Cryo-EM structures shows the mechanistic basis of pan-peptidase inhibition by human α2-macroglobulin

Daniel Luque[1,#], Theodoros Goulas[2,3,#], Carlos P. Mata[4,#,+], Soraia R. Mendes[2], F. Xavier Gomis-Rüth[2,*] and José R. Castón[5,*]

[1]Spanish National Microbiology Centre, Institute of Health Carlos III, Madrid, Spain; [2]Proteolysis Lab; Department of Structural Biology; Molecular Biology Institute of Barcelona (CSIC), Barcelona Science Park, Barcelona, Catalonia, Spain; [3]Department of Food Science and Nutrition; School of Agricultural Sciences; University of Thessaly, Karditsa, Greece; [4]AstburyCentre for Structural Molecular Biology, School of Molecular and Cellular Biology, Faculty of Biological Sciences, University of Leeds, Leeds, UK; [5]Department of Structure of Macromolecules, Centro Nacional de Biotecnología (CNB-CSIC), Campus de Cantoblanco, Madrid, Spain.
[+] Present address: Spanish National Microbiology Centre, Institute of Health Carlos III, Madrid, Spain

[#]These authors contributed equally to this work
[*]Co-corresponding authors; correspondence to: F. Xavier Gomis-Rüth, fxgr@ibmb.csic.es, and José R. Castón, jrcaston@cnb.csic.es

---

ABSTRACT

Human plasma α2-macroglobulin (hα2M) is a multidomain protein with a plethora of essential functions, including transport of signaling molecules and endopeptidase inhibition in innate immunity. Here we dissected the molecular mechanism of the inhibitory function of the ~720-kDa hα2M tetramer through eight cryo-EM structures of complexes from human plasma. In the native complex, the hα2M subunits are organized in two flexible modules with an expanded conformation, which encloses a highly porous cavity in which the proteolytic activity of circulating plasma proteins is tested. Cleavage of bait regions exposed inside the cavity triggers rearrangement to a compact conformation, which closes openings and entraps the prey proteinase. After the expanded-to-compact transition, which occurs independently in the subunits, the reactive thioester bond triggers covalent linking of the proteinase, and the receptor-binding domain is exposed on the tetramer surface for receptor-mediated clearance from circulation. These results depict the molecular mechanism of a unique suicidal inhibitory trap.

SIGNIFICANCE STATEMENT

Human plasma α2-macroglobulin (hα2M) is a ~720-kDa homotetrameric particle with pan-peptidase inhibitory functions that transits between an open native conformation and a closed induced state, in which endopeptidases are trapped upon cleavage of an accessible bait region. We determined the molecular mechanism of this function through eight cryo-EM structures, which revealed that the hα2M subunits are organized in two flexible modules that undergo independent expanded-to-compact transitions. In the induced state, a reactive thioester bond triggers covalent linking of the proteinase, and a receptor-binding domain is exposed on the tetramer surface for binding to its specific cellular receptor for internalization and clearance from circulation. These results elucidate the long-awaited molecular mechanism of a historical suicidal inhibitory trap.

INTRODUCTION

The α2-macroglobulins are large multi-domain proteins found in animals and selected colonizing bacteria (1-8). The best characterized is human α2-macroglobulin (hα2M), a 1451-residue protein built

of 11 domains (for hα$_2$M domain nomenclature, see Supplementary Fig. 1a), which is produced by several cell types including macrophages, astrocytes, and hepatocytes. Four protomers associate to a ~720-kDa polyglycosylated dimer of disulfide-linked homodimers, (hα$_2$M)$_4$ (Supplementary Fig. 1b), which is the largest non-immunoglobulin protein in human plasma and constitutes 2-4% of its total protein content (1, 9). Its multiple molecular functions include endopeptidase inhibition, as well as sequestration and transport of growth factors, cytokines, and hormones (10, 11). It is also an acute-phase reactant in rodents and a chaperone that binds misfolded or inactivated proteins, and has many more moonlighting functions such as transglutamination and zinc/copper binding (4, 8, 12).

These disparate functions explain the universal physiological significance of hα$_2$M: it is part of the innate immune response against pathogens (13, 14) and a major hemostatic regulator of the cardiovascular system through its anticoagulant, procoagulant, and antifibrinolytic activities (4). It is an early marker of cardiac hypertrophy, as well as a potential diagnostic marker for myocardial infarct and for HIV patients with cardiac pathologies (5). In amyloidoses, it is prophylactic, as it binds the major component of β-amyloid deposits, the Aβ peptide, and mitigates its neurotoxicity and fibrillogenic capacity (8). In addition, it has anti-inflammatory, signaling, and apoptotic properties, is engaged in growth and tissue remodeling, and protects joint cartilage (15, 16). Its deregulation contributes to most major human diseases including Alzheimer's disease, AIDS, inflammatory diseases, diabetes, arthritis, tumor growth and progression, and cardiovascular conditions (8, 15, 17). No total hα$_2$M deficiency has been described in humans, which supports the idea that its absence is embryonically lethal (4, 18-20).

The best-characterized molecular function of hα$_2$M is its indiscriminate inhibitory action on endopeptidases irrespective of catalytic class, whether endogenous or exogenous. This hallmarks it as a unique pan-endopeptidase inhibitor (1, 4, 5, 7, 8, 21). In the cardiovascular system, it inhibits thrombin, factor Xa, activated protein C, plasma kallikrein and kallikrein-related peptidases, and plasmin (4, 17). As part of the humoral defense barrier, it inhibits bacterial and viral proteolytic virulence factors such as pseudolysin, HIV-1 proteinase, clostripain, and vibriolysin, as well as snake-venom proteinases (22, 23). Finally, by inhibiting neutrophil elastase, matrix metalloproteinases, and ADAM/adamalysin metallopeptidases, it participates in inflammatory processes and tissue turnover (15, 16).

Whereas the vast majority of peptidase inhibitors act through reversible "lock-and-key" mechanisms that sterically block the active-site cleft of target enzymes (24, 25), hα$_2$M operates according to a unique suicidal Venus fly-trap (7, 26) or trap-hypothesis mechanism (1), by which prey peptidases diffuse into an unreacted or native (hα$_2$M)$_4$ tetramer (Fig. 1, step 1). Once inside, peptidases access and cleave a flexible, multitarget bait region (Fig. 1, step 2) within an exposed bait-region domain (BRD, see Supplementary Fig. 1). Cleavage leads to a nascent state of the inhibitor, in which a buried reactive β-cysteinyl-γ-glutamyl thioester bond becomes exposed on a thioester domain (TED). The bond is attacked by surface lysine amines of the prey peptidase, which thus becomes covalently bound to the inhibitor through an ε-lysyl-γ-glutamyl crosslink (Fig. 1, step 3). Bait-region cleavage also triggers a large, irreversible structural rearrangement of the tetramer, leading to a reacted, induced, or activated state, which engulfs the peptidase without disturbing its active site, similarly to insect capture by the Venus fly-trap plant (Fig. 1, step 4). A large reorganization of (hα$_2$M)$_4$ can also be observed by treatment of the native species with nucleophilic chemicals such as methylamine (MA), which yield an activated species with open thioester bonds but intact bait regions, unable to inhibit peptidases (27-30). Native and activated states can be distinguished because they have different sedimentation coefficients and mobilities in native gel electrophoresis (13).

Once within the closed trap, the peptidase no longer cleaves large protein substrates, but is still accessible to small inhibitors and substrates through openings in the tetrameric particle. Moreover, the four receptor-binding sequences of the receptor-binding domains (RBD), which are cryptic in the native and nascent species (31), become exposed on the tetrameric particle surface (32) and are recognized by cell-surface receptors such as the low-density lipoprotein receptor-related protein (LRP1) for receptor-mediated endocytosis (Fig. 1, step 4) (33). Inside the cell, the peptidase:inhibitor complex is cleared in the lysosomes within minutes of complex formation (34).

This sequence of events has been established through painstaking biochemical analyses for decades, but its molecular determinants remain unknown. Attempts have been undertaken to study the structure of tetrameric hα$_2$M and mammalian orthologs to provide the molecular determinants of this inhibitory mechanism. Nonetheless, owing to the large size and intrinsic flexibility of hα$_2$M, the lack of suitable recombinant expression systems for large-scale production of homogeneous functional protein, and the

structural heterogeneity of samples purified from natural sources, only two small domains (the RBD and the macroglobulin-like domain MG2) have so far been described by X-ray crystallography (35-39). For full-length assemblies, 20-40 Å resolution maps derived from electron microscopy (EM) of negative-stained or vitrified samples, and a crystallographic map of a hypothetically MA-activated $(h\alpha_2M)_4$ to 10 Å resolution, have shown morphological variations of the native and activated states of the $h\alpha_2M$ tetramer, with no details at the domain level (29, 40-43). Moreover, the 4.3-Å resolution crystal structure of activated $(h\alpha_2M)_4$ showed conserved structural features with the proteolytically activated C3b complement component (derived from the native C3 factor) (26). The most recent structural analyses of native $h\alpha_2M$ are limited to homology models calculated from the native C3 complement component and docking in low-resolution EM maps (26, 44).

Thus, this study addresses three questions relevant to $h\alpha_2M$ biochemistry: (i) the structure of the native $(h\alpha_2M)_4$ complex when poised to trap plasma endopeptidases and that of the peptidase-activated inhibitor; (ii) the conformational changes and the intermediates on the path between the native and activated states; and (iii) how large proteases are entrapped in a cavity with insufficient volume *a priori*. We used authentic human protein to determine eight cryo-EM structures that represent functional states of $(h\alpha_2M)_4$, unbound and in complex with physiologically relevant endopeptidases. The resulting structures show striking conformational rearrangement and provide detailed insight into the molecular mechanism of a unique sequential inhibitory mechanism.

RESULTS AND DISCUSSION

Cryo-EM structure analysis of the $(h\alpha_2M)_4$ complex. Authentic $(h\alpha_2M)_4$ from human plasma was vitrified and imaged by cryo-EM (Supplementary Fig. 2a). A total of ~1,625,000 particles collected on 300-keV FEI Titan Krios microscopes were automatically picked. Owing to the intrinsic flexibility and heterogeneity of the structures, they were subjected to several rounds of exhaustive 2D and 3D classification (Supplementary Fig. 2b, c). These resulted in five distinct cryo-EM structures at different states of reaction, to resolutions spanning 4.5-7.3 Å as estimated by the values at which the Fourier shell correlation (FSC) coefficient equals 0.143 (Supplementary Fig.3a-e, Supplementary Table 1).

To obtain a homogeneous population of fully activated $(h\alpha_2M)_4$ as a control, a purified sample of mostly native protein was incubated with trypsin and its structure determined to 3.6 Å resolution (Supplementary Figs. 3f and 4). This structure enabled unambiguous assignment of the polypeptide chain of the whole particle, which gives rise to a C2 tetramer featuring a compact cage with several openings of variable size and a large inner cavity, the prey chamber.

Based on protomer conformations, either expanded or compact, on their relative arrangement within the tetramers, and on the volume and occupancy of the prey chambers in the distinct structures (i.e., the presence of entrapped proteinases in its interior chamber), we identified five major states of $(h\alpha_2M)_4$, which we termed native I and II, semi-activated I and II, and fully activated, which corresponded to distinct reaction intermediates (Fig. 2a). With maximal dimensions of 210 x 185 x 150 Å, the native tetramer is substantially larger than the activated form, which spans 140 x 210 x 140 Å (Fig. 2a) and provides an explanation for the higher electrophoretic mobility of the latter (45). The native states have four large and four small openings of 70 x 50 Å and 30 x 20 Å, respectively, and enclose a prey chamber of ~600 nm^3, whereas the activated state has 12 small openings of 30 x 40 Å and encloses a ~300-nm^3 prey chamber (Fig. 2a, b, d).

Each $(h\alpha_2M)_4$ protomer has a vicinal, a disulfide-linked, and an opposite neighbor (Fig. 2a, left). In the native and semi-activated states, one vicinal dimer was solved to much better resolution than the other (Supplementary Fig. 3). For the native I and II states, resolution of the upper pair of vicinal protomers was sufficient to determine unambiguously their subunit organization (Fig. 2a, red and yellow). By contrast, the intrinsic flexibility of the lower vicinal protomers (~11-22 Å resolution) required development of a strategy that combined rigid and flexible domain refinement (Fig. 2a, green and blue). The semi-activated I state behaved similarly to native states, with the upper pair more flexible than the lower pair (8-25 Å vs. 3.5-12 Å resolution). Finally, the semi-activated II and activated states showed density with more features, which allowed accurate modeling of the four polypeptide chains (Supplementary Fig. 5; see Methods for details). Each $h\alpha_2M$ subunit has eight surface-located, protruding *N*-linked glycans at positions 55, 70, 247, 396, 410, 869, 991, and 1424 (Fig. 2c, indicated

for semi-activated state II, green sphere model; see UniProt entry P01023 for sequence numbering). Structural alignment of compact or expanded protomers in all states showed *rmsd* values <1.4 Å, indicating high structural similarity in the respective conformations.

The native (I and II), semi-activated (I and II), and activated states were captured at a ~1:2:1 ratio, although in a specific preparation (preparation P2), more than half the total particles were initially classified as native (Supplementary Fig. 2). These data underpin the existence of a substantial population (25-42%) of intrinsically activated $(h\alpha_2M)_4$ in preparations assumed to be mainly native, following long-established standard biochemical purification procedures (27).

The functional tetrameric states are built of expanded and compact protomer conformations. Each $h\alpha_2M$ subunit structure consists of 11 or 10 domains for the expanded and compact conformations, respectively, as the C-terminal RBD is flexible and thus not assignable in the latter (Fig. 3a; Supplementary Fig. 1). The first seven domains (MG1-MG7) are concatenated ~110-residue β-sandwiches with a three- and a four-strand antiparallel β-sheet (Fig. 3a). Domains MG1-MG6 form an N-terminal module, which is similarly arranged in all structures as a 1.5-turn right-handed superhelix that encircles an ellipsoidal opening of 30 x 20 Å (Fig. 2b, dashed ellipses). The 127-residue BRD is an extended, flexible domain inserted between the fourth and fifth β-strands of MG6 (Supplementary Fig. 1a); the domains downstream of MG6 form a C-terminal module that adopts two different conformations: expanded or compact. MG7 acts as a hinge domain that connects MG1-MG6 with the 116-residue CUB domain (acronym of complement protein subcomponents C1r/C1s, urchin embryonic growth factor, and bone morphogenetic protein 1), which consists of two four-stranded antiparallel β-sheets. The 315-residue helical TED domain is inserted between the third and fourth β-strands of CUB and includes the scissile thioester bond. TED adopts a thick disk-like structure that consists of six concentric β-hairpins arranged as a six-fold α-propeller around a central shaft (Fig. 3a). The thioester bond is formed between the cysteine side-chain sulfur and the glutamine side-chain carbonyl of the conserved motif Cys972-Gly-Glu-Gln975. Finally, the C-terminal domain of this module is the RBD, a variant of the MG fold with an extra α-β-α unit.

In all five functional states identified in the native preparation, the $h\alpha_2M$ monomers adopt either a compact or an expanded conformation for each pair of vicinal dimers. In native states I and II, all subunits are expanded (Fig. 2a). In the native I state, the TED domains of vicinal protomers (within the top dimer) contact each other through their respective fifth helices (Fig. 2a, left; subunits 1 and 3). By contrast, the protomers are dynamic in the bottom dimer (Fig. 2a, left; subunits 2 and 4) and swing around the interface between subunits 1 and 2 and subunits 3 and 4. These subunits encompass the inter-protomeric disulfide bond (Fig. 2b, indicated in the native II state, red and green spheres) as well as BRD-mediated contacts between opposite monomers of subunits 1 and 4 and subunits 3 and 2 (Fig. 2a, dashed rhombuses). As a result, the fifth helices of the bottom TED domains can either be separated by 36 Å (Fig. 2a,b, left; green and blue protomers) or in contact, as found in the top domains (Fig. 2a, center-left; green and blue protomers). Both expanded protomers, in a distal or proximal position, share an almost identical conformation (*rmsd* < 0.7 Å). This yields the native II state, which is quasi-symmetric for the top and bottom dimers as the bottom dimer is resolved at lower resolution, indicating it is more flexible or dynamic than the top dimer. In semi-activated states I and II, which correspond to the nascent state (46) (Fig. 1), the bottom vicinal subunits adopt a compact conformation (Fig. 2a, center and center-right; green and blue protomers), while the top vicinal subunits are dynamic and very similar to the native states. In the fully activated state, all subunits show a compact conformation (Fig. 2a, right, and Supplemental Movie 1).

Whereas the architecture of the N-terminal module is equivalent in the expanded and compact conformations and merely undergoes a 22°-rotation around MG4 (Fig. 3b), the C-terminal modules are coordinately shifted, with additional 45°- and 55°-rotations for CUB and TED, respectively, around the hinge-domain MG7 (Fig. 3c, Supplementary Movie 2). Following this conformational change, all domains except MG3 and MG4 are arranged as a triangular prism in the compact conformation, the hallmark structural motif of the activated state (Fig. 2d, dashed triangles). Vicinal TEDs interact through their fifth and third α-helix in the expanded and compact conformation, respectively. TED also contacts MG4 of the respective disulfide-linked subunit, as well as MG1s and BRDs of the respective vicinal subunit.

The conformational rearrangement between the expanded and compact conformations implies notable changes in the contact points within the top and bottom dimers of vicinal protomers

(Supplementary Table 2). Whereas the internal cavity in the native tetramer is framed by MG1-MG5, BRD, TED and a small part of RBD, in the activated tetramer only MG1-MG5 and TED face the prey chamber as the exposed surfaces of MG1, MG2 and MG5 are substantially reduced. All domains have a very similar fold in both conformations, except for TED. The four MG4 domains in the equatorial plane of the tetramer are nearly invariant in all functional states despite the massive structural changes that underlie the transition. Together with the BRD tethering loops (see next section), these domains form a constant belt responsible for the structural integrity of the tetramer in either state (Supplementary Fig. 6).

The major players: BRD, TED and RBD. The BRD spans 75 Å and contains little regular secondary structure. It consists of four regions in the expanded conformation (Fig. 4a, I-IV). Region I (Gln602-Leu627) is an N-terminal extended segment that is linked to the third β-strand of MG6 and contains two short α-helices. Region II (Thr628-Gln694) is a compact region that includes the tethering loop (Ile643-Cys689) and interacts with the symmetric region of the opposite protomer. Region III (Pro690-Thr728) is the 39-residue bait segment that is disordered and invisible in our maps. Finally, region IV (Glu729-Arg732) is the C-terminal segment connected with the fourth β-strand of MG6 on the outer surface of the protomer through one of the openings in the expanded conformation. This opening is framed by MG6, MG2, MG3, and the loop connecting MG2 with MG3, and it allows passage of an intact polypeptide chain (Fig. 4b, dashed circle; Supplementary Fig. 5). In the compact conformation (Fig. 4c), the beginning of the C-terminal segment at Thr730 is on the outer surface and interacts with the seventh strand of MG7. A small rearrangement in this region causes MG7 to fold over MG6, which almost occludes the opening (Fig. 4d; Supplementary Fig. 6). The BRD-MG6 connecting segment thus regulates the expanded-to-compact transition by acting as a trigger that must be cleaved for displacement of the new N terminus to the exterior surface. This is consistent with the assumption that MA-activated (hα$_2$M)$_4$, whose bait region is intact, adopts a structure distinct from the peptidase-activated state (32). These results further indicate that the reported crystal structure of induced (hα$_2$M)$_4$ (Supplementary Table 3), originally thought to be a non-cleaved MA-induced variant (26), cannot contain an intact bait region but is a cleaved, proteinase-induced species. Its overall conformation is equivalent to the structure of the fully peptidase-activated state (see below and Fig. 2a, right).

At the C terminus, protomers in expanded conformation, both in activated and semi-activated states, have a well-defined RBD, surrounded by MG7, CUB, and TED (Fig. 4e, dark grey; Supplementary Fig. 5). The second α-helix of RBD, which includes the receptor-binding region (47), points towards the prey chamber and is occluded (Fig. 4e,f, orange). Following activation by cleavage of the bait region, the substantial rearrangement of MG7-CUB-TED-RBD disrupts the interaction among these domains. In the activated (hα$_2$M)$_4$ structures, this causes the RBD to project away from the outer surface and become flexible and thus disordered, except for a small, blurred density for its first three or four residues (Supplementary Fig. 7). The receptor-binding region thus becomes accessible for interaction with LRP1. In previous structural studies, the RBD was visible in the crystal structure of activated (hα$_2$M)$_4$ for only one of the four subunits, probably due to tetramer-tetramer interactions within the crystal, and obviously in a physiologically irrelevant position (26).

The TED thioester bond is 16 Å from the second RBD α-helix in the expanded conformation of the native I state (Fig. 4e,f, yellow). Small nucleophiles such as MA can access and open the thioester bond without BRD processing. This would be sensed by the RBD, resulting in exposure of its receptor binding sequence to trigger removal of this inactivated hα2M complex. Alternatively, these two critical regions might be adjacent in the highly porous structure of the native complex to remain simultaneously inaccessible to the external environment. Access to TED is limited in the expanded conformation, as it is found in a cleft surrounded by bulky hydrophobic side chains (Phe1024, Leu1087, Tyr970, and His 1404) (Fig. 4f). Moreover, the glycan bound to Asn1424 is nearby (Fig. 4e, white), which might prevent thioester-bond opening by surface amines of circulatory proteins and other substances. In contrast, in the compact conformation, thioester bonds are accessible and face the prey chamber, and are thus prepared for reaction with surface lysines of the prey (see below).

We did not identify a native state in which both vicinal dimers displayed simultaneously non-interacting TED domains (Fig. 2a, left; green and blue protomers), which suggests that this intermediate is unstable or very short-lived and thus cannot be isolated by cryo-EM. To communicate the separated, interacting TED positions, two small contact areas between vicinal dimers are critical: the opposite BRD tethering loops and the MG3 and MG4 domains. They are symmetrically crosslinked through

intermolecular disulfide bonds (Cys278 of MG3 with Cys431 of MG4; see Supplementary Figs. 1 and 5). The structural integrity of the native tetramer resides only on the tethering loop-mediated, non-covalent interaction. This TED rigid-body rearrangement might capture non-specific circulatory proteins, which are conveyed to the prey chamber. Most of them will not interact with the hα$_2$M subunits and will be released directly through the 70 x 50 Å openings, which in turn might also be the direct entrance of many soluble proteins. By contrast, those with endopeptidolytic activity will cleave the bait region and trigger conformational rearrangement from the expanded to the compact conformation, which halves the volume of the prey chamber and restricts the openings to maximumally 30 x 40 Å, thus preventing prey from escaping. This mechanism facilitates scanning of many plasma proteins, and only endopeptidases that must be quickly withdrawn from the circulation would be "swallowed".

The symmetric activated state. Intrinsically activated (hα$_2$M)$_4$ was resolved as a single complex in which the heterogeneous proteinases trapped were resolved as a featureless density that occupied the prey chamber (Fig. 5a, red surface). Since the purified samples contained 25-55% native (hα$_2$M)$_4$, we analyzed a sample activated with bovine trypsin *in vitro* (Supplementary Fig. 4). The tetramer in the resulting compact conformation was refined to 3.6 Å resolution (Supplementary Fig. 3f), and the density corresponding to the trapped proteinases was resolved as two independent volumes (Fig. 5b, red surface). This result indicates that each disulfide-linked dimer participates in the binding of one trypsin molecule, in accordance with an inhibition stoichiometry of two proteinase molecules per inhibitor tetramer (13). Although the entrapped peptidases are probably not static inside the cavity, they could be tentatively docked (223 residues; PDB 1MTS) into these two densities (Fig. 5b,c, red ribbons). The density corresponding to the two peptidase moieties, in both intrinsically and trypsin-activated (hα$_2$M)$_4$, suggest intimate contact with the two symmetric MG4 loops, in particular with Gln435, and with Gln975 of the vicinal subunit thioester bonds for covalent binding (Fig. 5c).

The subunit-mediated activated state. (hα$_2$M)$_4$ was also activated *in vitro* with plasmin, a 791-residue protease that cannot be accommodated in the internal cavity of the activated tetramer, but is nevertheless inhibited efficiently by (hα$_2$M)$_4$ following cleavage after Arg719 (48). Whereas trypsin-incubated (hα$_2$M)$_4$ adopted an activated state in which the four subunits were in the same compact conformation, plasmin-incubated (hα$_2$M)$_4$ was resolved in two activated states, with four (plasmin-activated I state) and three (plasmin-activated II state) compact subunits (Supplementary Figs. 3g,h and 8), which account for 79 and 21% of the activated complexes, respectively (Fig. 6a). The density in the prey chamber attributable to a bound prey was much larger in the complex with three compact subunits (Fig. 6a, right), which suggests the absence of the large peptidase in the particles with four compact protomers. We thus hypothesize that acquisition of the compact conformation for the fourth subunit is prevented by steric hindrance through parts of the peptidase protruding from the inhibitor tetramer, even though the bait region is cleaved. Plasmin-activated II state therefore provides a structural explanation for the inhibitory potential of (hα$_2$M)$_4$ for endopeptidases larger than the prey chamber.

We sought to determine whether both subunits within a vicinal pair must simultaneously undergo the transition from the expanded to the compact conformation. We thus reanalyzed native (hα$_2$M)$_4$ complexes using a computational approach that included symmetry expansion of the previously calculated map, with imposed C2 symmetry and subsequent 3D classification to detect unique features of each subunit (Supplementary Fig. 2d, Supplementary Table 4). In addition to the two native states built of expanded subunits and to the fully activated state with four compact subunits, we isolated new tetramer arrangements with one, two, and three activated subunits, which made up 22% of the tetramers included in the 3D classification and were probably partially-activated intermediates (Fig. 6b). We hypothesize that once a first bait region is cleaved by the prey proteinase, the transition from the expanded to the compact conformation exposes the thioester bond to covalently fix the peptidase. This would cause the bait region of the vicinal subunit to be cleaved with higher probability, as the prey cannot escape from the trap, which would make the second protomer to become compact. In some cases, however, the peptidase would not be immediately bound by the thioester bond, and would thus gain access to the bait regions of the opposite and/or disulfide-linked subunit. These cleavages would produce species with three compact protomers. These intermediate structures indicate that conformational transition for each subunit is only dependent on cleavage of its own bait region, independently of the other subunits.

Molecular mechanism of (hα$_2$M)$_4$ function. Based on the structures obtained in this study, we propose a molecular mechanism of plasma endopeptidase inhibition by (hα$_2$M)$_4$ (Fig. 7). The native,

unreacted tetramer, which is built of subunits in expanded conformation, is extremely porous and flexible, and comprises four large 50 x 70 Å openings that would fuse to yield a large, irregular cavity. The openings are framed by five of the eight *N*-linked glycans, which contribute to thermal stability and high solubility in blood plasma but not to the inhibitory capacity of hα$_2$M (49). We hypothesize that this structure could act as a sieving complex in a crowded plasma environment, which is primed with four internal bait regions exposed to the particle lumen for testing of prey endopeptidases. Processing of the bait region leads to a conformational rearrangement of the particular protomer from an expanded to a compact conformation; this exposes the reactive thioester bond from TED for covalent entrapping of the peptidase, and the RBD for recognition by its cognate receptor for subsequent endocytosis. Native subunits become activated as they are cleaved, although for large trapped proteinases, steric constraints or clashes with the prey might prevent the structural transition in some of the processed hα$_2$M subunits. This dynamic complex is stabilized throughout this peptidase inhibition process by an inalterable structural belt formed by the MG4 domains and the BRD tethering loops.

A comparison with the mechanism derived for *Escherichia coli* α$_2$-macroglobulin (ECAM) shows substantial differences, despite a generally similar domain architecture in both inhibitors (50). While ECAM is monomeric and anchored to the cytosolic membrane of the bacterium facing the periplasm, hα$_2$M is tetrameric and secreted to the plasma in humans. In the former, covalent linkage and steric hindrance of peptidases thus inhibit activity, but only against very large substrates. This modus operandi has been dubbed a "snap-trap mechanism". In contrast, hα$_2$M inhibition is elicited through physical entrapment in a large cage following a "Venus-flytrap mechanism", in which preys are still active against small substrates and inhibitors that can enter the cage through several apertures.

In conclusion, our structural studies report the molecular determinants of the mechanism of action of (hα$_2$M)$_4$, a universal pan-peptidase inhibitor that has been the subject of considerable biochemical and biophysical study since its discovery in 1946 (51). We document a stepwise Venus-flytrap mechanism unique among peptidase inhibitors in its irreversibility and versatility, which enables to sequester endogenous and exogenous peptidases from plasma and remove them from the circulation. Given the role of these complexes in numerous essential processes, knowledge of their structure provides a starting point for further studies of the many functions of (hα$_2$M)$_4$.

MATERIALS AND METHODS

Cryo-EM data collection. Wild-type authentic native (hα$_2$M)$_4$ was isolated from thawed frozen plasma from healthy human donors, which was de-identified prior to use in this study. The protein was purified, assessed for peptidase-inhibition competence as described (15, 26, 49), and verified to be equivalent to protein purified from fresh plasma in functional and physiological assays. Aliquots of pure protein (5 μL) were diluted to 0.1 mg/mL, applied to R2/2 300 mesh acetone vapor-treated copper grids, and vitrified using a Leica EM CPC cryofixation unit. Data were collected on FEI Titan Krios electron microscopes operated at 300 kV, and images recorded with Gatan K2-summit cameras in counting mode using the EPU Automated Data Acquisition Software for Single Particle Analysis (ThermoFisher Scientific). The total number of recorded movies, nominal magnification, calibrated pixel size at the specimen level, total exposure, exposure per frame, and defocus range for each specimen are described in Supplementary Table 1.

Image processing. Movies were drift-corrected and dose-weighted with Motioncor2 (Zheng et al., 2017), and contrast transfer function (CTF) values were estimated with CTFFIND4.1 (52) using non-dose-weighted micrographs. All subsequent image processing was with RELION 2.1 (53, 54) within Scipion (55), unless otherwise stated. The data processing workflows for (hα$_2$M)$_4$ complexes purified from plasma, as well as for trypsin- and plasmin-treated complexes, are described in Supplementary Figs. 3, 5 and 9, respectively. Particle statistics are indicated in Supplementary Table 1. Class averages from preliminary datasets were used as templates for subsequent automated particle picking with Gautomatch (written by Kai Zhang, https://www2.mrc-lmb.cam.ac.uk/research/locally-developed-software/zhang-software/). Particles were then extracted, normalized, and subjected to several rounds of reference-free two-dimensional (2D) classification to discard particles from 2D classes that did not show secondary structural elements. Selected particles were 3D-classified, imposing C2 symmetry for plasma-purified and trypsin-treated (hα$_2$M)$_4$ complexes, and C1 symmetry for plasmin-treated

complexes. Classes representing equivalent conformational states were pooled and included in a 3D auto-refinement. Particles assigned to a native state were submitted to an additional round of 3D classification without alignment to identify and refine particles corresponding to native I and II states. To identify and classify intermediate conformations between the $(h\alpha_2M)_4$ major states in plasma, the C2 symmetry of these states was expanded (56) and particles were subjected to 3D classification without alignment (particle processing is indicated in Supplementary Table 4). Classes representing equivalent conformational states were pooled and included in a 3D auto-refinement. Local resolution was estimated using MonoRes (57) and unsharpened maps treated by local resolution-based sharpening in LocalDeblur (58).

Model building and refinement. The $h\alpha_2M$ crystal structure (26) (PDB 4ACQ) was first docked manually as a rigid body into the trypsin-activated locally-sharpened density map, and then subjected to real-space fitting with the Fit_in_Map routine of Chimera (59). A first step of real-space refinement was performed with Phenix (60) applying global minimization, local grid search and atomic displacement parameter refinement (ADP) protocols. The model was then rebuilt manually in Coot (61) to optimize the fit to the density for one set of disulfide-linked subunits (protomers 1 and 2). This asymmetric unit was then C2-symmetrized and further refined in Phenix with similar options as above, and with secondary structure, non-crystallographic symmetry (NCS), side chain rotamer, and Ramachandran restraints, as well as with hydrogens in riding positions. This model was used as starting model to build the compact subunits, and was fitted into the naturally-activated and plasmin-activated I sample maps, as well as in densities corresponding to protomers 2, 3, and 4 of plasmin-activated II, as a starting point for model building of these states. In Chimera, activated protomer 1 coordinates were flexible fitted in the native I protomer 1 density, considering each domain as a rigid body. These docked domains were used as the starting point for model building of the prototype of an expanded subunit, whose coordinates were fitted in expanded protomers of native I (protomer 2), native II (protomers 1 and 2), semi-activated I (protomer 1), semi-activated II (protomer 1) and plasmin-activated II (protomer 1). Fitted coordinates were checked manually in Coot and C2 symmetrized (to generate protomers 3 and 4). The first step of real-space refinement was performed in Phenix, with morphing and simulated annealing options, followed by the steps described above. Refinement statistics are listed in Supplementary Table 1.

Model validation and analysis. The FSC curves between model and map after local sharpening (Model *vs.* Map) are shown in Supplementary Fig. 10. For cross-validation against overfitting, the atoms in the final atomic models were displaced by 0.5 Å in random directions using Phenix. The shifted coordinates were then refined against one of the half-maps (work set) in Phenix using the same procedure as for refinement of the final model. The other half-map (test set) was not used in refinement for cross-validation. FSC curves of the refined shifted model against the work set (FSCwork) and against the test set (FSCtest) are shown in Supplementary Fig. 10. The FSCwork and FSCtest curves do not diverge markedly, consistent with the absence of overfitting in the final models. The quality of the atomic model was assessed by analysis of the basic protein geometry, Ramachandran plots, and clash analysis, and validated with Coot and MolProbity (62) as implemented in Phenix, and with the Worldwide PDB (wwPDB) OneDep System (https://deposit-pdbe.wwpdb.org/deposition/). Graphics were produced using UCSF Chimera.

Re-refinement of the crystallographic structure of peptidase-activated $h\alpha_2M$. To obtain a more accurate model of the reported crystal structure of peptidase-activated $h\alpha_2M$ with complete side chains and correct glycan structure, the X-ray diffraction data (PDB 4ACQ; (26)) were reprocessed to 4.2 Å resolution with current versions of the XDS (63) and XSCALE programs (64) (Supplementary Table 3). A test set for R_{free} monitoring was chosen in thin shells with SFTOOLS within the CCP4 suite of programs (65). The coordinates of the $h\alpha_2M$ part of the cryo-EM model of the complex with trypsin were superimposed onto the original crystallographic coordinates and refined in reciprocal space against the newly processed data with Phenix and Buster/TNT (66); this included hydrogens (set to zero occupancy) at riding positions for the protein residues, non-crystallographic symmetry restraints, and translation/libration/screw-rotation refinement. This refinement alternated with manual model building with Coot (see Supplementary Table 3). The eight glycosylation sites of each protomer were rebuilt to match the recently determined glycan structure of $h\alpha_2M$ (67).

Data availability. The atomic coordinates and cryo-EM density maps were deposited in the Protein Data Bank and EM Data Bank with codes 7OTL and EMD-12747 (native I); 7O7M and EMD-12748 (native II); 7O7N and EMD-12750 (semi-activated I); 7O7O and EMD-12751 (semi-activated II); 7O7P

and EMD-12752 (activated); 7O7Q and EMD-12753 (trypsin-activated); 7O7R and EMD-12754 (plasmin-activated I); 7O7S and EMD-12755 (plasmin-activated II). The cryo-EM density maps of intermediate structures with one, two (2 maps), and three activated monomers, corresponding to transient I-IV states, were deposited in the EM Data Bank with codes EMD-12941, EMD-12942, EMD-12943 and EMD-1294. The final X-ray crystallography model was likewise deposited at the Protein Data Bank (PDB 6TAV) and supersedes PDB entry 4ACQ.

ACKNOWLEDGEMENTS

AUTHOR CONTRIBUTION

T.G., S.R.M. performed biochemical experiments and purified complexes; D.L., T.G. and C.P.M. acquired the cryo-EM images; D.L. processed images and C.P.M. built the models; T.G., D.L., C.P.M., S.R.M., F.X.G.-R. and J.R.C. analyzed data; F.X.G.-R. and J.R.C. conceived and coordinated the project; F.X.G.-R. and J.R.C. wrote the paper with input from all authors.

CONFLICT OF INTERESTS

The authors declare no financial or non-financial conflicts of interest with the contents of this article.

REFERENCES

1. A. J. Barrett, P. M. Starkey, The interaction of α_2-macroglobulin with proteinases. Characteristics and specificity of the reaction, and a hypothesis concerning its molecular mechanism. *Biochem. J.* **133**, 709-724 (1973).
2. A. Budd, S. Blandin, E. A. Levashina, T. J. Gibson, Bacterial α_2-macroglobulins: colonization factors acquired by horizontal gene transfer from the metazoan genome? *Genome Biol.* **5**, R38 (2004).
3. N. Doan, P. G. W. Gettins, α-Macroglobulins are present in some Gram-negative bacteria: characterization of the α_2-macroglobulin from *Escherichia coli*. *J. Biol. Chem.* **283**, 28747-28756 (2008).

4. V. Ignjatovic, E. Mertyn, P. Monagle, The coagulation system in children: developmental and pathophysiological considerations. *Semin. Thromb. Hemost.* **37**, 723-729 (2011).

5. A. A. Rehman, H. Ahsan, F. H. Khan, α_2-Macroglobulin: a physiological guardian. *J. Cell. Physiol.* **228**, 1665-1675 (2013).

6. S. G. Wong, A. Dessen, Structure of a bacterial α_2-macroglobulin reveals mimicry of eukaryotic innate immunity. *Nat. Commun.* **5**, 4917 (2014).

7. T. Goulas *et al.*, Structural and functional insight into pan-endopeptidase inhibition by α_2-macroglobulins. *Biol. Chem.* **398**, 975-994 (2017).

8. S. Seddighi, V. Varma, M. Thambisetty, α_2-Macroglobulin in Alzheimer's disease: new roles for an old chaperone. *Biomark. Med.* **12**, 311-314 (2018).

9. P. M. Starkey, A. J. Barrett, Inhibition by α-macroglobulin and other serum proteins. *Biochem. J.* **131**, 823-831 (1973).

10. S. L. Gonias *et al.*, α2-macroglobulin and the α2-macroglobulin receptor/LRP. A growth regulatory axis. *Ann. N. Y. Acad. Sci.* **737**, 273-290 (1994).

11. C. T. Chu, S. V. Pizzo, Interactions between cytokines and α2-macroglobulin. *Immunol. Today* **12**, 249 (1991).

12. N.-M. Liu *et al.*, Transcuprein is a macroglobulin regulated by copper and iron availability. *J. Nutr. Biochem.* **18**, 597-608 (2007).

13. L. Sottrup-Jensen, α-Macroglobulins: structure, shape, and mechanism of proteinase complex formation. *J. Biol. Chem.* **264**, 11539-11542 (1989).

14. P. B. Armstrong, Role of α2-macroglobulin in the immune response of invertebrates. *Invertebrate Surviv. J.* **7**, 165-180 (2010).

15. C. J. Liu, The role of ADAMTS-7 and ADAMTS-12 in the pathogenesis of arthritis. *Nat. Clin. Pract. Rheumatol.* **5**, 38-45 (2009).

16. L. Troeberg, H. Nagase, Proteases involved in cartilage matrix degradation in osteoarthritis. *Biochim. Biophys. Acta* **1824**, 133-145 (2012).

17. J. Schaller, S. S. Gerber, The plasmin-antiplasmin system: structural and functional aspects. *Cell. Mol. Life Sci.* **68**, 785-801 (2011).

18. W. Borth, Alpha 2-macroglobulin, a multifunctional binding protein with targeting characteristics. *FASEB J* **6**, 3345-3353 (1992).

19. R. C. Roberts, Protease inhibitors of human plasma. Alpha-2-macroglobulin. *J Med* **16**, 129-224 (1985).

20. S. Zucker, R. M. Lysik, M. H. Zarrabi, J. J. Fiore, D. K. Strickland, Proteinase-alpha 2 macroglobulin complexes are not increased in plasma of patients with cancer. *Int J Cancer* **48**, 399-403 (1991).

21. I. Garcia-Ferrer, A. Marrero, F. X. Gomis-Rüth, T. Goulas, α_2-Macroglobulins: structure and function. *Subcell. Biochem.* **83**, 149-183 (2017).

22. A. F. Kisselev, K. von der Helm, Human immunodeficiency virus type 1 proteinase is rapidly and efficiently inactivated in human plasma by α_2-macroglobulin. *Biol. Chem. Hoppe Seyler* **375**, 711-714 (1994).

23. E. F. Sánchez, R. J. Flores-Ortiz, V. G. Alvarenga, J. A. Eble, Direct fibrinolytic snake venom metalloproteinases affecting hemostasis: structural, biochemical features and therapeutic potential. *Toxins (Basel)* **9**, 392 (2017).

24. M. Laskowski Jr., I. Kato, Protein inhibitors of proteinases. *Annu. Rev. Biochem.* **49**, 593-626 (1980).

25. N. D. Rawlings, Peptidase inhibitors in the MEROPS database. *Biochimie* **92**, 1463-1483 (2010).

26. A. Marrero *et al.*, The crystal structure of human α_2-macroglobulin reveals a unique molecular cage. *Angew. Chem. Int. Ed.* **51**, 3340-3344 (2012).

27. L. Sottrup-Jensen, T. E. Petersen, S. Magnusson, A thiol-ester in α_2-macroglobulin cleaved during proteinase complex formation. *FEBS Lett.* **121**, 275-279 (1980).

28. J. Travis, G. S. Salvesen, Human plasma proteinase inhibitors. *Annu. Rev. Biochem.* **52**, 655-709 (1983).

29. U. Qazi, P. G. Gettins, D. K. Strickland, J. K. Stoops, Structural details of proteinase entrapment by human a$_2$-macroglobulin emerge from three-dimensional reconstructions of Fab labeled native, half-transformed, and transformed molecules. *J. Biol. Chem.* **274**, 8137-8142 (1999).

30. P. B. Armstrong, J. P. Quigley, α_2-macroglobulin: an evolutionarily conserved arm of the innate immune system. *Dev. Comp. Immunol.* **23**, 375-390 (1999).
31. M. T. Debanne, R. Bell, J. Dolovich, Characteristics of the macrophage uptake of proteinase-α-macroglobulin complexes. *Biochim. Biophys. Acta* **428**, 466-475 (1976).
32. E. Delain *et al.*, The molecular organization of human α_2-macroglobulin. An immunoelectron microscopic study with monoclonal antibodies. *J. Biol. Chem.* **263**, 2981-2989 (1988).
33. O. M. Andersen *et al.*, Specific binding of α-macroglobulin to complement-type repeat CR4 of the low-density lipoprotein receptor-related protein. *Biochemistry* **39**, 10627-10633 (2000).
34. M. J. Imber, S. V. Pizzo, Clearance and binding of two electrophoretic "fast" forms of human α_2-macroglobulin. *J. Biol. Chem.* **256**, 8134-8139 (1981).
35. L. Jenner, L. Husted, S. Thirup, L. Sottrup-Jensen, J. Nyborg, Crystal structure of the receptor-binding domain of α_2-macroglobulin. *Structure* **6**, 595-604 (1998).
36. W. Huang, K. Dolmer, X. Liao, P. G. W. Gettins, NMR solution structure of the receptor binding domain of human α_2-macroglobulin. *J. Biol. Chem.* **275**, 1089-1094 (2000).
37. T. Xiao, D. L. DeCamp, S. R. Sprang, Structure of a rat α_1-macroglobulin receptor-binding domain dimer. *Protein Sci.* **9**, 1889-1897 (2000).
38. B. J. C. Janssen *et al.*, Structures of complement component C3 provide insights into the function and evolution of immunity. *Nature* **437**, 505-511 (2005).
39. N. Doan, P. G. W. Gettins, Human α_2-macroglobulin is composed of multiple domains, as predicted by homology with complement component C3. *Biochem. J.* **407**, 23-30 (2007).
40. G. R. Andersen, T. J. Koch, K. Dolmer, L. Sottrup-Jensen, J. Nyborg, Low resolution X-ray structure of human methylamine-treated α_2-macroglobulin. *J. Biol. Chem.* **270**, 25133-25141 (1995).
41. N. Boisset, J. C. Taveau, F. Pochon, J. Lamy, Similar architectures of native and transformed human α_2-macroglobulin suggest the transformation mechanism. *J. Biol. Chem.* **271**, 25762-25769 (1996).
42. S. J. Kolodziej, J. P. Schroeter, D. K. Strickland, J. K. Stoops, The novel three-dimensional structure of native human α_2-macroglobulin and comparisons with the structure of the methylamine derivative. *J. Struct. Biol.* **116**, 366-376 (1996).
43. S. J. Kolodziej, T. Wagenknecht, D. K. Strickland, J. K. Stoops, The three-dimensional structure of the human α_2-macroglobulin dimer reveals its structural organization in the tetrameric native and chymotrypsin α_2-macroglobulin complexes. *J. Biol. Chem.* **277**, 28031-28037 (2002).
44. S. L. Harwood *et al.*, Structural Investigations of Human A2M Identify a Hollow Native Conformation That Underlies Its Distinctive Protease-Trapping Mechanism. *Mol Cell Proteomics* **20**, 100090 (2021).
45. A. J. Barrett, M. A. Brown, C. A. Sayers, The electrophoretically 'slow' and 'fast' forms of the α_2-macroglobulin molecule. *Biochem. J.* **181**, 401-418 (1979).
46. L. Sottrup-Jensen, T. E. Petersen, S. Magnusson, Trypsin-induced activation of the thiol esters in alpha 2-macroglobulin generates a short-lived intermediate ('nascent' α_2-M) that can react rapidly to incorporate not only methylamine or putrescine but also proteins lacking proteinase activity. *FEBS Lett.* **128**, 123-126 (1981).
47. G. A. Jensen *et al.*, Binding site structure of one LRP-RAP complex: implications for a common ligand-receptor binding motif. *J. Mol. Biol.* **362**, 700-716 (2006).
48. L. Sottrup-Jensen *et al.*, Primary structure of the 'bait' region for proteinases in α_2-macroglobulin. Nature of the complex. *FEBS Lett.* **127**, 167-173 (1981).
49. T. Goulas, I. Garcia-Ferrer, S. Garcia-Pique, L. Sottrup-Jensen, F. X. Gomis-Ruth, Crystallization and preliminary X-ray diffraction analysis of eukaryotic alpha2 -macroglobulin family members modified by methylamine, proteases and glycosidases. *Mol Oral Microbiol* **29**, 354-364 (2014).
50. I. Garcia-Ferrer *et al.*, Structural and functional insights into Escherichia coli alpha2-macroglobulin endopeptidase snap-trap inhibition. *Proc Natl Acad Sci U S A* **112**, 8290-8295 (2015).
51. E. J. Cohn *et al.*, Preparation and properties of serum and plasma proteins; a system for the separation into fractions of the protein and lipoprotein components of biological tissues and fluids. *J. Am. Chem. Soc.* **68**, 459-475 (1946).

52. A. Rohou, N. Grigorieff, CTFFIND4: Fast and accurate defocus estimation from electron micrographs. *J Struct Biol* **192**, 216-221 (2015).

53. S. H. Scheres, RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J Struct Biol* **180**, 519-530 (2012).

54. R. Fernandez-Leiro, S. H. W. Scheres, A pipeline approach to single-particle processing in RELION. *Acta Crystallogr D Struct Biol* **73**, 496-502 (2017).

55. J. M. de la Rosa-Trevin *et al.*, Scipion: A software framework toward integration, reproducibility and validation in 3D electron microscopy. *J Struct Biol* **195**, 93-99 (2016).

56. S. H. Scheres, Processing of Structurally Heterogeneous Cryo-EM Data in RELION. *Methods Enzymol* **579**, 125-157 (2016).

57. J. L. Vilas *et al.*, MonoRes: automatic and accurate estimation of local resolution for electron microscopy maps. *Structure* **26**, 337-344 (2018).

58. E. Ramirez-Aportela *et al.*, Automatic local resolution-based sharpening of cryo-EM maps. *Bioinformatics* **36**, 765-772 (2020).

59. E. F. Pettersen *et al.*, UCSF Chimera - A visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605-1612 (2004).

60. P. D. Adams *et al.*, PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. sect. D* **66**, 213-221 (2010).

61. P. Emsley, K. Cowtan, COOT: model-building tools for molecular graphics. *Acta Crystallogr. sect. D* **60**, 2126-2132 (2004).

62. V. B. Chen *et al.*, MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. sect. D* **66**, 12-21 (2010).

63. W. Kabsch, XDS. *Acta Crystallogr. sect. D* **66**, 125-132 (2010).

64. W. Kabsch, Integration, scaling, space-group assignment and post-refinement. *Acta Crystallogr. sect. D* **66**, 133-144 (2010).

65. M. D. Winn *et al.*, Overview of the CCP4 suite and current developments. *Acta Crystallogr. sect. D* **67**, 235-242 (2011).

66. O. S. Smart *et al.*, Exploiting structure similarity in refinement: automated NCS and target-structure restraints in BUSTER. *Acta Crystallogr. sect. D* **68**, 368-380 (2012).

67. F. Clerc *et al.*, Human plasma protein N-glycosylation. *Glycoconj. J.* **33**, 309-343 (2016).

FIGURE LEGENDS

Figure 1. Overview of hα$_2$M action based on biochemical analysis – In native (hα$_2$M)$_4$, the four intact bait regions (black) are exposed in the enclosed cavity (prey chamber), while the thioester bonds (pink) and receptor-binding domains (RBD, grey oval) are hidden. Monomer pairs 1 and 2, as well as 3 and 4, are disulfide-linked protomers (dashed green line); pairs 1 and 3, as well as 2 and 4, are vicinal protomers; and pairs 1 and 4, as well as 2 and 3, are diagonally opposite protomers. Each protomer thus has a vicinal, a disulfide-linked, and an opposite protomer. The mechanism of action is illustrated in four steps that result in three states as described in the text.

Figure 2. Cryo-EM structures of (hα$_2$M)$_4$ functional states – (**a**) Five functional states of (hα$_2$M)$_4$ from human plasma are isolated from cryo-EM analysis, termed native I (left panel) and II (center-left), semi-activated I (center) and II (center-right), and activated (right). Protomer nomenclature [1 (red), 2 (green), 3 (yellow), and 4 (blue)] as described in Fig. 1. The red protomer has a disulfide-linked (green), a vicinal (yellow), and an opposite neighbor (blue). Native I and II states have protomers with expanded conformation and are in an equilibrium, at which vicinal dimers are in a distal (native I) or proximal (native II) position (curved arrows). After proteolytic activation, the native state becomes semi-activated states I and II, and one vicinal dimer is built of protomers in compact conformation. Semi-activated states I and II, which correspond to the nascent state described in the literature (46), evolve to the activated state, shown in "H-view", in which all protomers are compact. Openings are indicated as dashed ovals. Tethering loops of opposite protomers (1 and 4) are framed in a dashed rhombus. (**b**) Additional views of the native I (left) and II (right) complexes of panel a. The latter highlights the disulfide-linked residues between protomers 1 and 2 (red and green spheres). (**c**) The semi-activated II

complex highlights the N-linked glycosylation sites as green sticks (left). Magnified view of the red and blue boxes that correspond to the same glycan bound to MG4 Asn396 in the compact protomer 4 (blue) and in the expanded protomer 1 (red). (**d**) The activated state is shown in end-view (top) and X-view (bottom), in which the triangular prism profile for each protomer is framed.
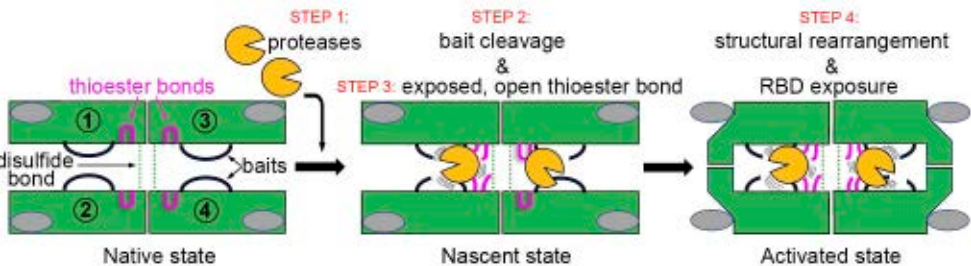
Figure 3. Functional states of (hα₂M)₄ are built of expanded and compact protomers – (**a**) Spatial domain organization of the expanded (top) and compact (bottom) protomer conformations in front view (as in Fig. 2a). The BRD (dark blue) contains the flexible, intact (top) and broken (bottom) bait region (dashed line). Insets, diagrams of the approximate domain organization of the conformers. (**b**, **c**) Regions equivalent to a, highlighting the N-terminal module (domains MG1-MG6) (**b**) and the C-terminal module (domains MG7-RBD) (**c**). 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.

Figure 4. The major players in the conformational shift: BRD, TED and RBD – (**a**) Unprocessed BRD (blue) seen from inside the prey chamber in the expanded conformation of the native I state (upper vicinal protomers 1 or 3). Regions I, II, III (dashed line) and IV are indicated. MG6 β4 (orange) is shown, bound to region IV (Glu729-Arg732). Glu729 is highlighted (grey sphere). (**b**) Close up of the unprocessed segment Glu729-Arg732 passing through the opening (dashed oval) framed by MG6 (pink), MG2 (yellow), MG3 (green) and the MG2-MG3 connecting loop (Glu729, grey sphere). (**c**) The proteolytically processed BRD seen from inside the prey chamber in the compact conformation of the activated state in any of the protomers (1-4), in which region III (dashed line) is discontinuous (similar view as panel a). The last visible residue is Thr730 (grey sphere). (**d**) Close up of the processed segment Thr730-Arg732 (blue), which interacts with the seventh β-strand of MG7 on the outer surface (Thr730, grey sphere). The opening seen in (b) is occluded. (**e**) In the expanded conformation of the native I state, the TED thioester bond (yellow) and α2 helix of the RBD (orange) are 16 Å apart, and are buried in the structure. (**f**) Close up of the region shown in (**e**). The thioester bond is in a local cavity surrounded by the large hydrophobic side chains of Tyr970, Phe1028, and Leu1087 (from TED), and His1404 (from RBD). Helix α2 from the RBD is partially hidden by segment Ser1428-Thr1432 from the domain's seventh β-strand.
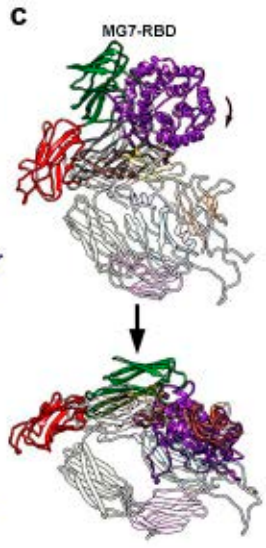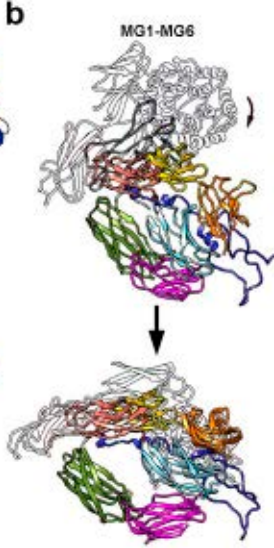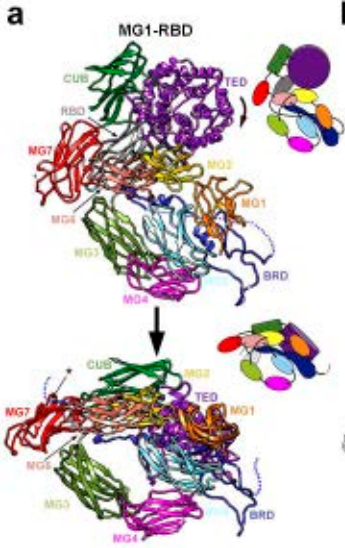
Figure 5. The symmetrically activated state – (**a**) Intrinsically activated (hα₂M)₄ from plasma. The front half of the map (transparent surface) was removed to visualize the heterogeneous proteinase density (red). (**b**) Trypsin-activated (hα₂M)₄ in a similar view as in (a), with two trypsin molecules tentatively docked into the density (dark red), which is clearly resolved in two separate volumes (clear red). However, the exact orientation of the molecules cannot be determined due to the lowish resolution of the map. (**c**) Close-up of the boxed region in (b). To highlight that the caged trypsin molecules are dynamically oriented, wavy lines are shown around the corresponding red ribbons. Location of major functional residues are indicated in the activated complex. Trypsin is fixed to the TED thioester bonds formed between Cys972 and Gln975 from vicinal protomers. Residues Gln435 of each MG4 moiety are indicated. Color code: MG2, yellow; MG3, green; MG4, pink; TED, purple; trypsin, red.
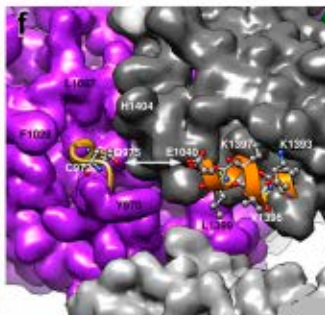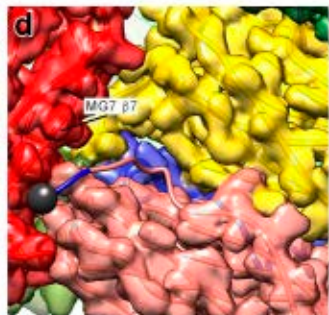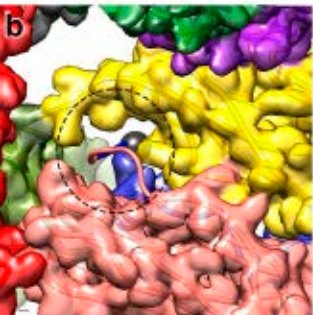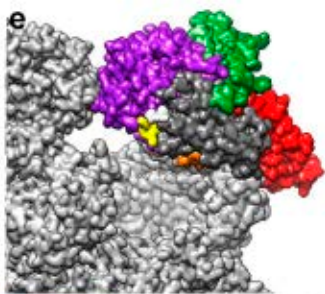
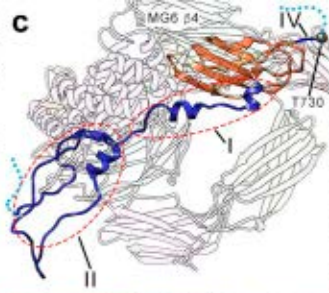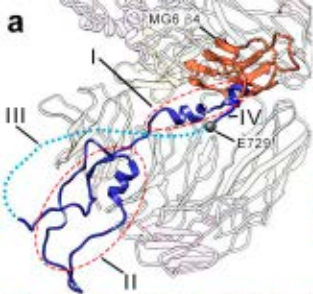**Figure 6. The subunit-mediated activated state** – (**a**) State I of plasmin-activated (hα₂M)₄ with three compact (blue) and an expanded (yellow) subunit (left). After removing the front half of the map, an asymmetric, large density corresponding to plasmin becomes evident (red) (right). (**b**) Analysis by symmetry expansion of native (hα₂M)₄ reveals arrangements with one (top, left), two (top, right and bottom, left) and three (bottom, right) compact subunits (blue).
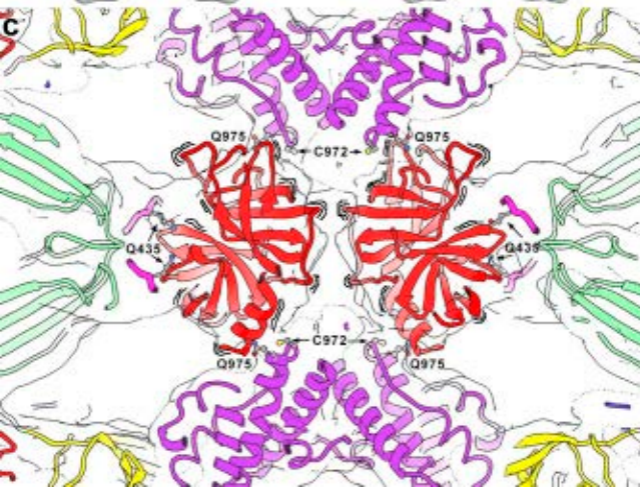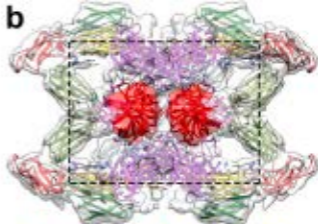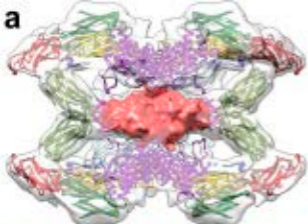
Figure 7. Mechanism of irreversible protease inhibition by (hα₂M)₄ – In the initial stage, native (hα₂M)₄ (four yellow expanded subunits) is a molecular sieve of plasma proteins (dark green, light green, and gold). Only peptidases (gold and light green) are retained in the internal cavity after processing of the bait region (red star), which can include two small peptidases (gold) or one large peptidase (light green). The thioester bond (red dot) becomes accessible to covalently fix the prey peptidase, which triggers the expanded-to-compact transition of the protomer (yellow-to-blue) and exposure of the flexible RBD (grey) on the surface, ready for cognate receptor binding (compact blue subunits). As the prey remains active within the trap, it can still cleave a second, third, and fourth protomer to yield the distinct intermediates and ultimately, the final fully compact stage 1 for one or two small peptidases. In the case of large peptidases (bottom row), peptidase regions protruding from the prey chamber prevent complete transition, giving rise to an asymmetric final stage 2 that corresponds to the penultimate step of the pathway for small peptidases.
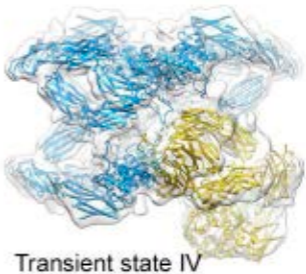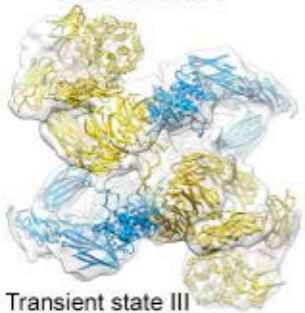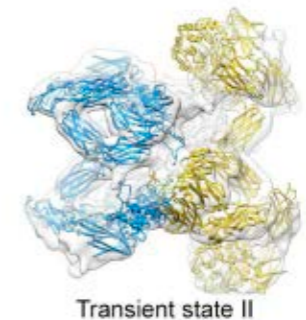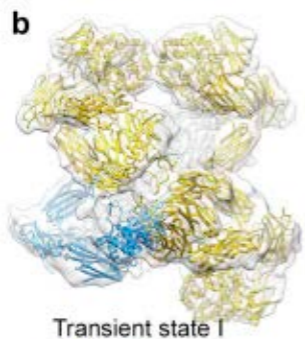
STEP 1: proteases

STEP 2: bait cleavage & STEP 3: exposed, open thioester bond

STEP 4: structural rearrangement & RBD exposure

thioester bonds

① ③

disulfide bond

baits

② ④

Native state

Nascent state

Activated state

a Native I — Native II — Semi-activated I — Semi-activated II — Fully activated

① ③ ② ④

70x50 Å
210 Å
185 Å
175 Å
15x15 Å
30x40 Å
210 Å
140 Å
210 Å

b 50° 90° 90° 90°

70x50 Å
30x20 Å

c

d 90°

a MG1-RBD

CUB
TED
RBD
MG7
MG2
MG4
MG1
MG5
MG3
BRD
MG6

CUB
MG7
MG2
TED
MG1
MG5
MG3
MG4
BRD

b MG1-MG6

c MG7-RBD

Labels visible in figure:

a
I, III, IV, MG6 β4, E729

b

c
I, II, IV, MG6 β4, T730

d
MG7 β7

e

f
F1029, L1057, H1404, C973, C975, E1040, K1397, K1393, Y975, L1390, T1385

a

b

Transient state I

Transient state II

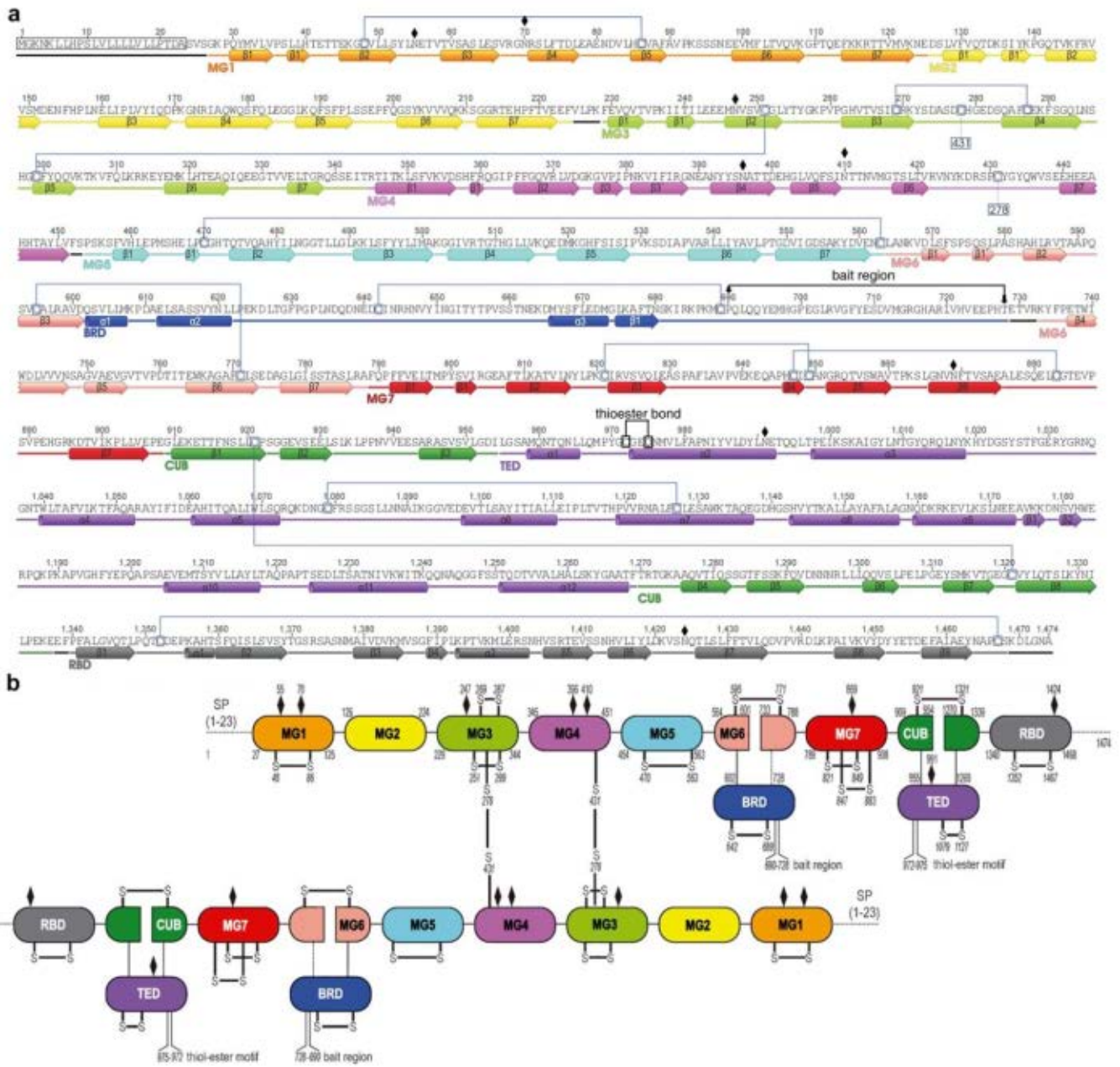Transient state III

Transient state IV

Initial stage

Final stage 1

Final stage 2

Supplemental Information

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

Daniel Luque[#], Theodoros Goulas[#], Carlos P. Mata[#], Soraia dos Reis-Mendes,

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

Supplementary Figure S1: Sequence, secondary structure elements, and domain organization of hα2M. (a) Sequence and secondary structure elements of the 1474-residue hα2M (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)**.
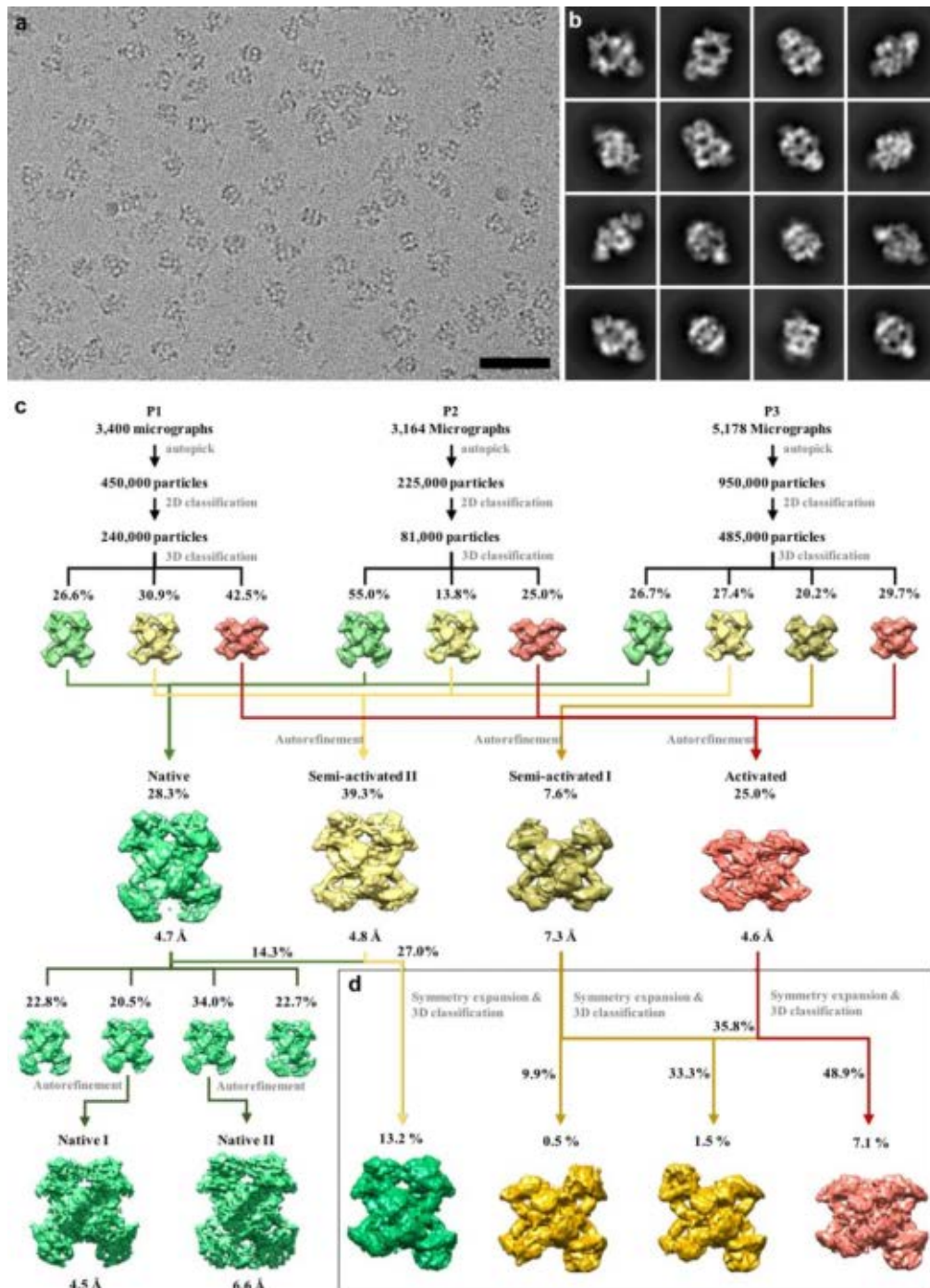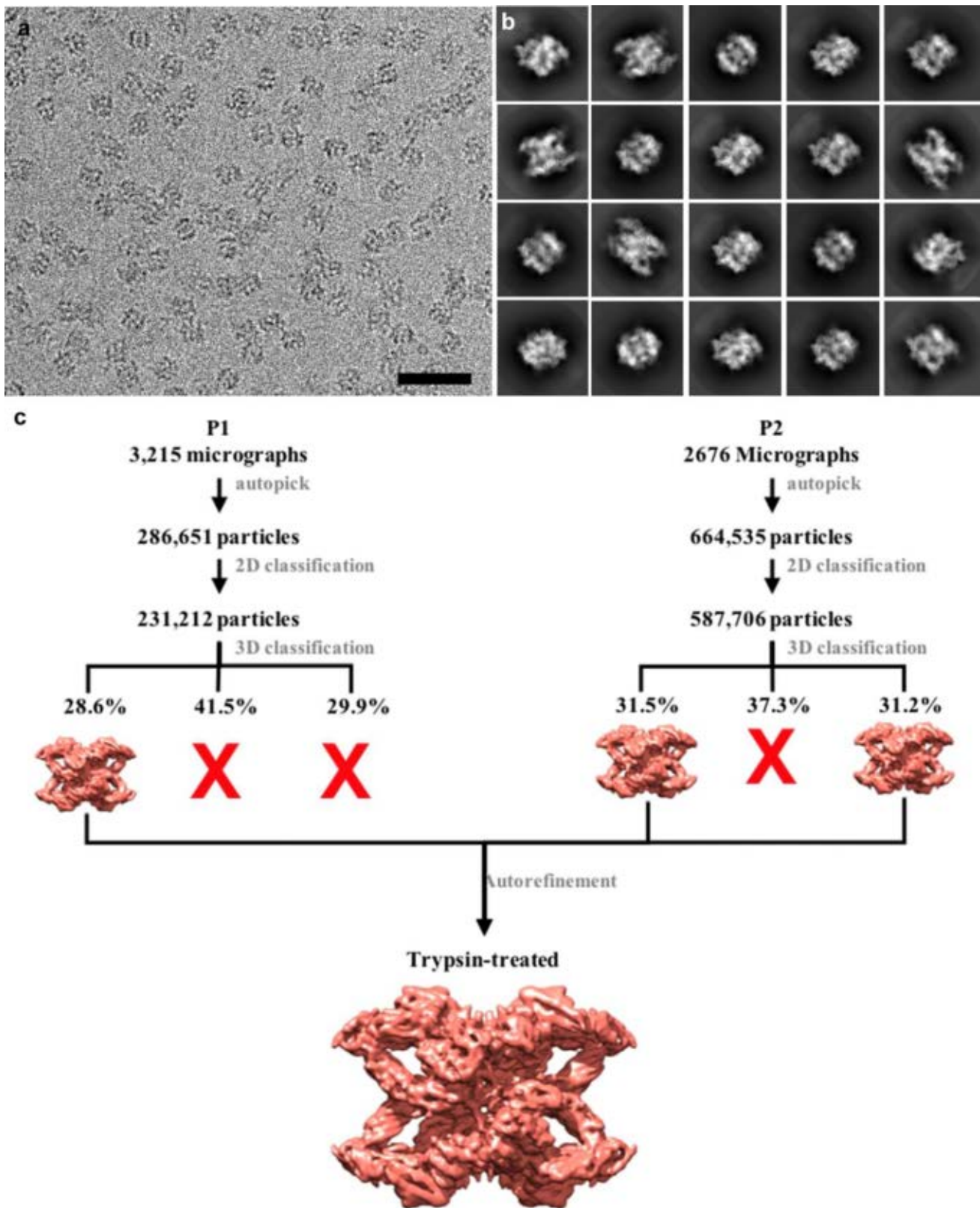
Supplementary Figure S2: Cryo-EM data processing of (hα₂M)₄ corresponding to the native untreated fraction. (**a**) Representative cryo-EM image of (hα₂M)₄ complexes (bar, 500 Å). (**b**) 2D class averages of (hα₂M)₄. (**c**) Data processing workflow and structure determination of the five functional states of (hα₂M)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 (hα₂M)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.
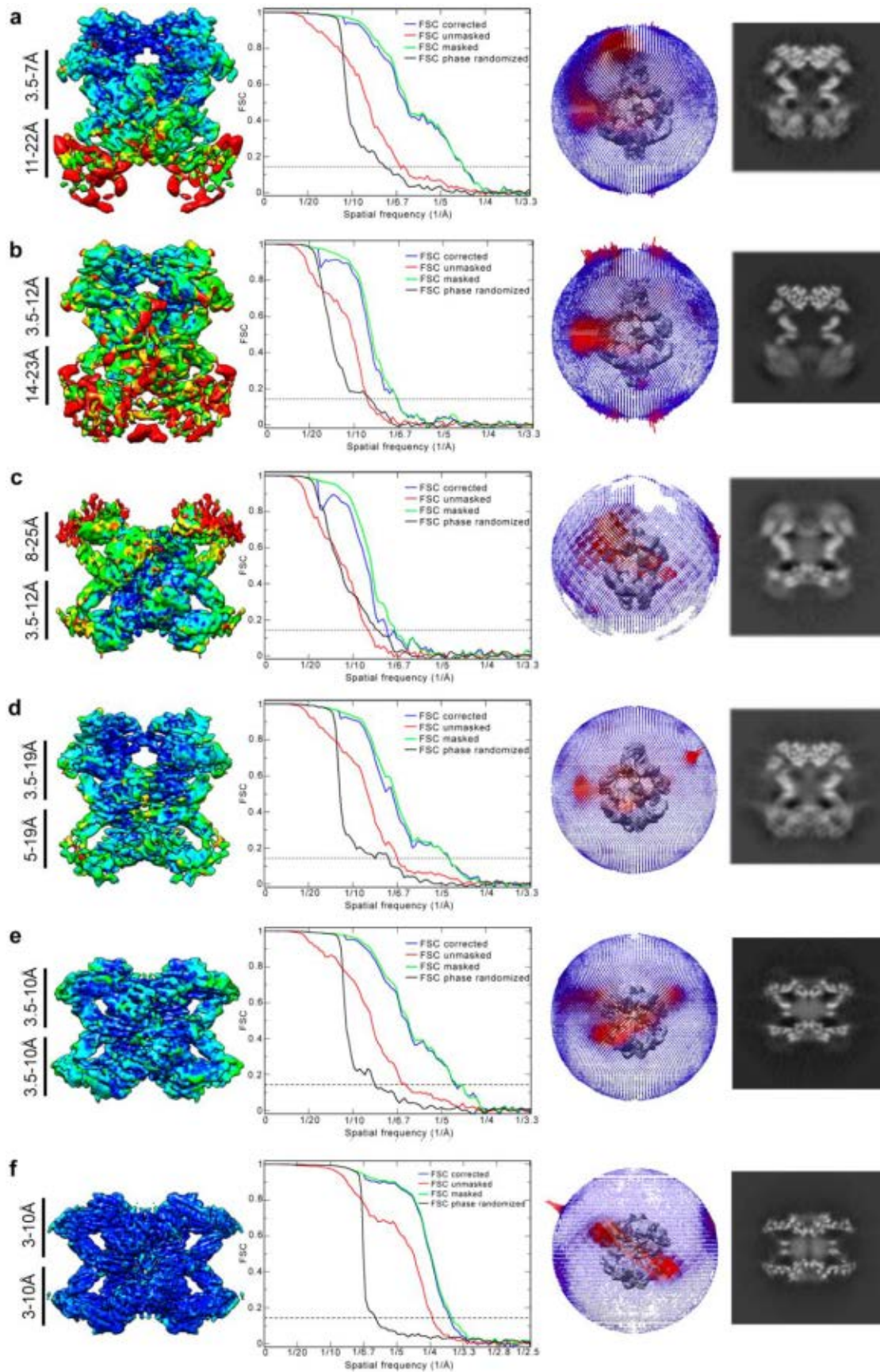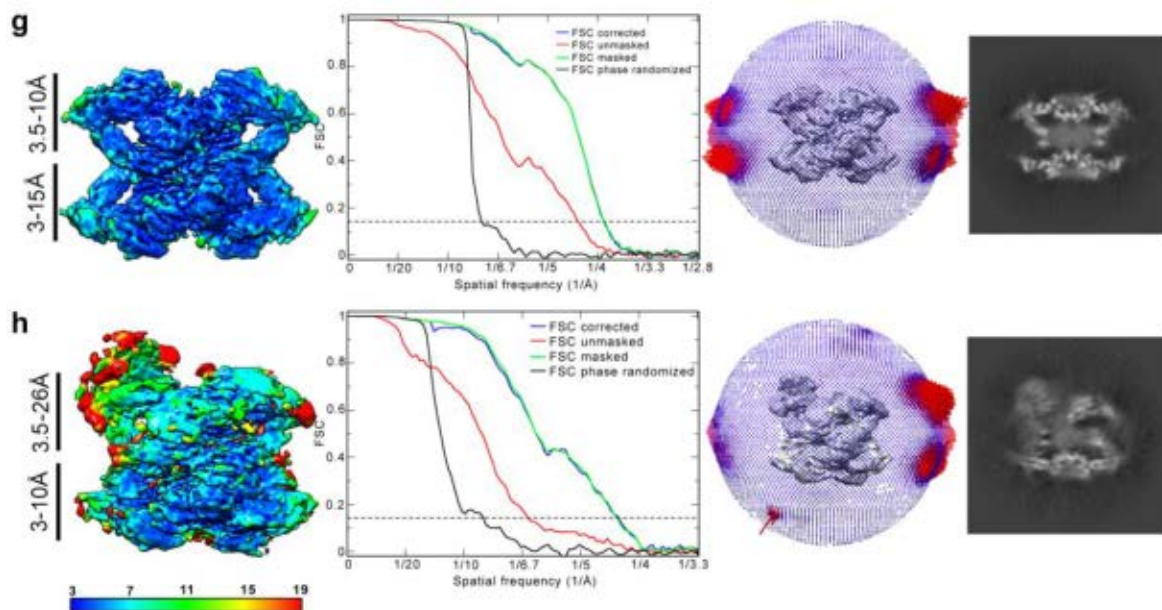
Supplementary Figure S3: Global and local resolution of cryo-EM maps. Local resolution assessment and Fourier shell correlation (FSC) curves for the eight maps calculated in this study: (**a**) native I, (**b**) native II, (**c**) semi-activated I, (**d**) semi-activated II, (**e**) fully activated state, (**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 (hα₂M)₄. (**a**) Cryo-EM image of (hα₂M)₄ complexes after trypsin treatment (bar, 500 Å). (**b**) 2D class averages of trypsin-treated (hα₂M)₄ particles. (**c**) Data processing workflow and structure determination of the trypsin-treated, activated (hα₂M)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.
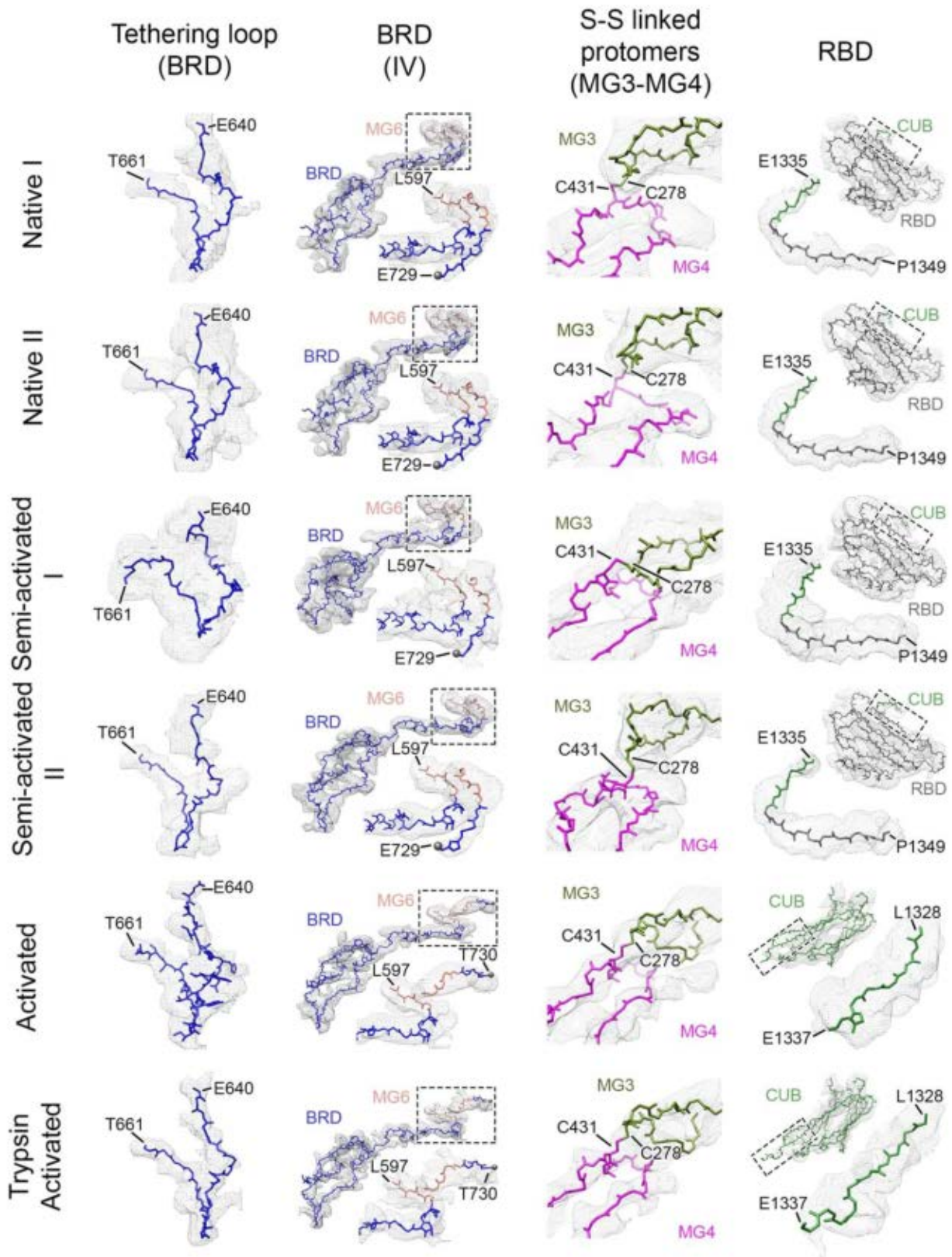
Supplementary Figure S5: Quality of the (hα₂M)₄ density maps. Cryo-EM density maps around four different regions (the tethering loop, the BRD region IV, the disulfide-linked MG3-MG4, and the RBD) of (hα₂M)₄ structures of native I, native II, semi-activated I, semi-activated II, naturally activated and 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.
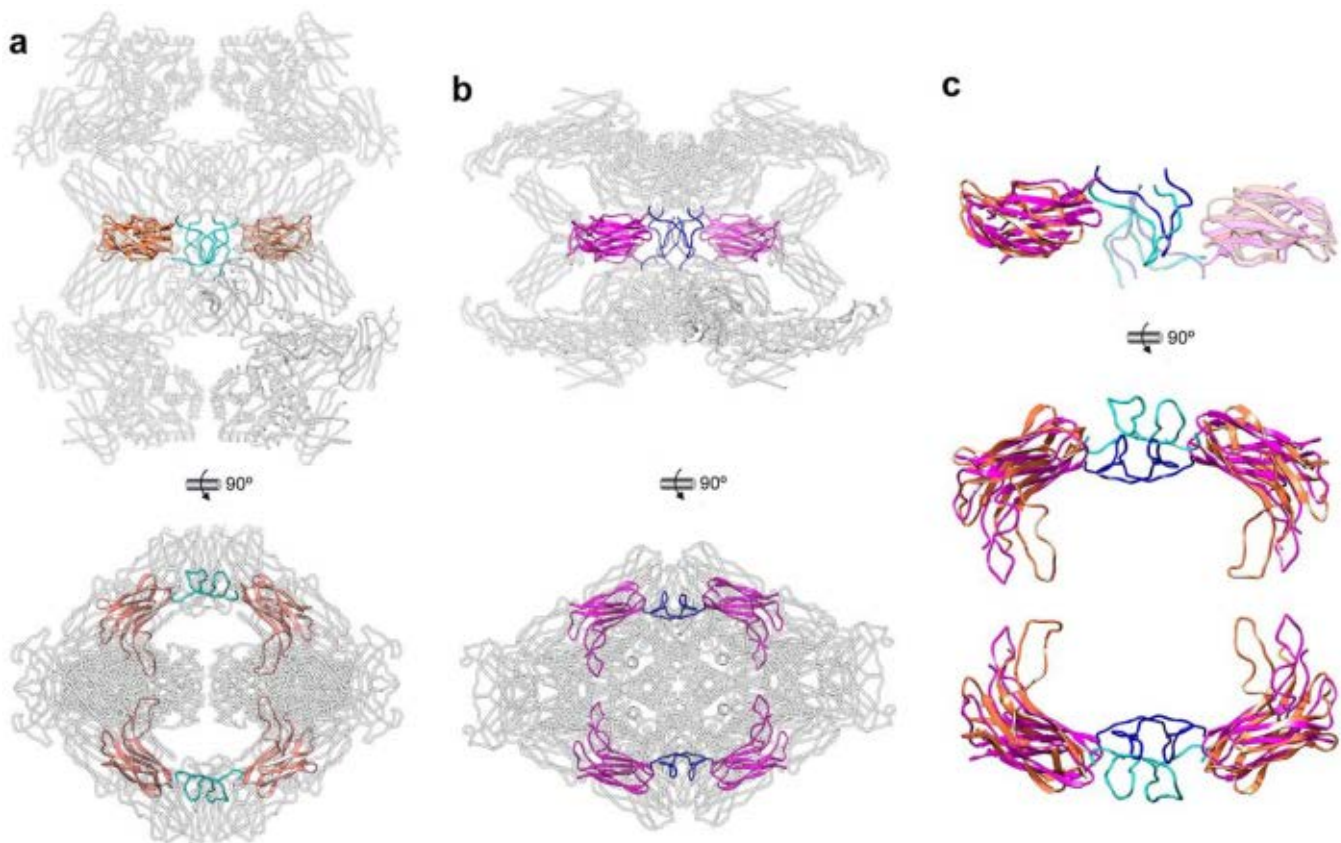
Supplementary Figure S6: Structural belt of MG4 domains and BRD tethering loops. (**a**) Two orthogonal views of the native II state in which the MG4 domains (orange) and the tethering loops (light blue) are highlighted. (**b**) Views as in (**a**) of the activated state with highlighted MG4 domains (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**).
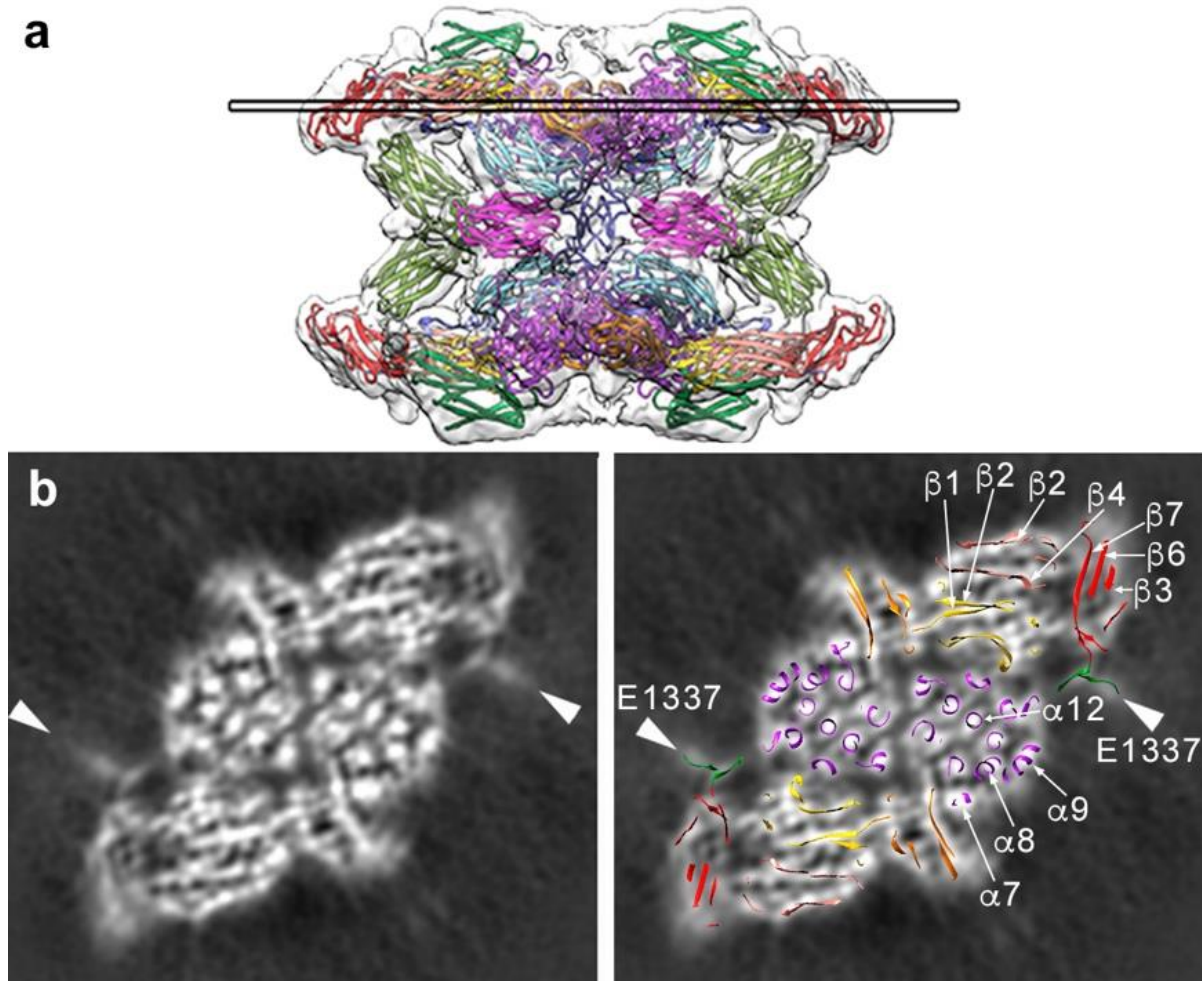
Supplementary Figure S7: The RBD is flexible in the cryo-EM density map of activated (hα2M)4. (**a**) Structure of the activated state (domain colors as in Figure S1). The thin rectangle indicates the transversal section shown in (**b**). (**b**) Section across the cryo-EM map without (left) and with atomic model regions of (hα2M)4 (right). The last visible hα2M residue in the model is Glu1337, located in the connecting loop between CUB and RBD, and indicated with two molecules (arrowheads).
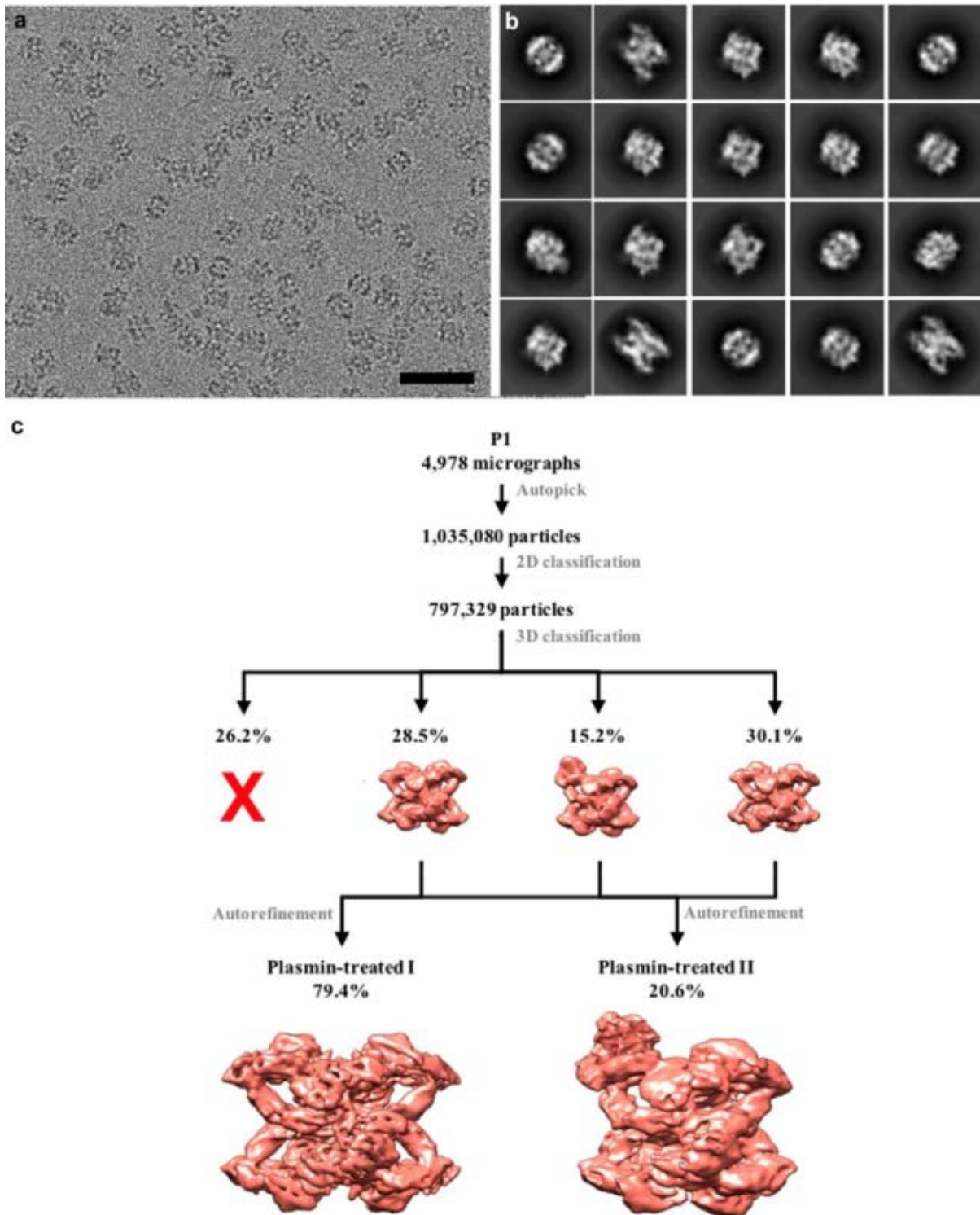
Supplementary Figure S8: Cryo-EM data processing of plasmin-treated (hα₂M)₄. (**a**) Cryo-EM image of (hα₂M)₄ complexes after plasmin treatment (bar, 500 Å). (**b**) 2D class averages of (hα₂M)₄ plasmin-treated particles. (**c**) Data processing workflow and structure determination of the plasmin-treated (hα₂M)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).
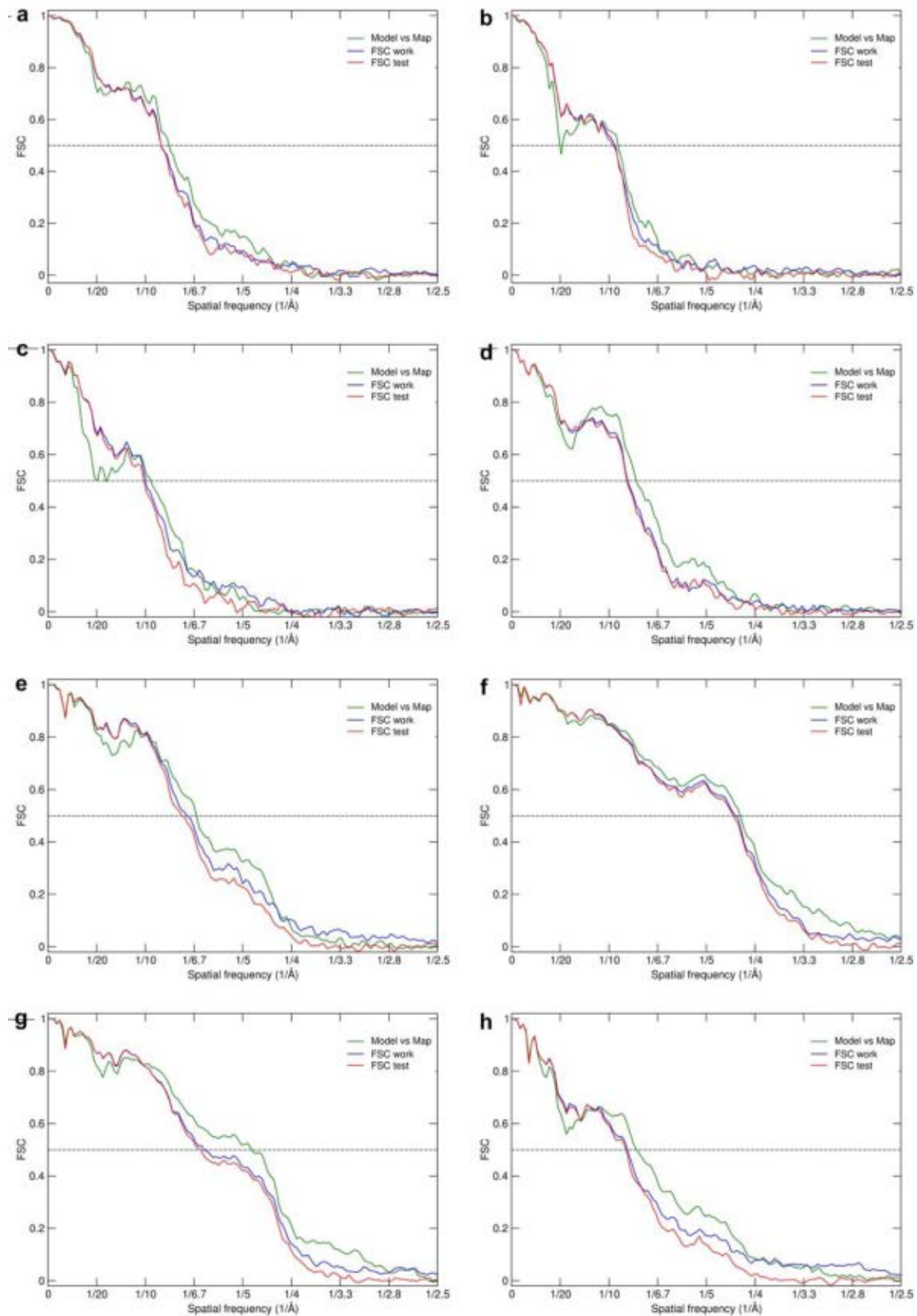
Supplementary Figure S9: Cryo-EM map quality and model validation. FSC of the refined model versus the map (green curve) and FSC work/FSC test validation curves (blue and red curves, respectively) for the eight maps calculated in this study: (**a**) native I, (**b**) native II, (**c**) semi-activated 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	Native II	Semiactivated I state	Semiactivated II state	Activated	Trypsin-activated	Plasmin-activated I state	Plasmin-activated II state
	EMD-12747 PDB: 7O7L	EMD-12748 PDB: 7O7M	EMD-12750 PDB: 7O7N	EMD-12751 PDB: 7O7O	EMD-12752 PDB: 7O7P	EMD-12753 PDB: 7O7Q	EMD-12754 PDB: 7O7R	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

	Extended subunit	Compact subunit
Supplementary Table S2. Analysis of intra-subunit domains interactions.		
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 $(h\alpha_2 M)_4$
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).