure supplement 2 continued on next page

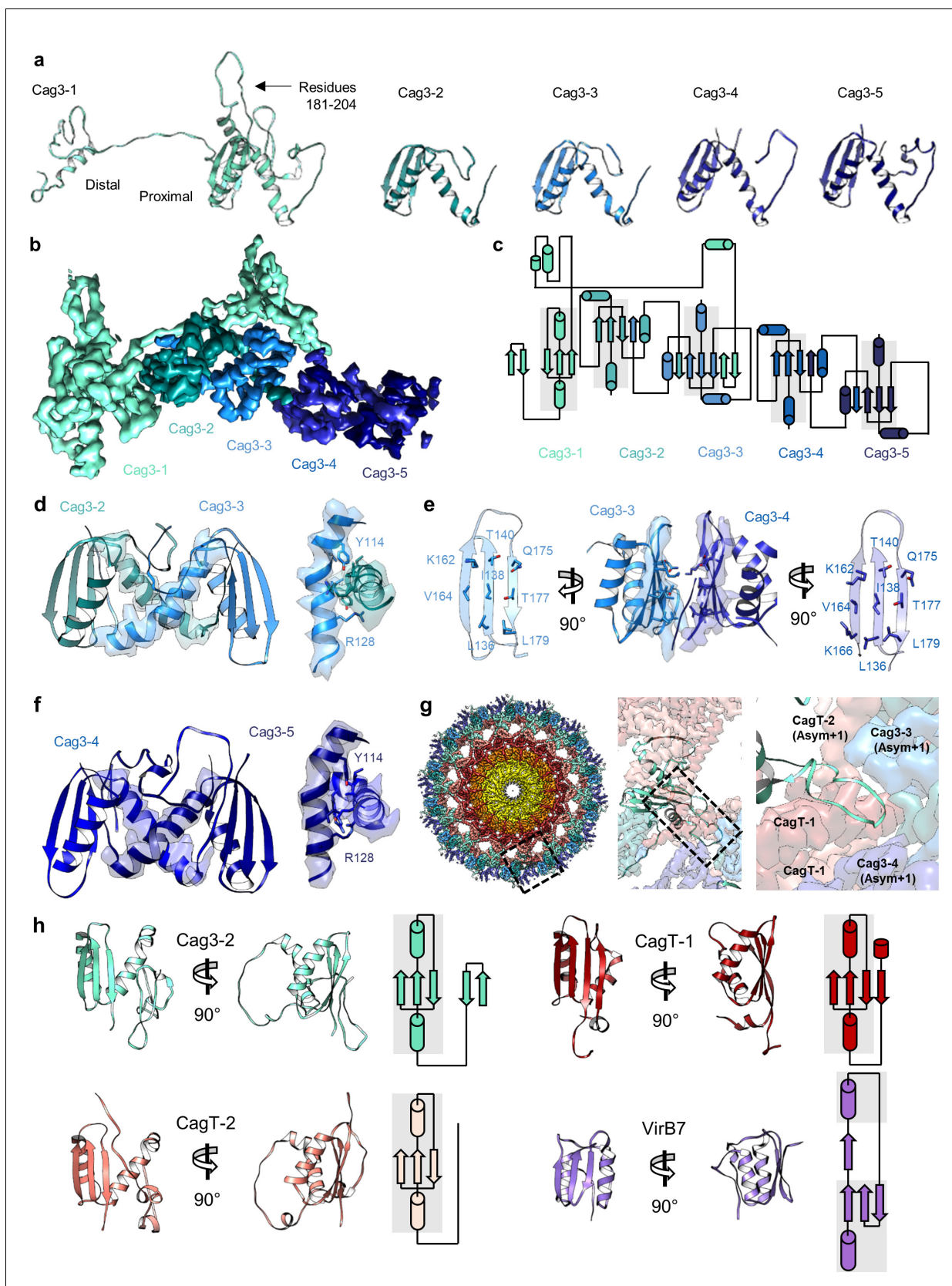
*Figure 4—figure supplement 2 continued*

three additional helices within the C-terminus of the protein that mediate interactions with adjacent asymmetric units. **(b)** The three C-terminal helices of CagT-1 and CagT-2 observed within the WT map are shown with experimental density displayed as a mesh (top). Landmark residues demonstrating the accuracy of the register are shown. The two C-terminal helices of CagT-1 that were observed within the  $\Delta$ Cag3 map are shown with experimental density displayed as a mesh (bottom). Landmark residues demonstrating the accuracy of the register are shown.





**Figure 5.** Conservation of protein structures in WT and  $\Delta$ Cag3 Cag T4SS complexes. (a) Despite global differences between the overall WT and  $\Delta$ cag3 maps, the structure of CagT-1 shares a similar core VirB7-like fold in both maps. (b–c) Similarly, CagX and CagY adopt nearly identical orientations within both maps. The illustrated portions of CagX and CagY are the domains localized to the OMC.

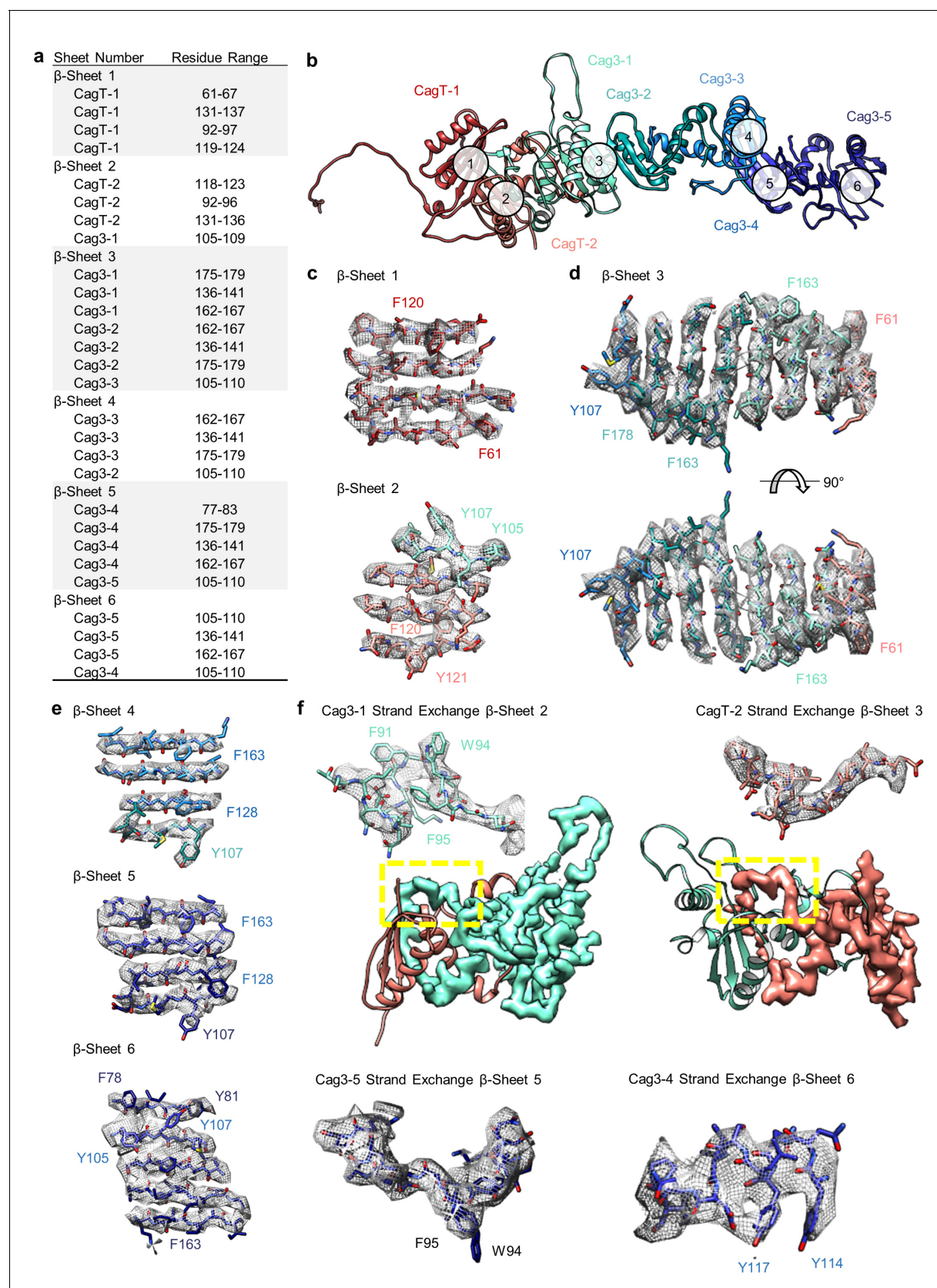


**Figure 6.** There are five copies of Cag3 within the OMC asymmetric unit. (a) Five different molecules of Cag3 are present within the asymmetric unit. The general architecture of Cag3 can be described as two domains, the proximal domain (residues 62–232) and the distal domain (residues 252–309). *Figure 6 continued on next page*



## Figure 6 continued

(b) All five of the Cag3 proteins share a heavily interwoven architecture in which  $\beta$ -sheets are formed between adjacent molecules within the asymmetric unit. (c) A topology diagram showing the general architecture of all copies of Cag3 within the asymmetric unit. (d) The interface of Cag3-2 (cyan) and Cag3-3 (light blue) is formed predominantly by two helices with contact mediated by Y114 and R128 of each molecule. (e) The interface of Cag3-3 (light blue) and Cag3-4 (blue) is dramatically different from the other Cag3 interfaces and includes an extensive hydrophobic interaction that is formed by two adjacent beta sheets. (f) Cag3-4 (blue) and Cag3-5 (navy blue) share a similar interface as Cag3-2 and Cag3-3, as shown above. (g) A view of the Cag T4SS is shown from the top-down (left) and indicates the position of a loop within the Cag3-1 proximal domain (amino acids 181–204, black box) that mediates contacts between asymmetric units. (h) The proximal domain of Cag3 contains a fold that is structurally similar to folds within CagT-1 (shown in red), CagT-2 (shown in salmon) and *X. citri* VirB7 (PDB 6GYB, shown in purple).

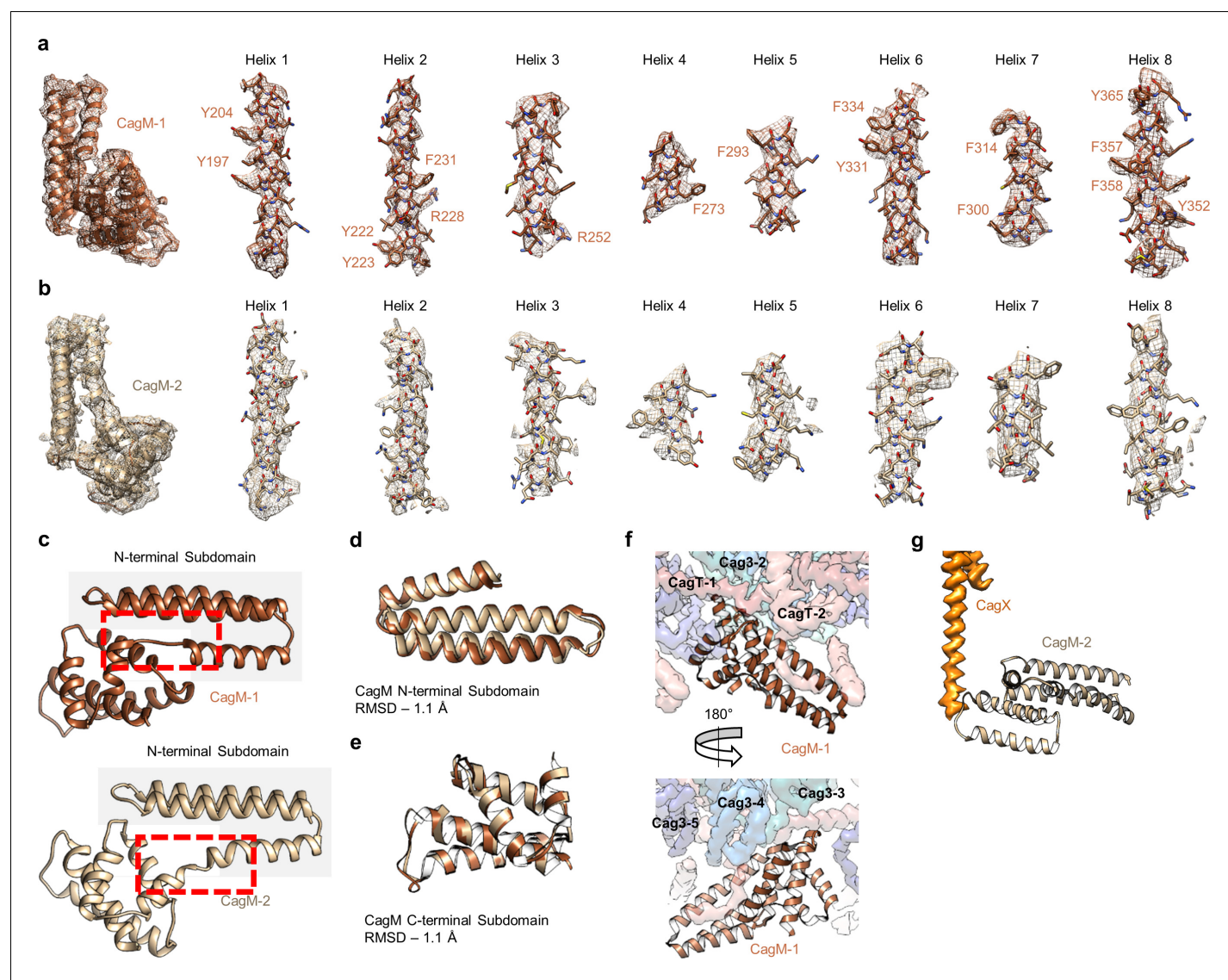


**Figure 6—figure supplement 1.** The repetitive β-sheet structure of CagT and Cag3. (a) The table lists residues that contribute to six β-sheets formed by copies of CagT and Cag3 within the OMC. The location of these six β-sheets is indicated in panel (b). (c–e) Representative density for all β-sheets

*Figure 6—figure supplement 1 continued on next page*

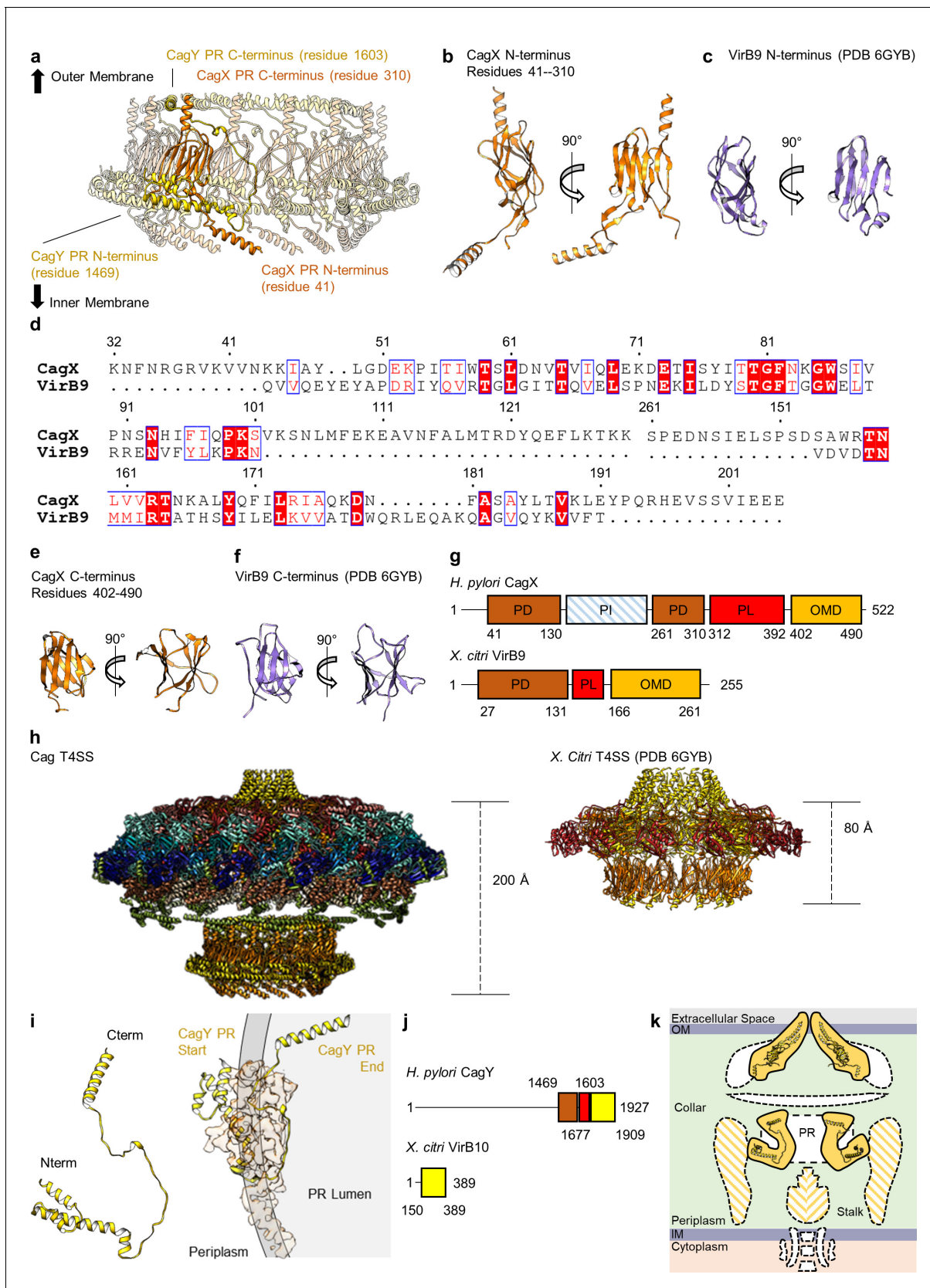
Figure 6—figure supplement 1 continued

with landmark residues noted for each. (f) Representative density for  $\beta$ -sheet strand exchange is shown for  $\beta$ -sheet 3 (top, left),  $\beta$ -sheet 4 (top, right),  $\beta$ -sheet 5 (bottom, left), and  $\beta$ -sheet 6 (bottom, right).



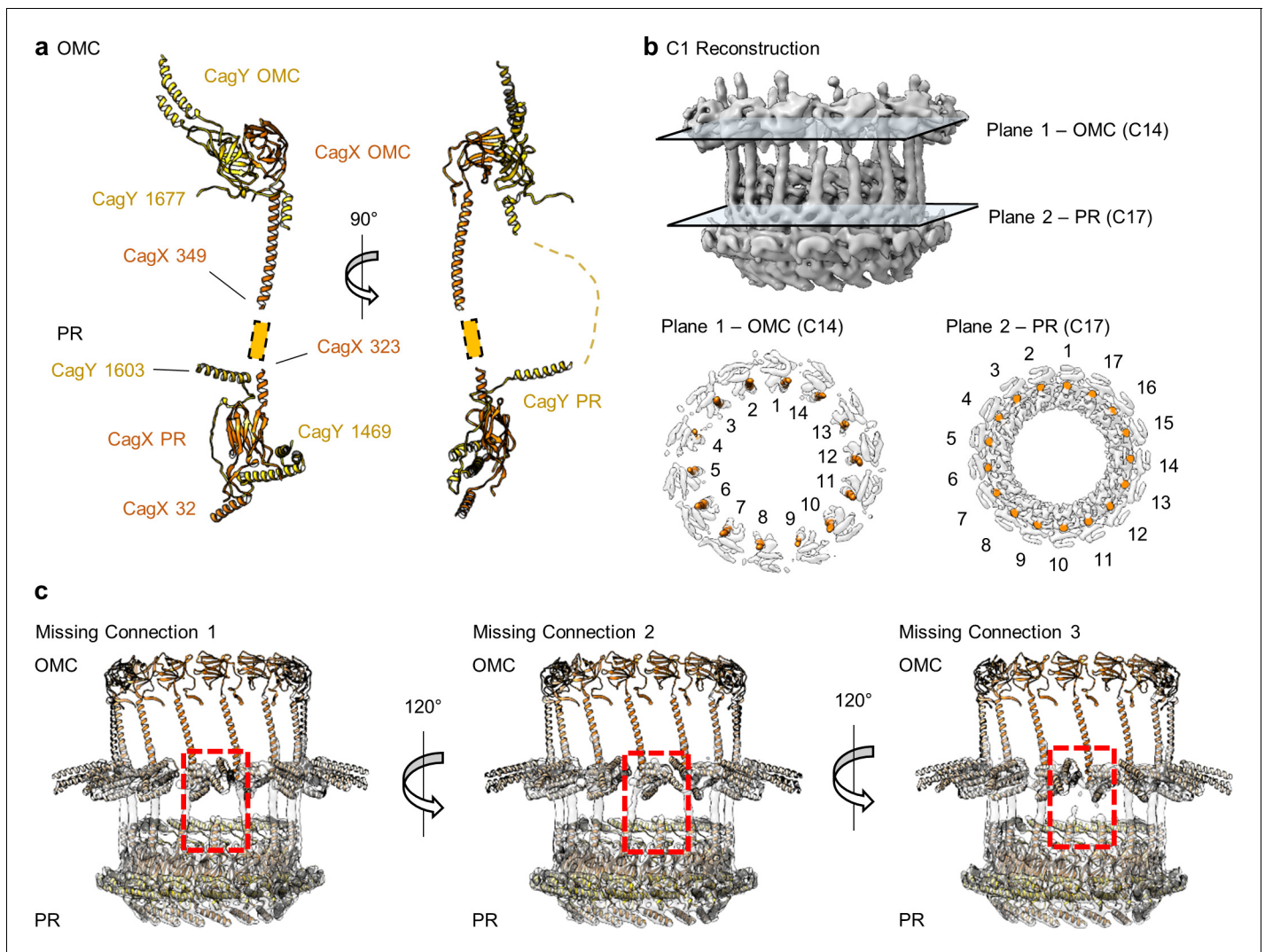
**Figure 7.** Localization of CagM within the I-layer of the OMC. (a) We modeled a single domain of CagM-1, consisting of 8 helices within the I-layer of the OMC. Representative density for all eight helices is shown with landmark residues indicated. (b) Although the local resolution within the I-layer is notably lower than that of the O-layer, several structural features within the I-layer density are consistent with the sequence of CagM (designated CagM-1). The correlation of all 8 helices of CagM-2 to the experimental map is shown. (c) The two subdomains within CagM are connected by a hinge (noted by the red dotted line). When the CagM-1 subdomains are aligned independently, we note RMSDs of 1.1 Å for both the N-terminal subdomain (d) and the C-terminal domain (e). The differences in sub-domain orientation are likely the result of differences in interacting partners of CagM-1 (f) and CagM-2 (g) within the OMC.





## Figure 8 continued

around CagX, starting from the periplasm and winding into the lumen of the PR. (b) The N-terminal domain of CagX (residues 41–310) is similar to that of VirB9 (c) from *X. citri* in both structure and sequence (d). The C-terminal domain of CagX (e) is similar to that of the C-terminal domain of VirB9 from *X. citri* (f). (g) The periplasmic and outer membrane domains (PD and OMD, respectively) are similar in structurally characterized VirB9 homologs, though they are separated by a periplasmic linker (PL) that is variable in length. CagX contains an additional insertion (residues 102–153, periplasmic insertion or PI) that is unique when compared to other homologs. The structure corresponding to the PI was not observed within any of our cryo-EM reconstructions. (h) The spacing of the two CagX/VirB9 domains varies depending on the organism and appears to be highly dependent on the length of the periplasmic linker (CagX and VirB9 are shown in orange). (i) The segment of CagY within the PR (residues 1469–1603) adopts a highly elongated fold that consists of four  $\alpha$ -helices (left). The periplasmic portion of CagY (as shown on the left) starts within the periplasm and wraps around the globular domain of CagX (shown in orange), eventually ending in the lumen of the PR (gray, right). (j) The periplasmic segment of CagY, which we call the periplasmic domain (PD, brown), is unique to CagY and is not present in other VirB10 homologs such as VirB10 from *X. citri* (yellow indicates a VirB10 like domain, and red represents the unstructured linker). (k) The N-terminus of CagY was not observed within any of these cryo-EM reconstructions. The N-terminal portion of CagY in the model that was constructed (about residue 1469) is positioned so that it might continue outward into the periplasmic space, possibly contributing to the structural feature known as the collar, as well as the stalk.



**Figure 9.** CagX and CagY span the symmetry mismatch between the OMC and PR. (a) Models of CagX and CagY within the OMC and PR asymmetric units. (b) In the asymmetric reconstruction from a focused refinement of the interface between the OMC and PR of the WT map, we note the presence of helical density predicted to be a portion of CagX (top). We have modeled 14 copies of CagX into density within the OMC (bottom, left) and 17 copies of CagX into the PR (bottom, right). (c) Though we observe clear connections between 14 subunits of the OMC and 14 subunits of the PR, three copies of CagX in the PR do not show any obvious connection to the OMC.