SI APPENDIX

Cryo-EM shows structural basis of pan-peptidase inhibition by human α2-macroglobulin

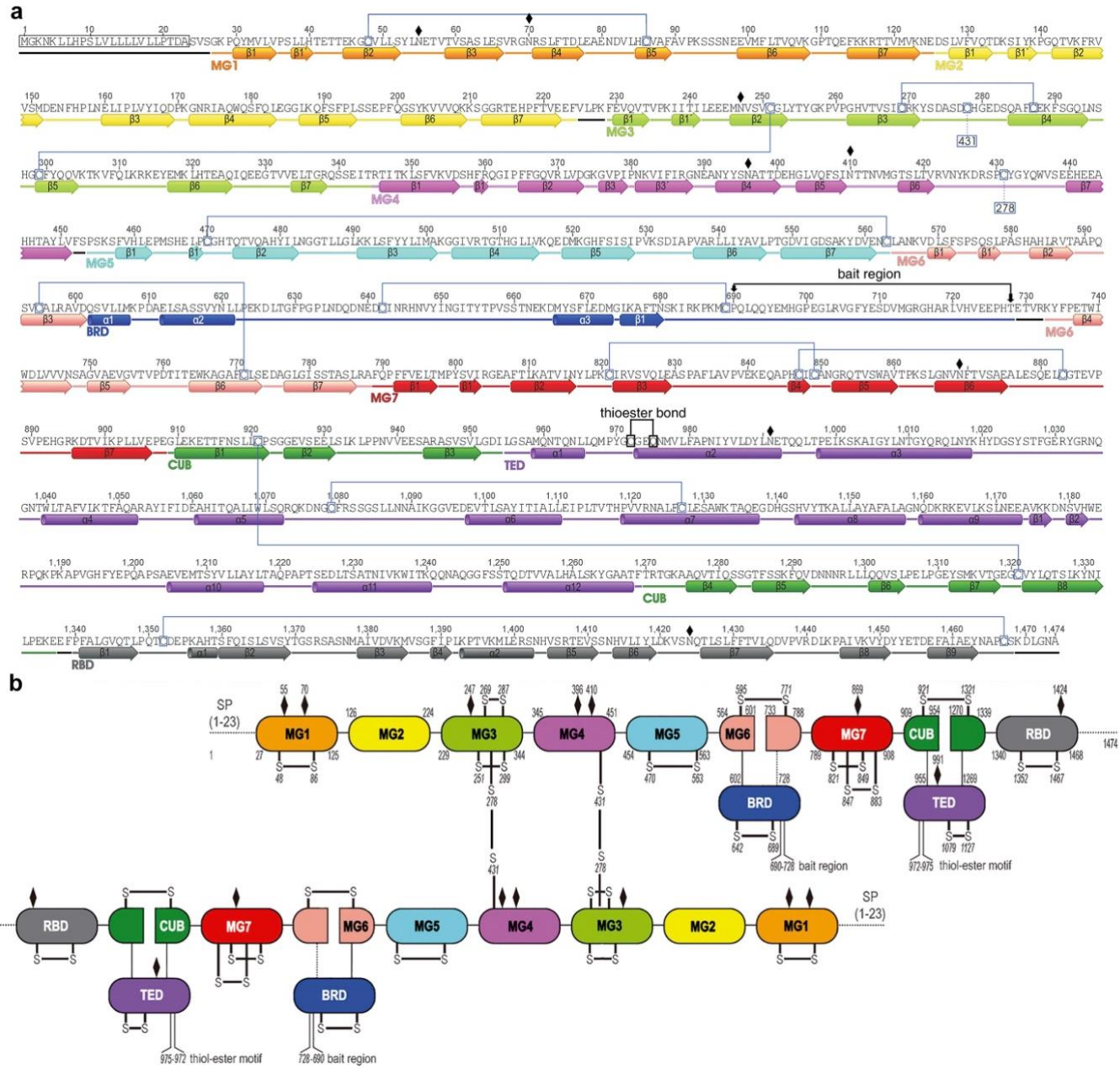
Daniel Luque[#], Theodoros Goulas[#], Carlos P. Mata[#], Soraia R. Mendes,

F. Xavier Gomis-Rüth[*] and José R. Castón[*]

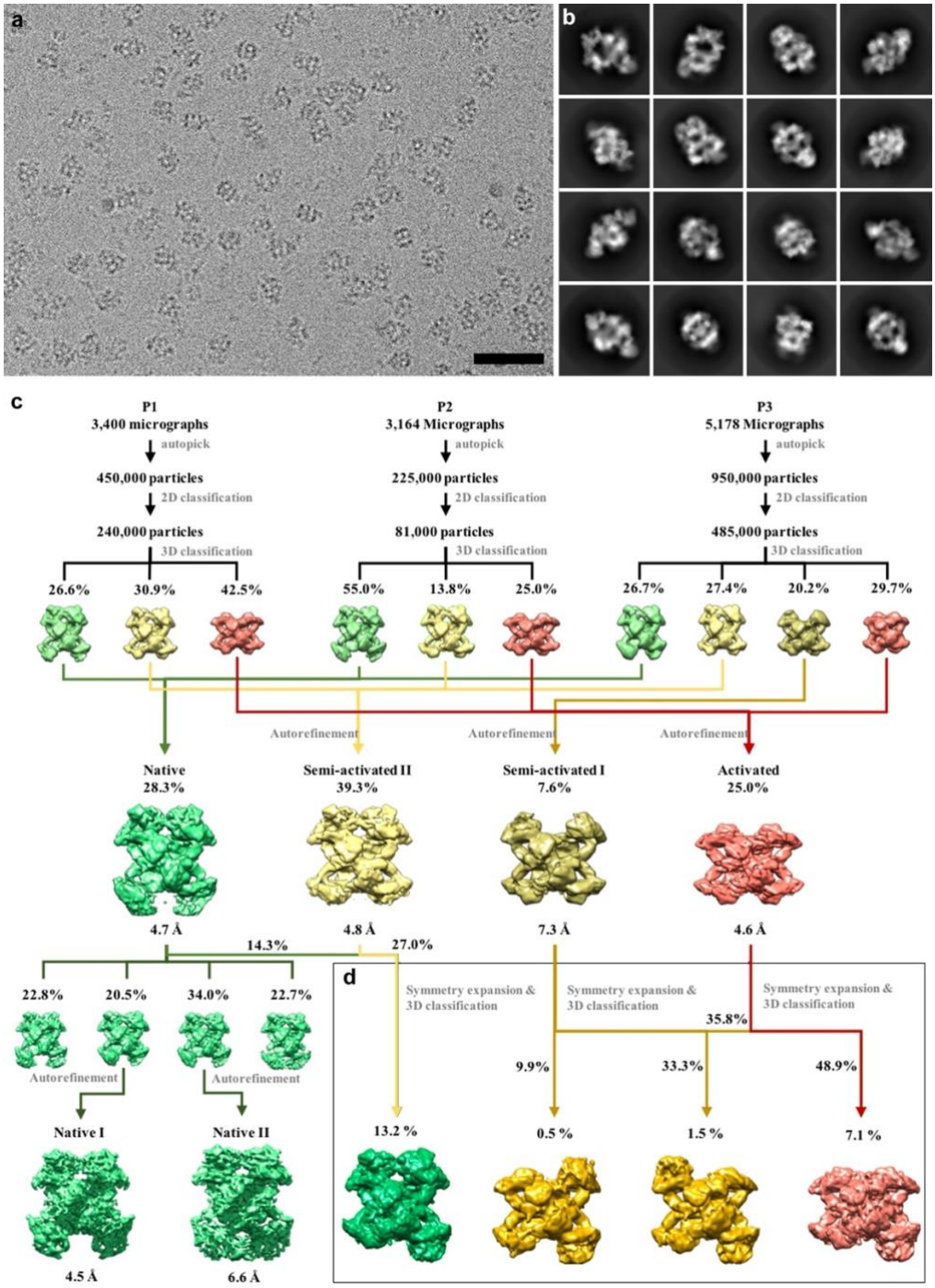
Supplementary Figures: 9

Supplementary Tables: 4

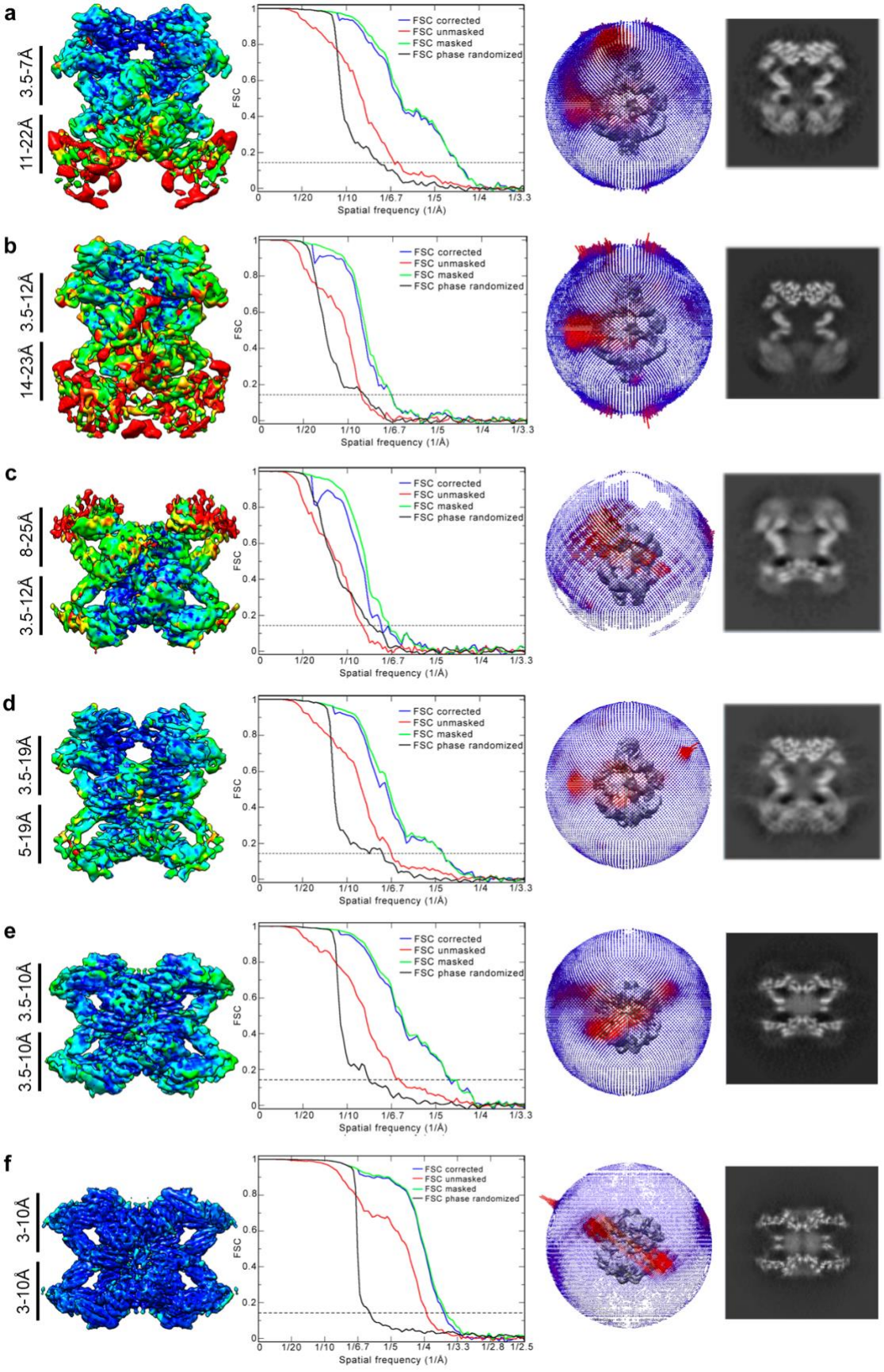
Legends to Supplementary Movies 1 and 2 (uploaded separately)

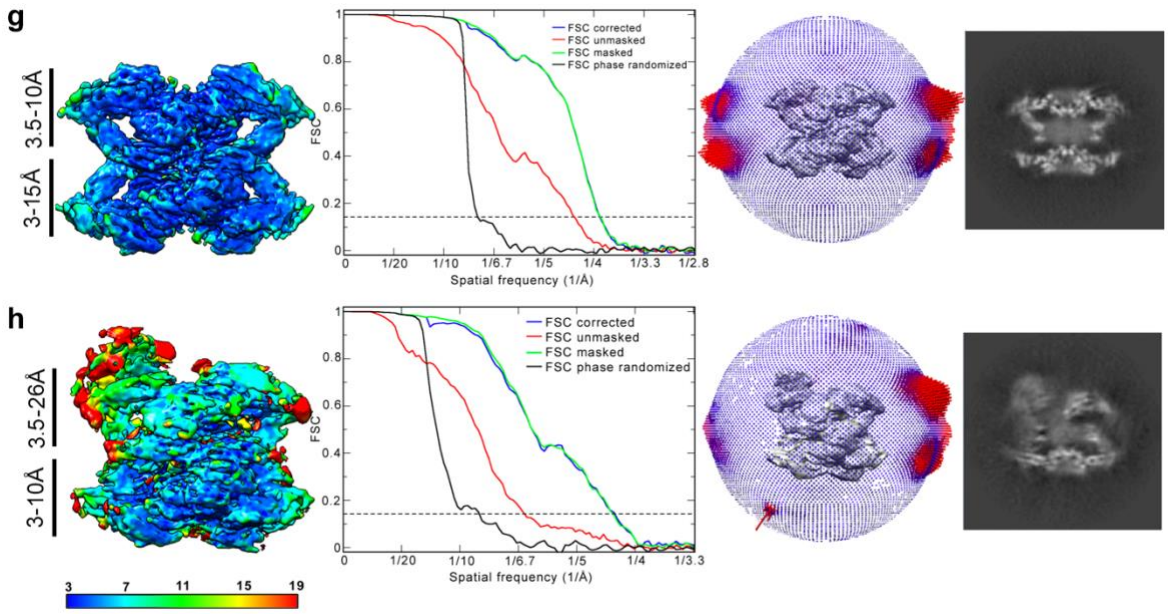


Supplementary Figure S1: Sequence, secondary structure elements, and domain organization of hα_2M. (**a**) Sequence and secondary structure elements of the 1474-residue hα_2M (UniProt code P01023) spanning a 23-residue signal peptide and the 1451-residue secreted protein. In the expanded native conformation, the first and last residues assigned were Ser26 and Ser1468, respectively; the compact activated conformation spans Lys28-Glu1335. The α-helices and β-strands are represented as cylinders and arrows, respectively, colored distinctly for the eleven domains. The bait region and the thioester bond are highlighted. Intra- and intermolecular disulfide bonds are designated with square boxes linked with a solid or dotted line (blue), respectively. N-glycosylation sites are indicated by rhombuses. MG, macroglobulin-like domains; BRD, bait region domain; CUB, domain found in C1r/C1s, urchin embryonic growth factor, and bone morphogenetic protein 1; TED, thioester domain; RBD, receptor binding domain. (**b**) Diagram of the domain organization of the disulfide-linked homodimer. The protomers are bound through two disulfide bonds. Symbols and colors as in (**a**).

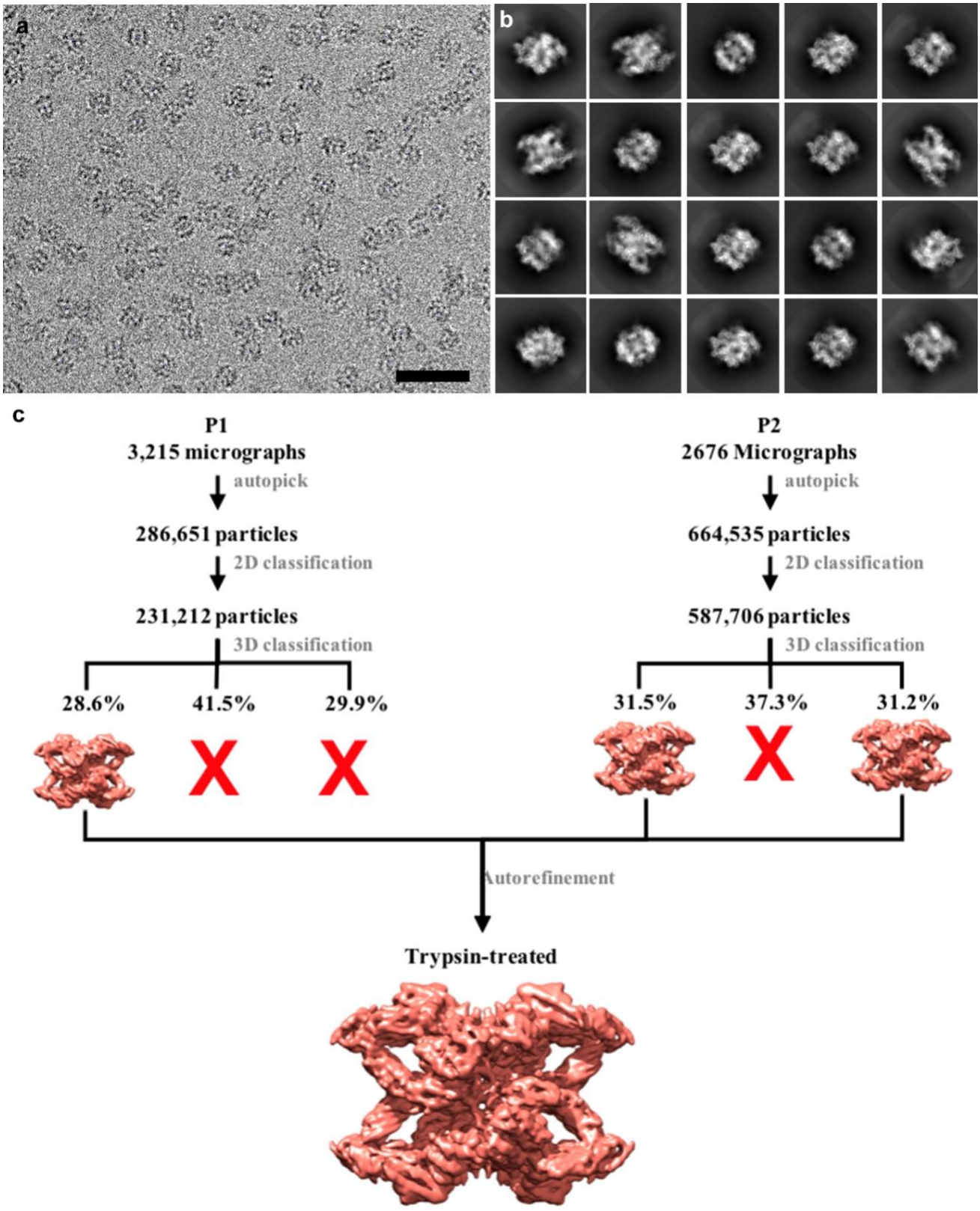


Supplementary Figure S2: Cryo-EM data processing of (hα$_2$M)$_4$ corresponding to the native untreated fraction. (**a**) Representative cryo-EM image of (hα$_2$M)$_4$ complexes (bar, 500 Å). (**b**) 2D class averages of (hα$_2$M)$_4$. (**c**) Data processing workflow and structure determination of the five functional states of (hα$_2$M)$_4$ using three serum preparations (P1-P3). After 3D classification, three (green, yellow, and red for preparations P1 and P2) and four (green, yellow, gold, and red for P3) distinct conformational states were identified (percentages indicated relative to total 3D selected particles of each preparation). Particles of each of the four states were combined into four separate datasets and further refined. The global resolution of each of the resulting 3D reconstructions was 4.7, 4.8, 7.3 and 4.6 Å, respectively, and the percentages relative to the sum of particles in the four classes was 28, 39, 8, and 25%, respectively. The dataset corresponding to truly native (hα$_2$M)$_4$ (green) was further classified and two conformational states were refined, yielding resolutions of 4.5 and 6.6 Å for native I and II states, respectively. (**d**) The C2 symmetry of the native, semi-activated I plus II and activated states was expanded and the particles from each state were subjected to additional 3D classification without alignment. Classes representing equivalent conformational states (percentages relative to the parental state are indicated) were pooled, resulting in 3D reconstructions of four new intermediate transient states.

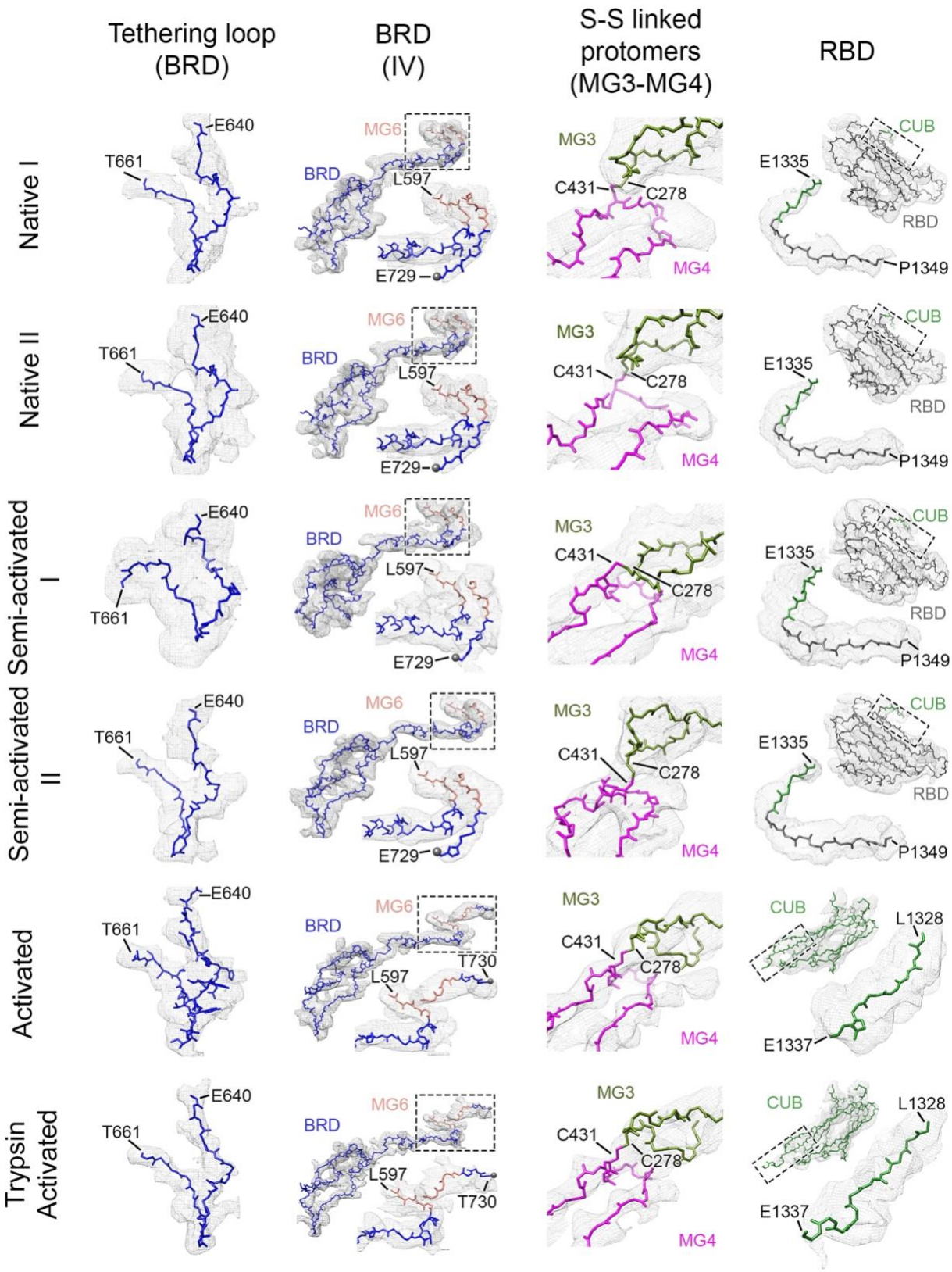

a
3.5-7Å
11-22Å
FSC corrected
FSC unmasked
FSC masked
FSC phase randomized
FSC
Spatial frequency (1/Å)
b
3.5-12Å
14-23Å
c
8-25Å
3.5-12Å
d
3.5-19Å
5-19Å
e
3.5-10Å
3.5-10Å
f
3-10Å
3-10Å



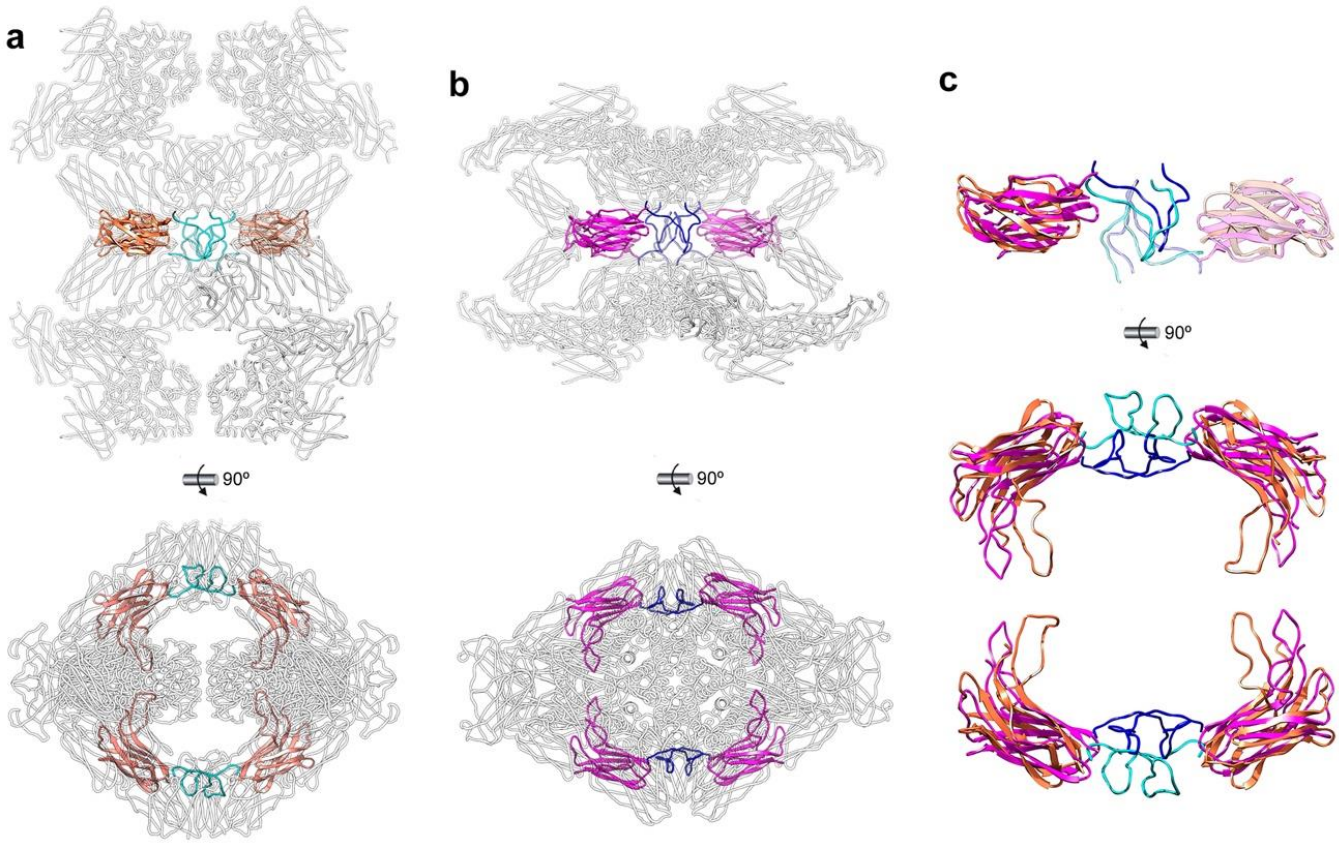
Supplementary Figure S3: Global and local resolution of cryo-EM maps. Local resolution assessment and Fourier shell correlation (FSC) curves for the eight maps calculated in this study: (**a**) native I, (**b**) native II, (**c**) semi-activated I, (**d**) semi-activated II, (**e**) fully activated state, (**f**) trypsin-activated state, (**g**) plasmin-activated I state, and (**h**) plasmin-activated II state. The bar indicates the resolution in Å (bottom). The angular distributions of particles used to compute the final three-dimensional maps are shown (top, right), as well as the longitudinal central section of each map (protein is white, bottom, right).



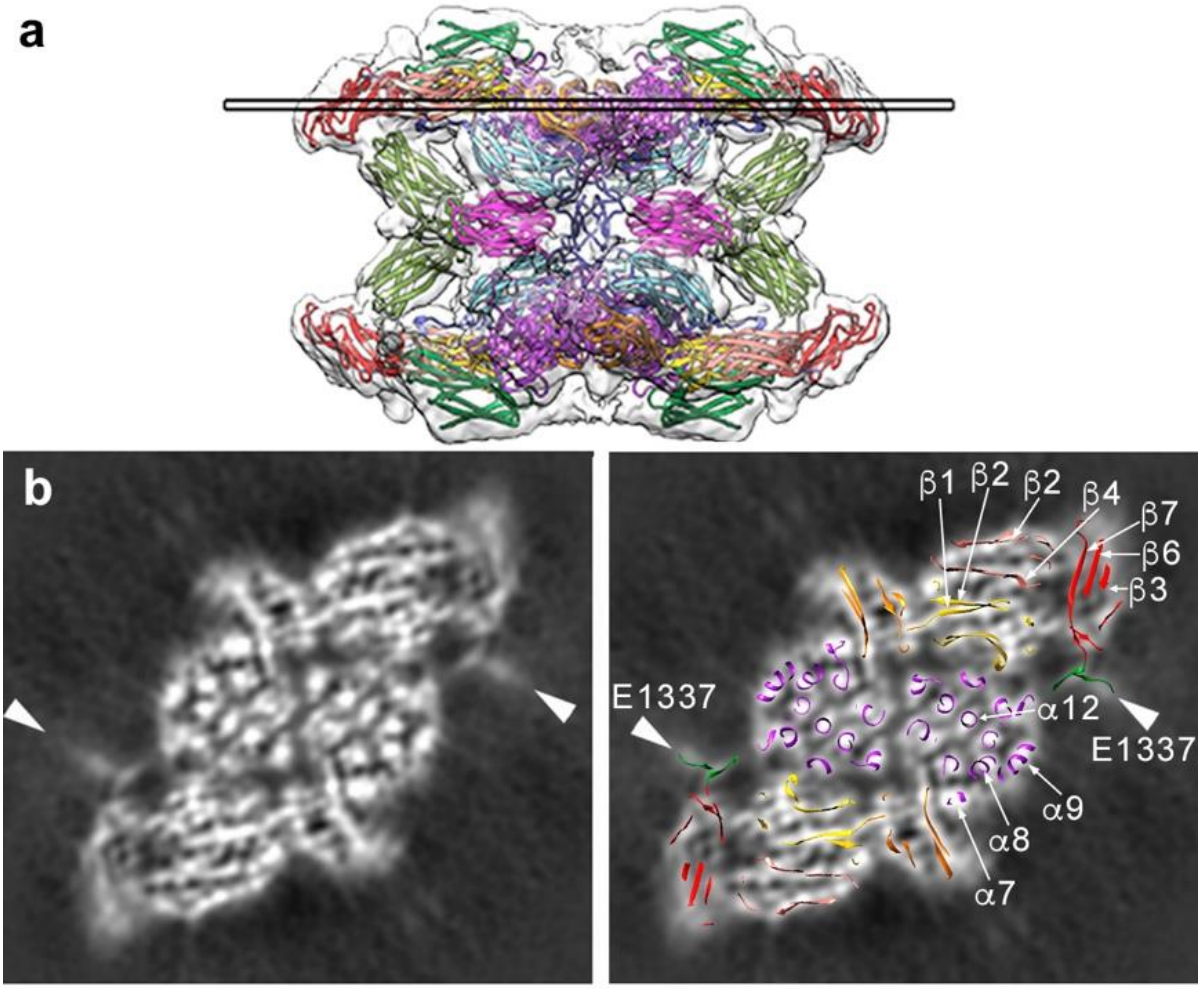
Supplementary Figure S4: Cryo-EM data processing of trypsin-treated (hα$_2$M)$_4$. (**a**) Cryo-EM image of (hα$_2$M)$_4$ complexes after trypsin treatment (bar, 500 Å). (**b**) 2D class averages of trypsin-treated (hα$_2$M)$_4$ particles. (**c**) Data processing workflow and structure determination of the trypsin-treated, activated (hα$_2$M)$_4$ state using two preparations (P1 and P2). After 2D and 3D classifications, a homogenous population of particles was selected for each preparation. Particles for each of these pools were combined and autorefined. Percentages relative to total 2D selected particles of each preparation are indicated. See Methods for further details on cryo-EM data processing.



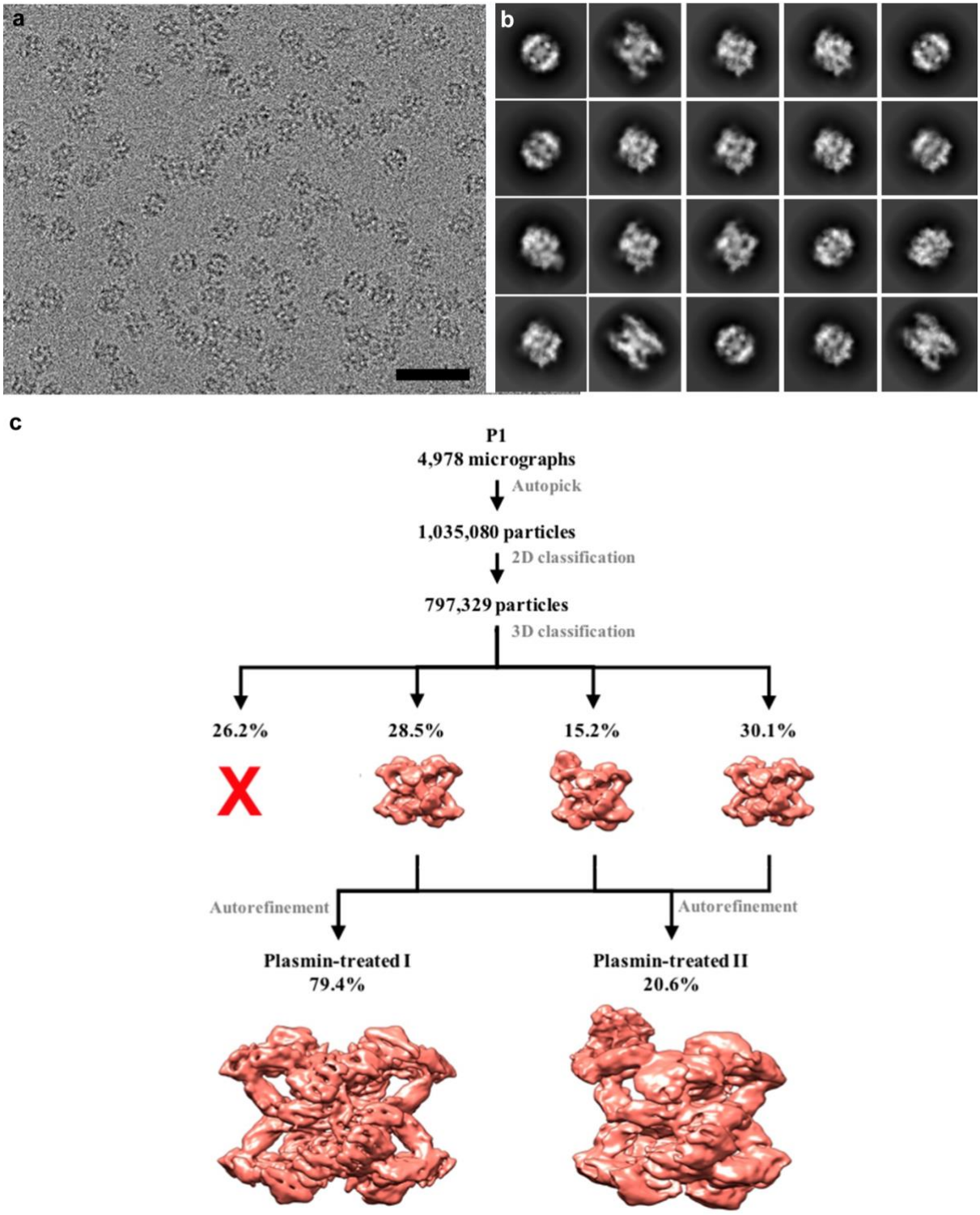
Supplementary Figure S5: Quality of the $(h\alpha_2M)_4$ density maps. Cryo-EM density maps around four different regions (the tethering loop, the BRD region IV, the disulfide-linked MG3-MG4, and the RBD) of $(h\alpha_2M)_4$ structures of native I, native II, semi-activated I, semi-activated II, naturally activated and trypsin-activate states. The cryo-EM density map is shown as a grey mesh with the corresponding atomic model with some selected residues indicated. The box of BRD region IV is shown below in a magnified view for each structure.



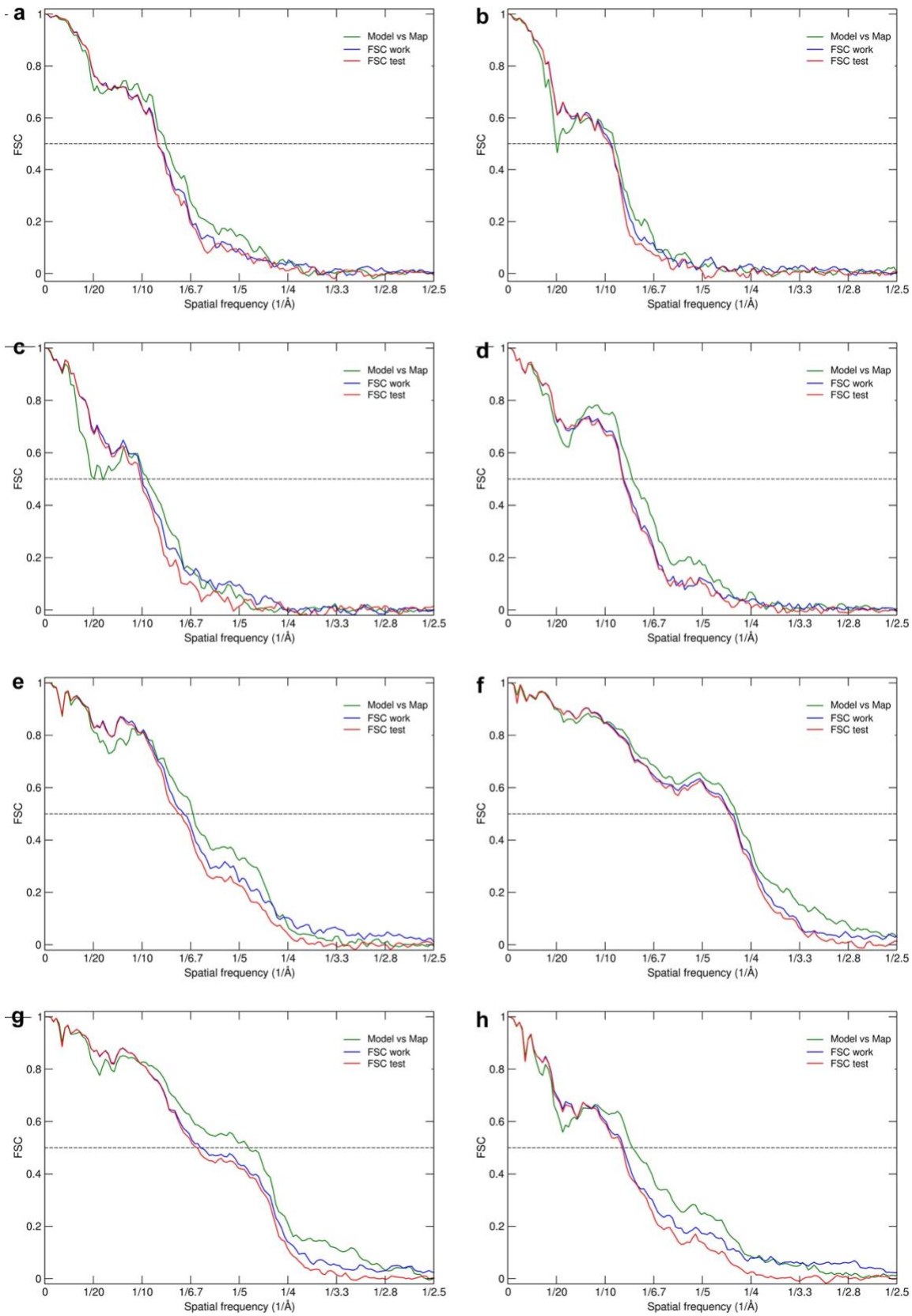
Supplementary Figure S6: Structural belt of MG4 domains and BRD tethering loops. (**a**) Two orthogonal views of the native II state in which the MG4 domains (orange) and the tethering loops (light blue) are highlighted. (**b**) Views as in (**a**) of the activated state with highlighted MG4 domains (magenta) and tethering loops (dark blue). (**c**) Superimposed MG4 domains and tethering loops of native II and activated states (color code as in **a** and **b**).



Supplementary Figure S7: The RBD is flexible in the cryo-EM density map of activated (hα$_2$M)$_4$. (**a**) Structure of the activated state (domain colors as in Figure S1). The thin rectangle indicates the transversal section shown in (**b**). (**b**) Section across the cryo-EM map without (left) and with atomic model regions of (hα$_2$M)$_4$ (right). The last visible hα$_2$M residue in the model is Glu1337, located in the connecting loop between CUB and RBD, and indicated with two molecules (arrowheads).



Supplementary Figure S8: Cryo-EM data processing of plasmin-treated (hα$_2$M)$_4$. (**a**) Cryo-EM image of (hα$_2$M)$_4$ complexes after plasmin treatment (bar, 500 Å). (**b**) 2D class averages of (hα$_2$M)$_4$ plasmin-treated particles. (**c**) Data processing workflow and structure determination of the plasmin-treated (hα$_2$M)$_4$ activated state. After 3D classification, two conformational states (I and II) were identified (percentages relative to total 2D selected particles of each preparation are indicated). Particles for each of these states were combined and further refined (percentages indicated relative to total 3D selected particles).



Supplementary Figure S9: Cryo-EM map quality and model validation. FSC of the refined model versus the map (green curve) and FSC work/FSC test validation curves (blue and red curves, respectively) for the eight maps calculated in this study: (**a**) native I, (**b**) native II, (**c**) semi-activated I, (**d**) semi-activated II, (**e**) fully activated, (**f**) trypsin-activated, (**g**) plasmin-activated I, and (**h**) plasmin-activated II states.

Supplementary Table S1. Cryo-EM data collection and refinement statistics.

	Native I **EMD-12747** **PDB: 7O7L**	**Native II** **EMD-12748** **PDB: 7O7M**	**Semiactivated I state** **EMD-12750** **PDB: 7O7N**	**Semiactivated II state** **EMD-12751** **PDB: 7O7O**	**Activated** **EMD-12752** **PDB: 7O7P**	**Trypsin-activated** **EMD-12753** **PDB: 7O7Q**	**Plasmin-activated I state** **EMD-12754** **PDB: 7O7R**	**Plasmin-activated II state** **EMD-12755** **PDB: 7O7S**
Data collection and processing								
Microscope	FEI Titan Krios	FEI Titan Krios	FEI Titan Krios	FEI Titan Krios	FEI Titan Krios	FEI Titan Krios	FEI Titan Krios	FEI Titan Krios
Detector	K2	K2	K2	K2	K2	K2	K2	K2
Magnification	47.755x	47.755x	47.755x	47.755x	47.755x	47.755x	130,000x	130,000x
Voltage (kV)	300	300	300	300	300	300	300	300
Electron exposure ($e^-/Å^2$)	39.6	39.6	39.6	39.6	39.6	40.0	38,7	38,7
Exposure per frame ($e^-/Å^2$)	0.99-1.27	0.99-1.27	0.99-1.27	0.99-1.27	0.99-1.27	1.12-1.25	0.96	0.96
Defocus range (μm)	-1.00 to -3.25	-1.00 to -3.25	-1.00 to -3.25	-1.00 to -3.25	-1.00 to -3.25	-0.70 to -2.5	-1.30 to -3.70	-1.30 to -3.70
Pixel size (Å)	1.047	1.047	1.047	1.047	1.047	1.047	1.052	1.052
Micrographs collected (no.)	12,143	12,143	12,143	12,143	12,143	6,514	4,978	4,978
Initial particles (no.)	1,625,000	1,625,000	1,625,000	1,625,000	1,625,000	933,186	1,035,080	1,035,080
Final particles (no.)	45,669	30,618	35,993	185,640	118,333	434,851	121,437	466,082
Symmetry imposed	C2	C2	C2	C2	C2	C2	C1	C1
Map resolution (Å)	4.5	6.6	7.3	4.8	4.6	3.6	3.9	4.3
FSC threshold	0.143	0.143	0.143	0.143	0.143	0.143	0.143	0.143
Map resolution range (Å)	3.5 – 22.4	3.5 – 23.7	3.5 – 25.8	3.5 – 19.7	3.5 – 9.8	2.2 – 10.1	2.8 – 15.3	2.8 – 25.9
Refinement								
Model resolution (Å)	7.51	8.65	8.17	6.95	6.37	4.10	4.32	6.83
FSC threshold	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
Mask correlation coefficient	0.64	0.57	0.57	0.69	0.69	0.76	0.75	0.58
Map sharpening B factor	LocalDeblur	LocalDeblur	LocalDeblur	LocalDeblur	LocalDeblur	LocalDeblur	LocalDeblur	LocalDeblur
Model composition								
Non-hydrogen atoms	44,614	44,614	42,372	42,642	40,295	40,550	40,530	41,466
Protein residues	5,640	5,640	5,370	5,376	5,100	5,126	5,126	5,236
Ligands								
NAG	38	38	34	42	36	36	36	38
BMA	4	4	0	6	0	4	4	5
MAN	2	2	0	4	0	0	0	1
ADP (B-factors)								
min								
max								
mean								
Protein	143.00	135.21	63.37	164.96	165.62	116.08	131.87	124.45
	819.33	999.99	821.45	806.53	490.80	279.28	415.19	564.41
	375.62	502.63	311.57	328.20	278.28	159.94	201.79	280.43
Ligand	293.91	293.91	238.72	293.91	239.72	164.29	164.29	168.41
	471.01	471.01	446.90	483.22	347.75	211.87	211.87	448.26
	374.75	374.75	329.92	396.34	299.90	187.07	186.77	247.15
R.m.s. deviations								
Bond lengths (Å)	0.007	0.006	0.006	0.010	0.009	0.009	0.013	0.009
Bond angles (°)	1.154	1.019	1.003	1.231	1.143	1.494	1.328	1.276
Validation								
MolProbity score	2.33	2.22	2.09	2.44	2.20	2.16	2.55	2.60
Clashscore	13.26	10.43	8.78	17.31	11.58	9.18	20.68	21.36
Rotamer outliers (%)	0.49	0.36	0.30	0.43	0.13	0.18	1.07	1.09
Ramachandran plot								
Favored (%)	82.49	83.74	87.26	82.15	87.04	84.27	81.39	79.18
Allowed (%)	16.83	15.73	12.33	17.39	12.65	14.44	18.28	20.13
Outliers (%)	0.68	0.53	0.41	0.47	0.31	1.29	0.33	0.69

Supplementary Table S2. Analysis of intra-subunit domains interactions.

	Extended subunit	Compact subunit
MG1	MG3, MG5, BRD	MG2, MG5, BRD
MG2	MG1, MG6, BRD, TED	MG1, MG6, BRD, MG7, CUB, TED
MG3	MG4, BRD-MG6 loop, MG7	MG4, MG6, MG7, BRD
MG4	MG3, MG5, BRD	MG3, MG5, BRD
MG5	MG1, MG4, BRD	MG1, MG4, BRD
MG6	MG2, MG3 (BRD-MG6 loop), BRD, MG7	MG2, MG3, BRD, MG7
BRD	MG1, MG2, MG4, MG5, MG6	MG1, MG2, MG3, MG4, MG5, MG6
MG7	MG3, MG6, CUB, RBD	MG2-MG3 loop, MG3, MG6, CUB
CUB	TED, RBD	MG2, TED
TED	MG2, CUB, RBD	MG2, CUB
RBD	MG7, CUB, TED	

Supplementary Table S3. Crystallographic data reprocessing and model re-refinement parameters.

Dataset	Peptidase-activated $(h\alpha_2M)_4$
PDB Access code	6TAV
Space group	$P2_12_12_1$
Cell constants (a, b, c, in Å)	130.8, 260.3, 281.8
Wavelength (Å)	1.0000
No. of measurements / unique reflections	339,260 / 70,972
Resolution range (Å)	95.9 – 4.20 (4.45 – 4.20) [a]
Completeness (%)	99.1 (96.8)
R_{merge}	0.095 (1.620)
R_{meas}	0.107 (1.865)
$CC^{1/2}$	0.998 (0.446)
Average intensity	10.0 (1.0)
B-Factor (Wilson) (Å^2)	198.6
Aver. multiplicity	4.8 (3.9)
No. of reflections used in refinement [in test set]	69,422 [831]
Crystallographic R_{factor} / free R_{factor}	0.235 / 0.282
Correlation coefficient F_{obs}-F_{calc} [test set]	0.939 [0.911]
No. of protein residues / non-hydrogen atoms /	5,234 / 40,826 /
covalent ligands	47 NAG, 9 BMA, 5 MAN [b]
Rmsd from target values	
bonds (Å) / angles (°)	0.011 / 1.25
Average B-factors (Å^2) (overall // mol. A/ B/ C/ D)	257 // 253 / 269 / 249 / 256
All-atom contacts and geometry analysis [c]	
Protein residues	
in favored regions / outliers / all residues	4,394 (84%) / 220 (4%) / 5,214
with outlying rotamers / bonds / angles / chirality / torsion	601 (13%) / 0 / 3 / 0 / 0
All-atom clashscore	4.3

[a] Data processing values in round brackets are for the outermost resolution shell. Model refinement parameters in square brackets are for the test set of reflection. [b] NAG, *N*-acetyl-D-glucosamine; BMA, β-D-mannose; and MAN, α-D-mannose. [c] According to the wwPDB X-ray Structure Validation Service.

Supplementary Table S4. Cryo-EM data collection.				
	Transient I state EMD-12941	**Transient II state EMD-12942**	**Transient III state EMD-12943**	**Transient IV state EMD-12944**
Data collection and processing				
Microscope	FEI Titan Krios	FEI Titan Krios	FEI Titan Krios	FEI Titan Krios
Detector	K2	K2	K2	K2
Magnification	47.755x	47.755x	47.755x	47.755x
Voltage (kV)	300	300	300	300
Electron exposure ($e^-/Å^2$)	39.6	39.6	39.6	39.6
Exposure per frame ($e^-/Å^2$)	0.99-1.27	0.99-1.27	0.99-1.27	0.99-1.27
Defocus range (μm)	-1.00 to -3.25	-1.00 to -3.25	-1.00 to -3.25	-1.00 to -3.25
Pixel size (Å)	1.047	1.047	1.047	1.047
Micrographs collected (no.)	12,143	12,143	12,143	12,143
Initial particles (no.)	1,625,000	1,625,000	1,625,000	1,625,000
Final particles (no.)	213,866	7,131	23,998	116,074
Symmetry imposed	C1	C1	C1	C1
Map resolution (Å)	5.2	12.0	9.1	9.3
FSC threshold	0.143	0.143	0.143	0.143
Map resolution range (Å)	2.3 – 37.2	4.9 – 47.9	2.1 – 55.9	4.4 – 55.9

Legend to Supplementary Movie S1:

Movie depicting the flexible arrangement of the distinct expanded and compact subunits within the native, intermediate and activated tetramers.

Legend to Supplementary Movie S2:

Movie depicting the transition between expanded and compact conformations of a single entire protomer (left panel), the MG1-MG6 block (central panel), and the MG7-CUB-TED-RBD block (right panel).