Supplementary Information

Supplementary Figure Legends

Supplementary Fig. S1. Sequence and subdomain organization of MUC2 N-terminus and MUC2D3 proteins. The subdomains present in MUC2 N-terminus and MUC2D3 protein variants were identified by comparison with the recently established VWF subdomain assignment [1]. The MUC2 N-terminus was composed of several D assemblies (D1 (orange), D2 (yellow), D´ (blue) and D3 (blue)). The D1, D2 and D3 assemblies were each composed of four subdomains: VWD, C8, TIL and E, respectively. The D´ assembly was composed of TIL´ and E´. The different subdomains are colour-coded and displayed above the sequence alignment. The cysteine residues are highlighted in yellow colour within the sequence. The subdomain borders are indicated by vertical bars. The MUC2D3CysD1-MG construct included CysD1 (red), Myc tag (black) and GFP (light green) whereas MUC2D3 did not include them and MUC2D3CysD1-M was lacking GFP. The VWF N-terminal sequence was human (UniProt/Swiss-Prot P04275) and corresponded to amino acid residues 23-1270 excluding the signal sequence. The MUC2 N-terminal sequence was also human (UniProt/Swiss-Prot Q02817) and corresponded to amino acid residues 21-1395 excluding the signal sequence. The sequences of the three MUC2D3 protein constructs were of human MUC2 origin. The three MUC2D3 protein constructs are lacking the D1 assembly but included TIL-2 (half a domain) and E-2 from the D2 assembly. The signal sequences for the different constructs are not shown.

Supplementary Fig. 2. Electron micrographs of MUC2D3. Parts from representative electron micrograph of negatively stained MUC2D3CysD1-MG (A), MUC2D3CysD1(B), MUC2D3 (C) used for image processing. Examples of donut-like particles are indicated by arrows. 2048

x 2048 pixels were cut out from the original micrographs and the pixel size was adjusted to 2.7 Å in all micrographs to render them the same magnification. From the micrographs, a total of 8635 MUC2D3CysD1-MG, 7248 MUC2D3CysD1-M and 7321 MUC2D3 particles were selected for processing. Particles have been boxed out at the same sizes as used in the image processing after the CTF correction, i.e. 120, 112 and 100 pixels respectively. (D) Representative electron micrograph showing MUC2D3CysD1-MG labeled with Gold-labelled PurαMUC2N3 antibody (example indicated by arrow) in order to specify that the particles consisted of the D´D3 domain. Scale bar 50 nm. (E) Representative particles from electron micrographs of a MUC2D3 where the half TIL-2 domain and the E-2 domain have been removed (H-MUC2D3DTIL2E2-M). Images were recorded in the same way as for the other larger particles. Scale bar 10 nm.

Supplementary Fig. S3. 2D classification of the raw images of MUC2D3CysD1-MG (A), MUC2D3CysD1-M (B) and MUC2D3 (C) using the *refine2d.py* program in EMAN1. Examples of densely populated classes showing approximate 2-fold and 3-fold symmetry are marked with red and yellow boxes.

Supplementary Fig. S4. 3D projection maps of the three different constructs MUC2D3CysD1-MG (A) MUC2D3CysD1-M (B) and MUC2D3 (C). The correspondence to the 2D classification above is significant.

Supplementary Fig. S5. Euler angle distribution for the classified particles used for 3D reconstructions based on class means of MUC2D3CysD1-MG (A), MUC2D3CysD1-M (B) and MUC2D3 (C). All Euler angles were represented in the asymmetric triangle in the three

different constructs. Each dot represents one class of particles in a certain orientation. Euler angles were calculated based on projection matching between each boxed particle image and a series of projected images calculated from the 3D refined volume in 6-degree intervals. The tilt angle is 0°-90° (from top to bottom). The orientation of the tilt axis in the x-y plane is 0°-60°, measured from the y axis (from left to right).

Supplementary Fig. S6. Plots of the FSC curves were used to determine the resolution of the 3D reconstructions of MUC2D3CysD1-MG (A), MUC2D3CysD1-M (B) and MUC2D3 (C) using the FSC 0.5 threshold criterion. The obtained resolution was 18.5Å for MUC2D3CysD1-MG (A); 17.5Å for MUC2D3CysD1-M (B) and MUC2D3 (C). According to the convergence, the errors at the above resolutions were 10-12%, for the two last convergence curves shown as thin dotted lines and solid lines (comparison between second last and third last, and between the last and second last iteration respectively).

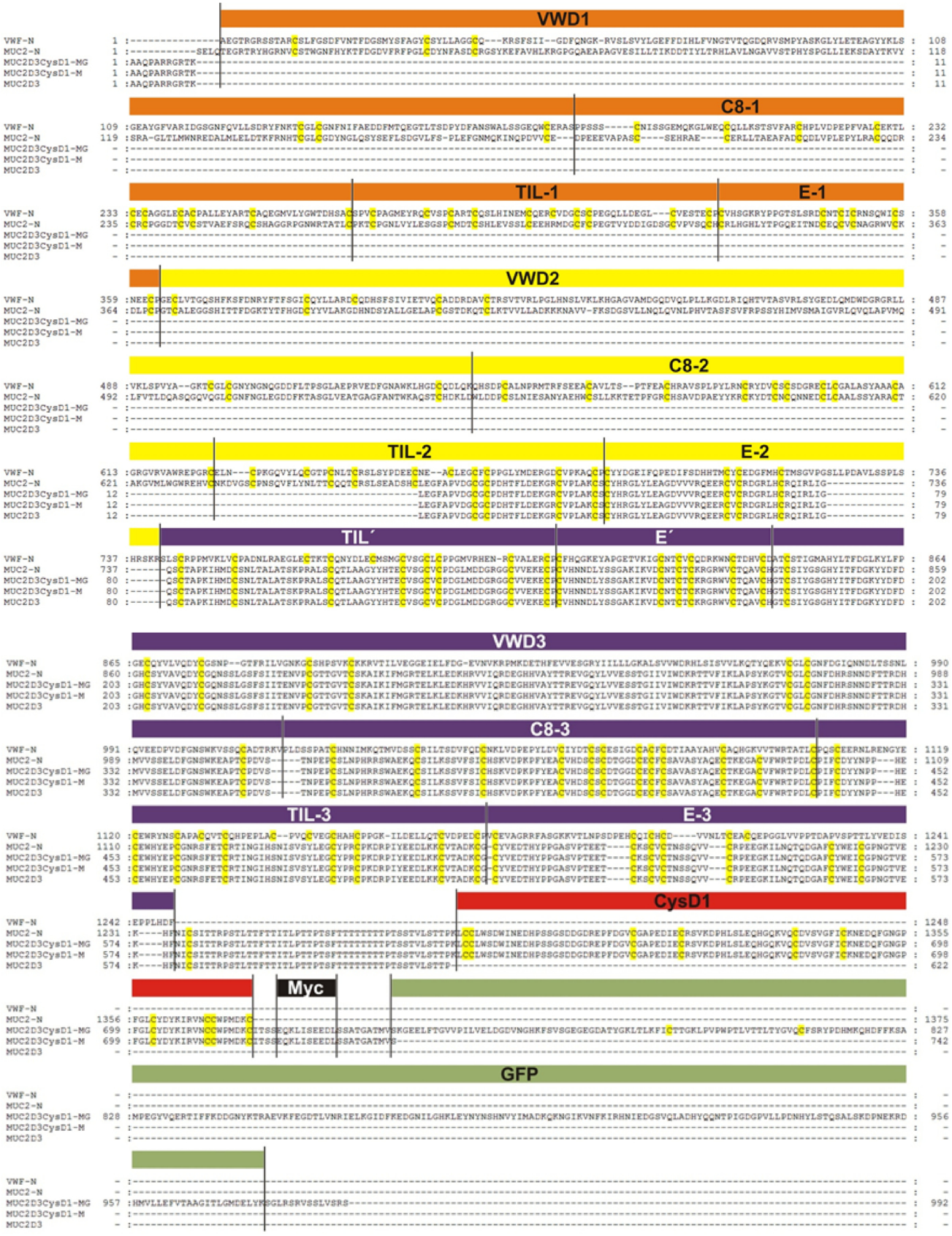
Supplementary Fig. S7. 2D crystals of MUC2D3CysD1-MG. (A) Electron micrograph from frozen hydrated specimens. Scale bar 500 Å. (B) Projection map from a 2D-crystal of MUC2D3CysD1-MG. Four unit cells, a=b=118 Å, γ=120°, as calculated in p1 are shown. The overlay is a representation of the 3D map of this complex obtained from the single particle analysis as seen along the two-fold axis and rotated to fit the orientation of the protein in the crystals.

Supplementary Fig. S8. Organization of the mucin and D1D2 linker regions relatively to the symmetry operations of the D'D3 hexamers, MUC2D3CysD1-MG (mesh representation) and MUC2D3CysD1-M (surface rendered with one monomer segmented as in Fig. 5). (A) Side

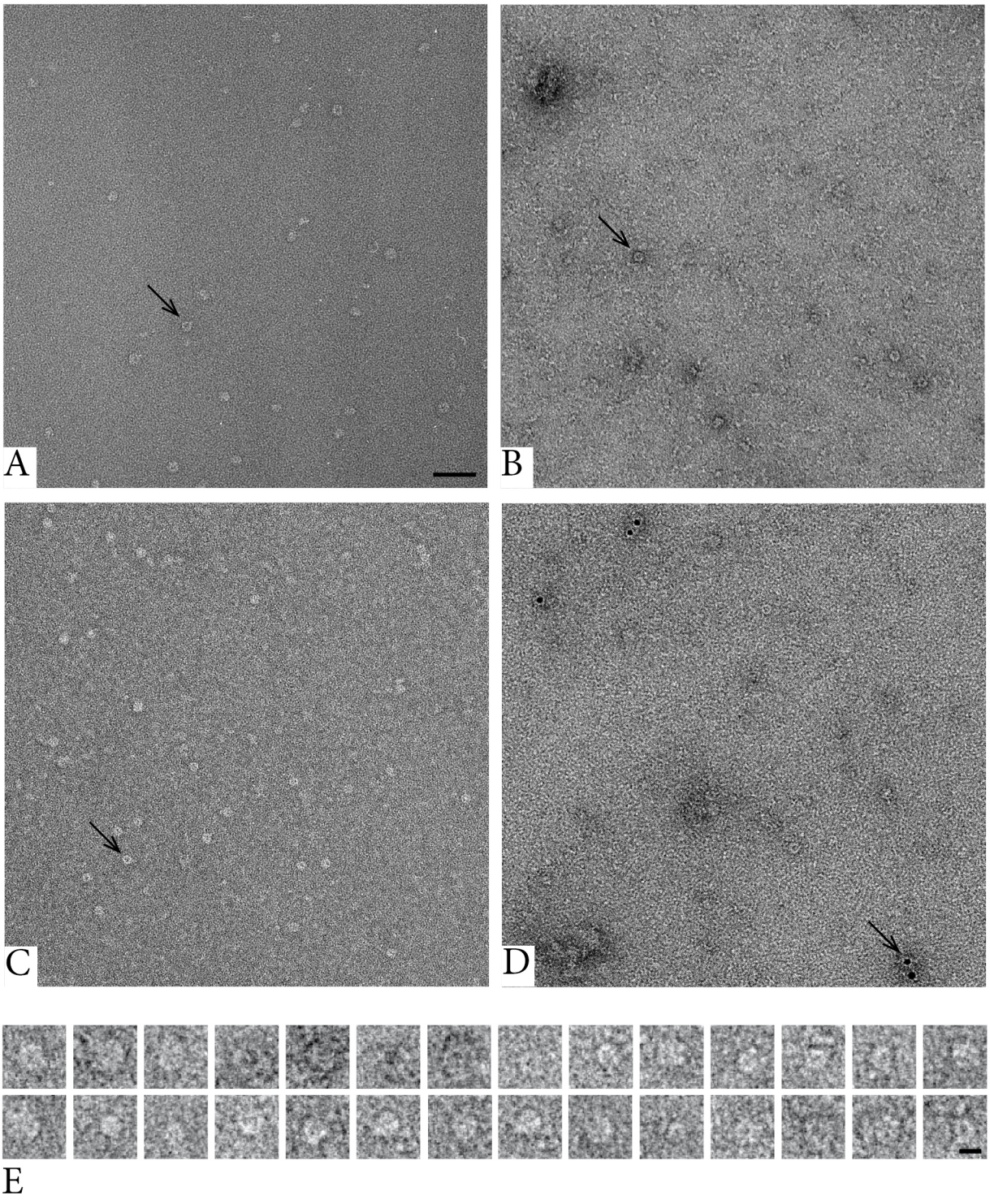
view along the two-fold symmetry axis marked by the ⊗ symbol. Two antiparallel mucin domains have been placed with preservation of the symmetry and are shown as arrows with open arrow heads. The N-terminal ends of mucin PTS domains are close to the CysD1 and Myc tails of the hexamers (orange in the segmented monomer). (B) At the opposite face of the particle, i.e. the side view also along the two-fold axis but in the opposite direction as compared to A, most of the mass of GFP is located (mesh representation). (C) Top view along the three-fold axis (marked with a triangle) with the two-fold symmetry axes shown as small arrows running through the center of the particles. The large arrows with filled arrow heads represent schematically three pairs of D1D2 domains in opposite directions. The C- and N-termini of D1D2 (filled arrow heads) connect identical hexamers in polygons as depicted in Fig. 6. The starting points of the D1D2 domains close to the three-fold axis correspond to their C-termini connected to the TIL', E', VWD3-domains. Only D1D2 pairs attached to one of the leaflets of the hexamer are shown. On the opposite leaflet related by the two-fold axis similar pairs of D1D2 will be attached.

Reference

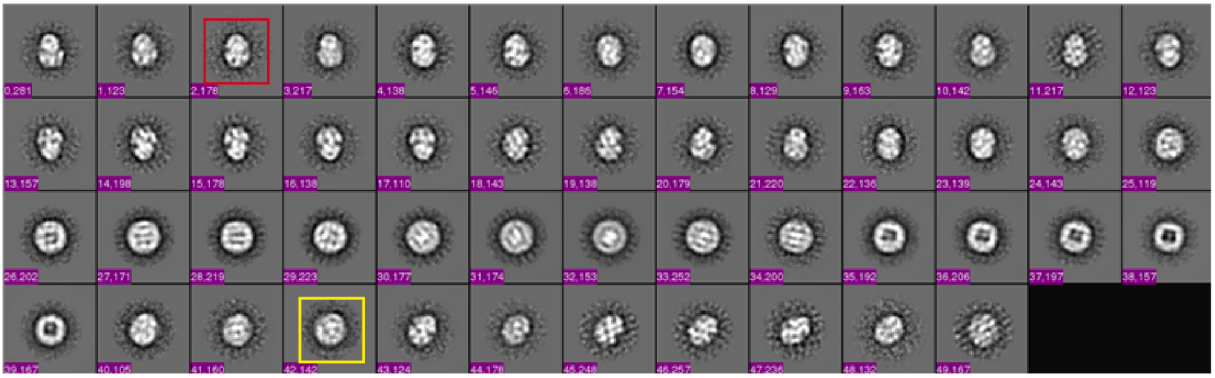
1. Zhou Y .F., Eng, E. T., Zhu, J., Lu, C., Walz, T. & Springer, T. A. (2012) Sequence and structure relationships within von Willebrand factor. *Blood* 120(2):449-458.



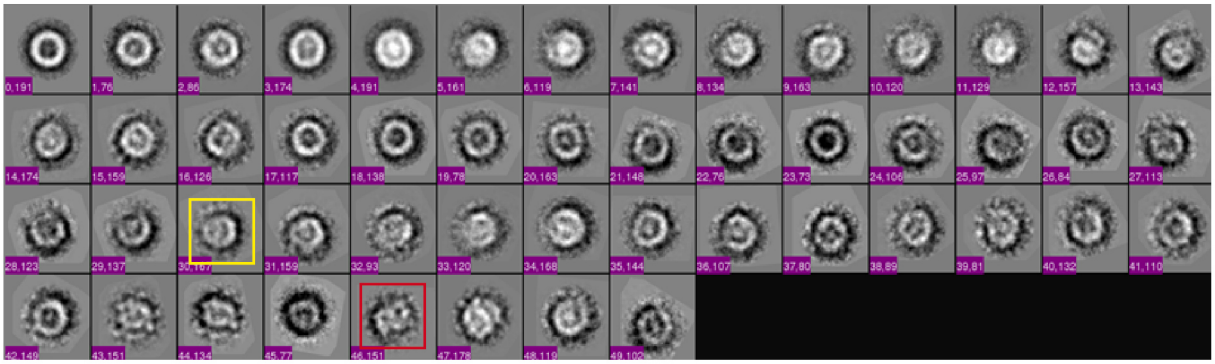
Supplementary Fig. S1



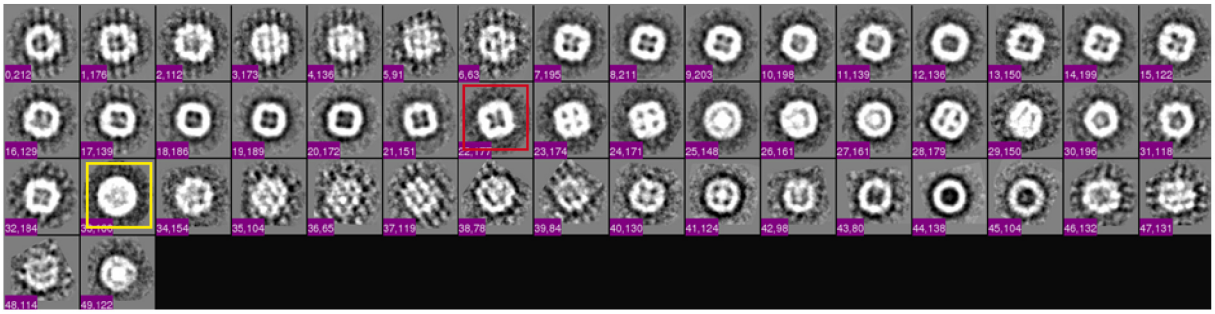
Supplementary Fig. S2



A

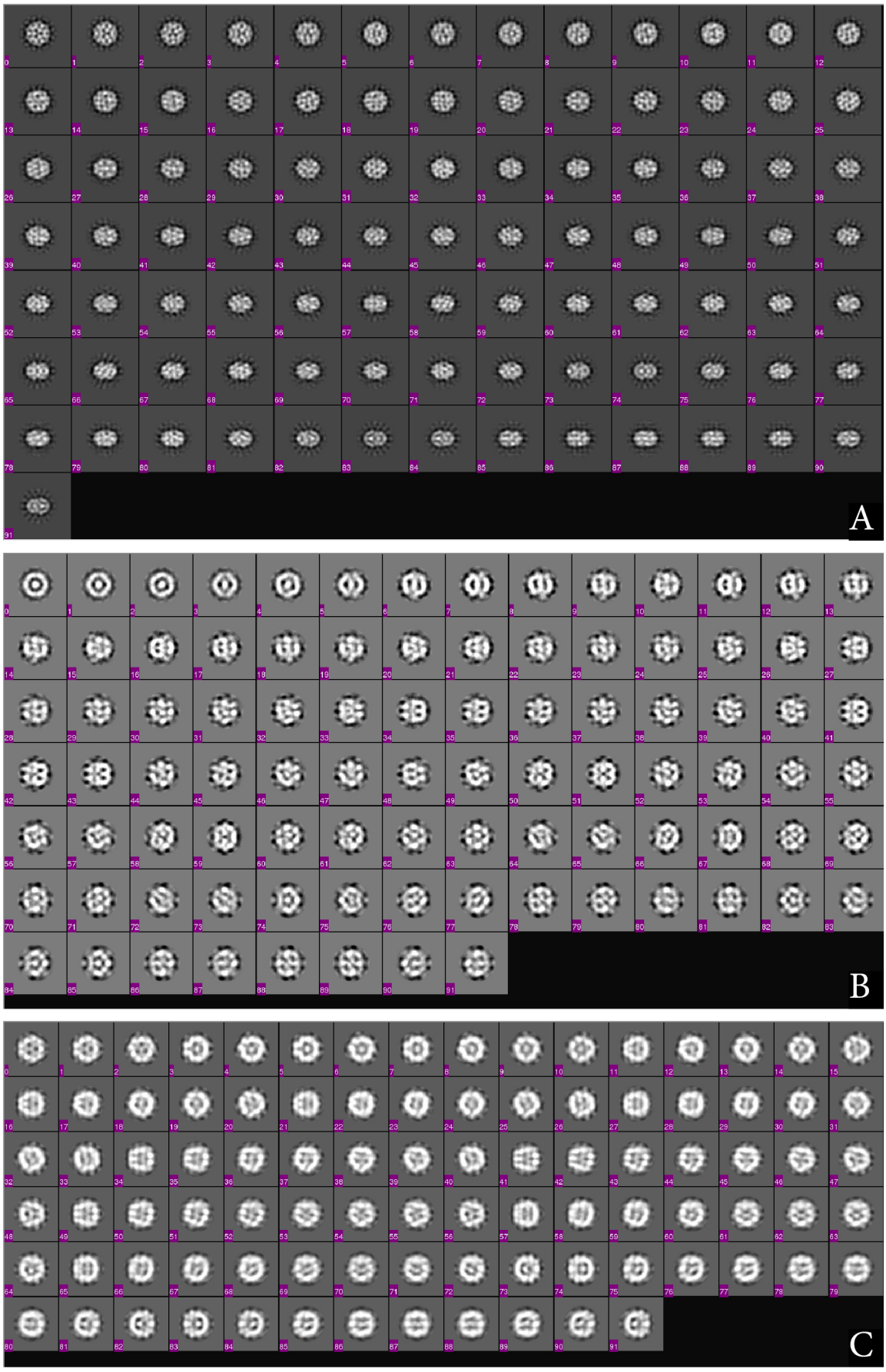


B

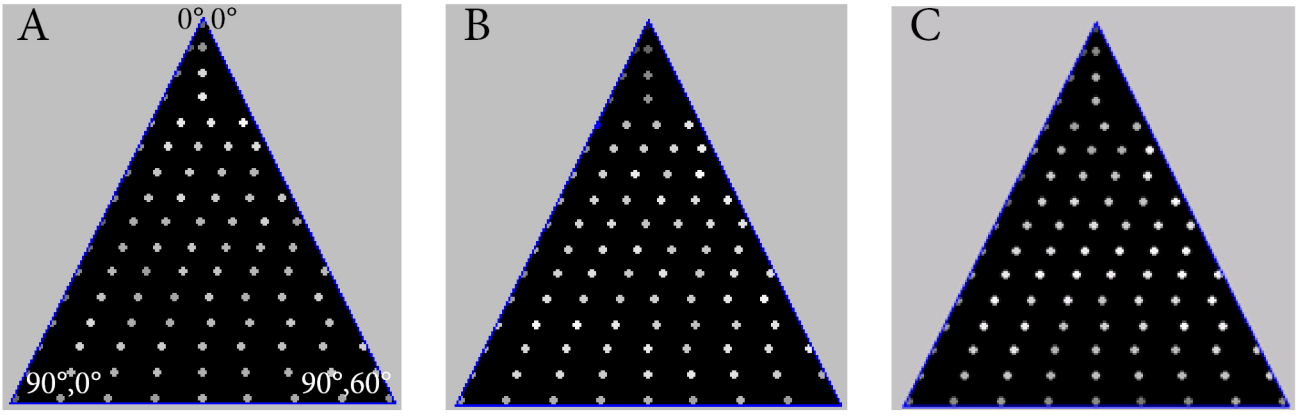


C

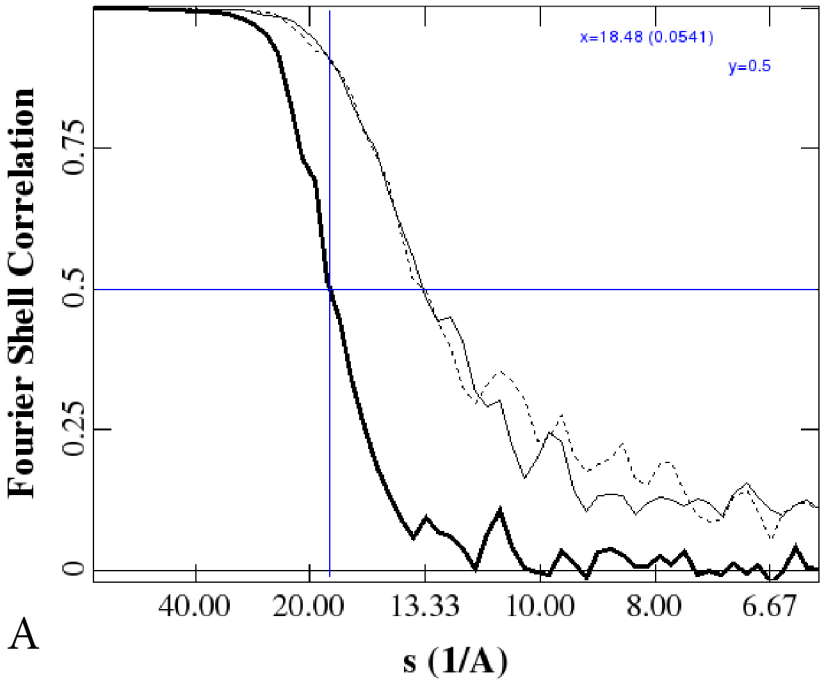
Supplementary Fig. S3



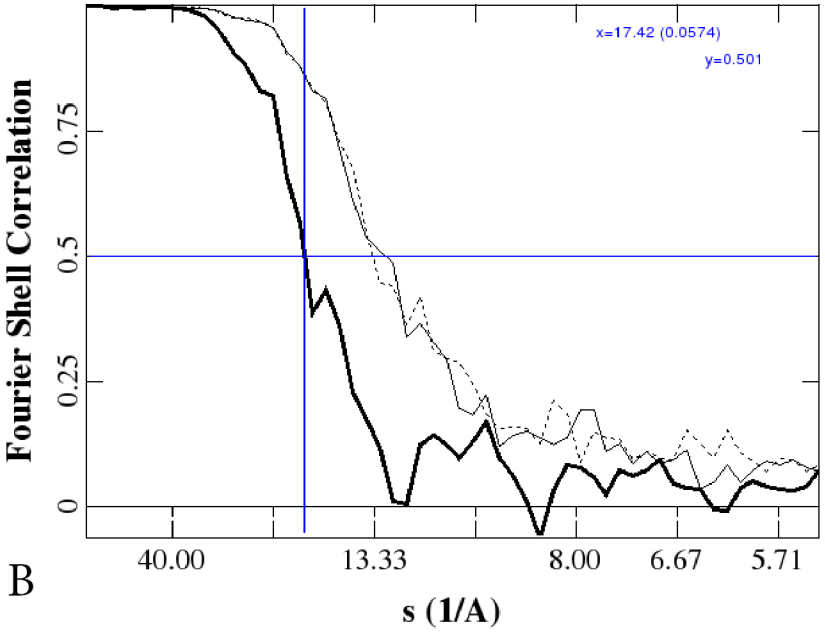
Supplementary Fig. S4



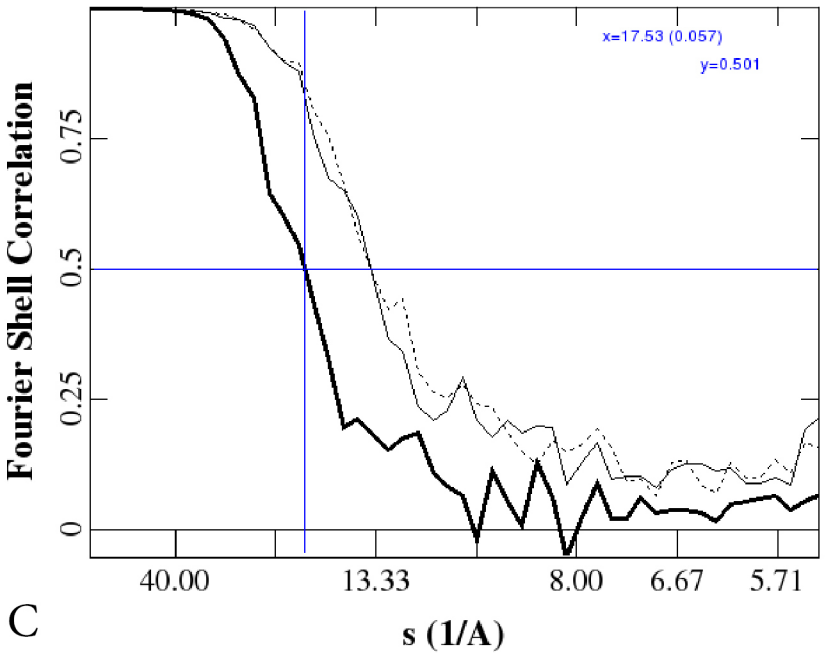
Supplementary Fig. S5



A

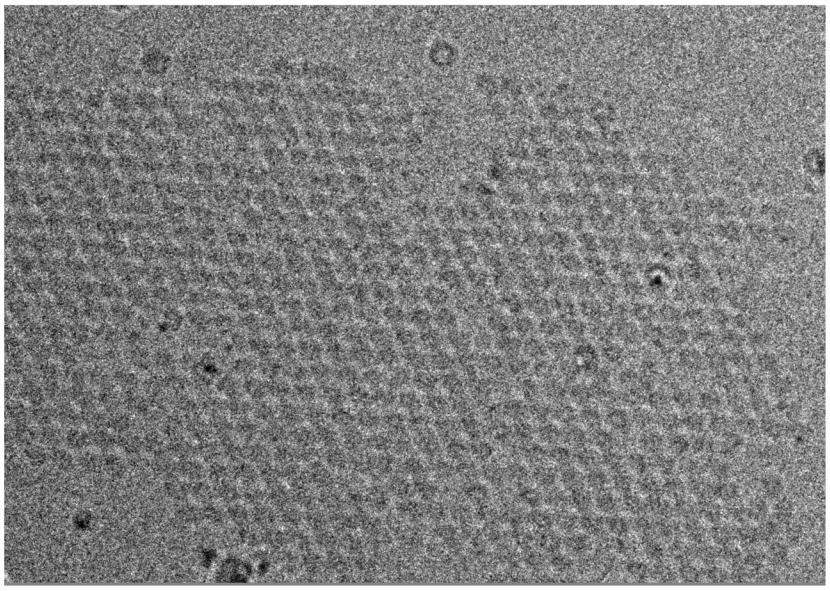


B

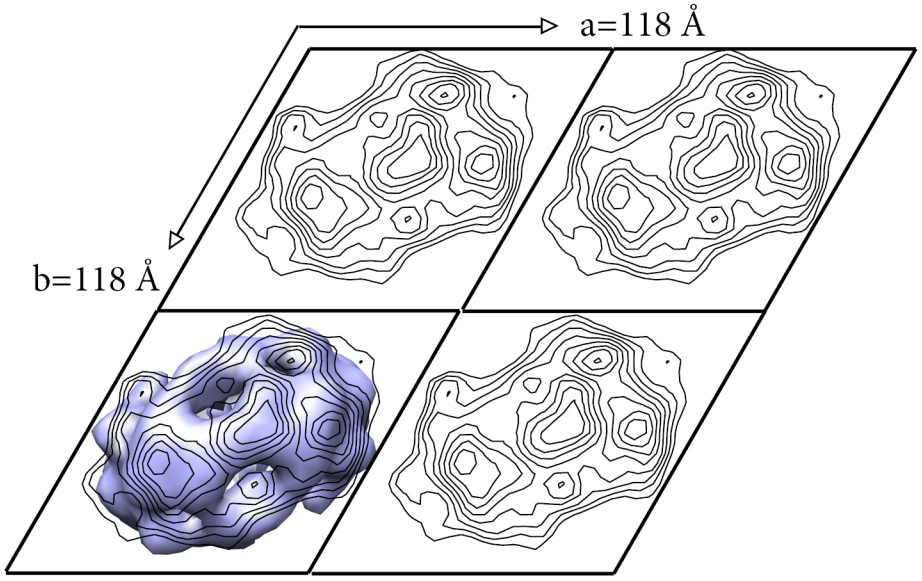


C

Supplementary Fig. S6

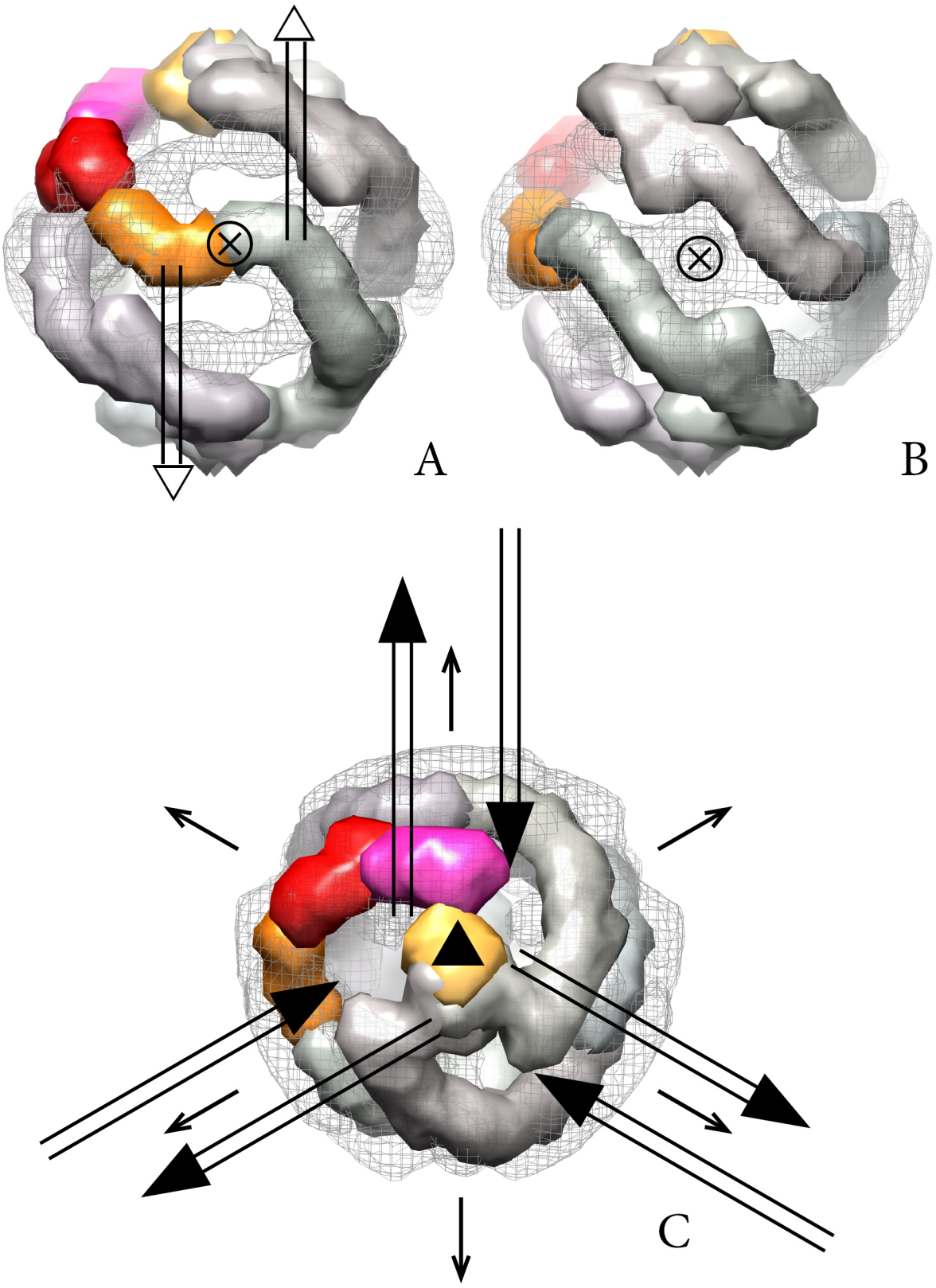


A



B

Supplementary Fig. S7



Supplementary Fig. S8