Structural and functional insights into *Escherichia coli* α2-macroglobulin endopeptidase snap-trap inhibition.

Irene Garcia-Ferrer 1, Pedro Arède 1, Josué Gómez-Blanco 2, Daniel Luque 2,3, Stephane Duquerroy 4,5, José R. Castón 2, Theodoros Goulas 1,2 and F. Xavier Gomis-Rüth 1,2

1 Proteolysis Lab; Department of Structural Biology (“Maria de Maeztu” Unit of Excellence); Molecular Biology Institute of Barcelona (CSIC); Barcelona Science Park; c/Baldiri Reixac, 15-21; 08028 Barcelona (Spain). 2 Department of Structure of Macromolecules; Centro Nacional de Biotecnología (CSIC); Cantoblanco; c/Donavín, 3; 28049 Madrid (Spain). 3 Spanish National Microbiology Centre; Institute of Health Carlos III; Carretera Majadahonda-Pozuelo, km. 2.2. Majadahonda; 28220 Madrid (Spain). 4 Unité de Virologie Structurale; Département de Virologie et CNRS Unité de Recherche Associee (URA) 3015; Institut Pasteur; 25 Rue du Dr Roux; 75724 Paris Cedex 15 (France). 5 Département de Biologie; Faculté des Sciences d’Orsay; Université Paris-Sud; 91405 Orsay (France).

Submitted to Proceedings of the National Academy of Sciences of the United States of America

The survival of commensal bacteria requires them to evade host peptidases. Gram-negative bacteria from the human gut microbiome encode a relative of the human endopeptidase inhibitor, α2-macroglobulin (α2M). *Escherichia coli* α2M (ECAM) is a ~180-kDa multi-domain membrane-anchored pan-peptidase inhibitor, which is cleaved by host endopeptidases in an accessible bait region. Structural studies by electron microscopy and crystallography reveal that this cleavage causes major structural rearrangement of over half the 13-domain structure from a native to a compact induced form. It also exposes a reactive thioester bond, which covalently traps the peptidase. Subsequently, peptidase-laden ECAM is shed from the membrane and may dimerize. Trapped peptidases are still active except against very large substrates, so inhibition potentially prevents damage of large cell-envelope components but not host digestion. Mechanistically, these results document a novel monomeric “snap trap.”

The human microbiome plays a crucial role in host health and disease (1). Successful commensalism requires microorganisms to neutralize damaging host factors but the mechanisms to maintain symbiosis are only poorly understood (2). In particular, their habitat is rich in host proteolytic enzymes, which are generally held in check by protein inhibitors (3). Several Gram-negative proteolysins—including human pathogens—contain genes similar to the widespread metazoan α2-macroglobulins (α2Ms) (4). These are large multidomain glycoproteins that uniquely function as broad-spectrum endopeptidase inhibitors and mostly contain a reactive β-cysteinyl-g-glutamyl thioester bond (5). The potential bacterial α2Ms (hα2Ms) occur in two independent forms: one is provided with a thioester bond (represented by *Escherichia coli* α2M; ECAM) and co-transcribed with penicillin-binding protein 1C; and the other lacks a thioester bond and is transcribed from an operon further encoding other proteins (represented by *E. coli* YfαS).

In humans, α2M (hα2M) circulates mostly in blood plasma as an abundant “native” ~720-kDa tetramer. After cleavage in a “bait region” (6), the tetramer closes under large conformational rearrangement to yield an “induced” form, which encages the peptidase following an irreversible “Venus flytrap” mechanism (5, 7, 8). Inside the cage, within a large “central prey chamber,” peptidases still cleave small-to-medium substrates (<10-20kDa (9)), which enter the tetramer through any of 16 entrances (8), but not large substrates. In some cases, prey lysines may be covalently bound through the thioester bond of mammalian α2Ms. However, other α2M-family inhibitors such as ovostatins lack thioester bonds and only encage, but they are as efficient inhibitors as hα2M (10). Induction of tetrameric hα2M exposes C-terminal receptor-binding domains (RBDS), which are bound by specific cell-surface receptors. This exposure triggers receptor-mediated endocytosis and clearance of the inhibitor and its prey from the circulation (11). For successful encaging, at least two protomers are required to wrap around a standard-size endopeptidase (12), but the detailed molecular mechanism of tetrameric α2M inhibition is unknown as only the molecular structure of induced hα2M is available (8). Little is also known yet about the physiology and function of hα2Ms, as only a YfαS-ortholog from *P. aeruginosa* and ECAM have been partially studied to date (13-16). The crystal structure of native α2M from *Salmonella enterica* (SEAM) is available (16), but its working mechanism is also unknown so far.

To shed light on the structure and function of α2Ms, we studied ECAM functionally, biophysically and structurally by X-ray crystallography and cryo-electron microscopy (cryo-EM). We found that cleavage at the bait region of ECAM triggers major conformational rearrangement and covalent binding of the prey peptidase following a novel monomeric snap-trap mechanism, which differs from the encaging Venus flytrap mechanism of tetrameric hα2M.

Results and Discussion

Native and induced forms of ECAM — In thioester proteins in general, the reactive thioester bond is protected in native forms to prevent precocious opening (17). In hα2M, treatment with small proteolytic enzymes are inhibited in vivo by protein inhibitors. Such inhibitors are employed by symbiotic bacteria in our gut to protect themselves from digestive peptidases. This is the case for *Escherichia coli*, which has acquired a large multidomain inhibitor of broad inhibitory spectrum (ECAM). We studied ECAM and found it is cleaved by host peptidases, which triggers a large conformational rearrangement of the inhibitor—shown by protein crystallography and electron microscopy reconstructions—and covalent binding of the peptidase. The latter is inhibited like a prey by a snap trap, which prevents damage to the bacterial envelope. Prey peptidases, however, are still active in the digestion of intact proteins.

Significance

Reserved for Publication Footnotes

www.pnas.org ——— PNAS | Issue Date | Volume | Issue Number | 1–77
nucleophiles such as methyamine (MA) opens the thioester bond and rearranges the tetramer. This rearrangement is equivalent to the status induced by prey peptidases (18); following which the thioester loop is exposed on the inner protein surface of the cage and becomes accessible to surface lysines of the prey (7, 8). Native and peptidase- or MA-induced hτ2M differ in their biophysical properties (19). In contrast, recombinant native ECAM (nECAM) and MA-treated ECAM (MA-ECAM) were equivalent in size-exclusion chromatography (SEC), native PAGE, thermofluor assays, and circular dichroism spectroscopy (CD), and they were both monomeric (SI Appendix, Fig. S1A-D). This indicates that MA-treatment of nECAM, which opens the thioester bond as shown by the emergence of free cysteines (SI Appendix, Table S2), does not produce an induced species. This, in turn, is consistent with the structural equivalence of native and MA-treated SEAM (16). Induced ECAM (iECAM) was obtained only by incubation with endopeptidases such as the physiologically-relevant mammalian host digestive enzyme trypsin (20). Unlike nECAM, iECAM formed monomers and dimers, both diverging from nECAM in SEC, native PAGE, thermofluor assays, and CD, which is consistent with conformational rearrangement upon induction (SI Appendix, Fig. S1A-D,J). MA-ECAM, in turn, was transformed into an induced form similar to iECAM after cleavage in the bait region. Other endopeptidases capable of ECAM cleavage were thermolysin, chymotrypsin and subtilisin (SI Appendix, Fig. S1E).

Analysis of trypsin-induced iECAM by denaturing SDS-PAGE and N-terminal Edman degradation showed that induction entailed cleavage at R^162^-D^163 (numbering follows UniProt [UP] P76578), which falls into the “bait-region domain” (BRD; see Fig. 1A for domain organization and acronyms) of the inhibitor. Further cleavages contributed to a complex band pattern (SI Appendix, Fig. S1F), which occasionally gave rise to high-molecular mass products of proteinase-generated fragments, as previously described for hτ2M (7). To simplify the picture, we produced an ECAM mutant (TEV-ECAM), in which bait-region positions R^162 and D^163 were replaced with a recognition sequence for tobacco-etch virus (TEV) peptidase. This protein was cleaved by TEV peptidase at a single site (SI Appendix, Fig. S2A) and gave rise to an induced form (TEV-iECAM; SI Appendix, Fig. S2B,C). We also observed cleavage in the bait region of ECAM by uylisin, chymotrypsin, pancreatic elastase, subtilisin and thermolysin (SI Appendix, Fig. S2F). In all cases, this cleavage was efficient and showed that the bait region contains accessible recognition sites for peptidases with distinct substrate specificities despite being shorter than in hτ2M (~25 residues, Q^{184/G^193} [see below] vs. 39 residues, see (6)). These results indicate that ECAM is a pan-proteinase target protein and that cleavage in the bait region is required and suffices to generate iECAM.

**Fig. 1. — Monomeric iECAM structure.** (A) Domain organization of ECAM with spanning residues (see also SI Appendix, Fig. S7). SP, signal peptide; A, flexible segment; MG0-MG7, macroglobulin domains 0-to-7; NIE, N-terminal domain of induced ECAM; L, linker; BRD, bait-region domain; GUB, domain type described in proteins C1r/C1s, Uegf, Bmp1; TED, thioester domain; and RBD, receptor-binding domain. The bait region and the thioester segment are depicted as green and purple handles, respectively. Red arrows point to the primary trypsin cleavage sites for induction (R^162^-D^163) and solubilization/dimerization (R^162^-G). Domain naming is based on hτ2M, which lacks A, MG0 and NIE (B). (C) Scheme of iECAM in front view. (C) Ribbon-plot of monomeric iECAM in front, lateral, and back views. In (B) and (C), the domain colors are as in (A) and the visible polypeptide chain ends upstream (u) and downstream (d) of the cleaved bait region, as well as the thioester (cyan ellipse; blown up in the inset in (C)), are indicated.

**Induced ECAM is released as monomers and dimers.** Among the trypsin cleavage sites of ECAM was also R^182^-D^183, which falls between the first two N-terminal domains, MG0 and NIE (see Fig. 1A). This linker was likewise targeted by pancreatic elastase and thermolysin (SI Appendix, Fig. S1G,H). As the N-terminal flexible segment A (Fig. 1A) of ECAM is anchored to the periplasmic side of the inner membrane through a "lipoxin"-mediated lipidic linkage of the N-terminal cysteine residue of the secreted protein (C^14^-C^17) through post-translational modification, cleavage at MG0-NIE removes the membrane anchor, thus yielding soluble iECAM. This strongly suggests shedding is a relevant step of the working mechanism of ECAM after induction. Noteworthy, this cleavage was also responsible for freshly purified monomeric iECAM being slowly transformed by trypsin to a non-covalent dimer. Evidence for this came from ECAM mutant R^162,G, which was not cleaved by trypsin at MG0-NIE and remained largely monomeric even after extended incubation periods (SI Appendix, Fig. S1G). In addition, TEV peptidase, which transformed native TEV-ECAM into an induced form (see above), did not cleave at MG0-NIE nor did it form dimers (SI Appendix, Fig. S2B). Extended dimerization was mainly observed at high trypsin:nECAM ratios or at prolonged incubation times (SI Appendix, Fig. S1I). Dimers of iECAM were stable and separable from monomers, did not revert to monomers under any condition assayed, and were unaffected by high salt, detergents or reducing agents. In addition, dimers were conformationally equivalent to iECAM monomers in CD and thermofluor assays (SI Appendix, Fig. S1C,D), and in cryo-EM reconstructions (see below), which supports that once induction has occurred on monomeric ECAM, dimerization just entails association of two pre-formed moieties. We conclude that shedding under removal of the first ~140 residues of secreted ECAM (segment A and domain MG0, Fig. 1A) after its induction is required for dimerization by trypsin and uylisin. However, other peptidases targeting the MG0-NIE linker such as pancreatic elastase and thermolysin did not significantly form dimers (SI Appendix, Fig. S1G). Thus, iECAM dimerization is restricted to treatment with particular peptidases, possibly depending on their size and shape.
ECAM inhibits cleavage of very large substrates and cell-wall components — We assessed the inhibitory activity of ECAM in vitro against a cohort of model serine- and metallopeptidases of differing specificity in the presence of a wide range of substrates (see SI Appendix, Supplementary Experimental Procedures [SEP] section §1.2). We did not detect inhibition against low- or medium-molecular mass substrates. In contrast, proteolytic activity was inhibited for trypsin against thyroglobulin (660kDa) and aldolase (160kDa), for chymotrypsin against thyroglobulin, and for subtilisin against thyroglobulin and fumarase (200kDa), indistinguishably by both monomeric and dimeric iECAM (SI Appendix, Fig. S3A-C). These experiments were complemented with assays against cell-envelope extracts prepared from E. coli K12 cells (SI Appendix, SEP §1.3), which are rapidly processed by endopeptidases that separate the outer membrane and the peptidoglycan. These experiments revealed that digestion by trypsin, chymotrypsin and subtilisin was inhibited by ECAM in a concentration-dependent manner (SI Appendix, Fig. S3D). In addition, we also found that ECAM is a cell-wall protector in vivo, as its absence results in diminished cell viability in the presence of host peptidases (see SI Appendix, Suppl. Results and Discussion [SRD] §2.1). Accordingly, ECAM inhibits proteolysis of large globular proteins and proteins embedded in the cell envelope, but not of isolated peptides and medium-to-large-sized proteins.

ECAM is a pan-peptidase covalent inhibitor — We also found that the inhibitory mechanism of ECAM further requires covalent bonding of prey peptidases by means of an intact thioester bond targeted by a lysine from the prey. Covalent linkage was shown in a zymogram, which yielded ECAM cleavage fragments showing tryptic activity against casein, similarly for monomeric and dimeric trypsin-induced iECAM (SI Appendix, Fig. S2D). In addition, purification of trypsin-treated monomeric and dimeric iECAM by SEC revealed presence of the peptidase in the elution peaks of the latter, as shown by activity against small fluorescent trypsin substrates and peptide-mass fingerprinting. Evidence for the involvement of the thioester bond came from the finding that iECAM showed free cysteines when compared with trypsin-untreated nECAM (SI Appendix, Table S2), i.e. the bond had been broken during induction. To verify that lysines were the targets of the thioester bond, we probed lysine-methylated TEV peptidase and found that, in contrast to the untreated protein, the enzyme was not bound by TEV-ECAM upon cleavage induction (SI Appendix, Fig. S2C).

In contrast, incubation of MA-ECAM—with a broken thioester bond—with trypsin underwent cleavage in the bait region and the MG0-NIE linker (SI Appendix, Fig. S1J) and was induced, but it did not bind the peptidase as shown by lack of peptidolytic activity of SEC-purified peptidase-induced MA-ECAM and absence of cross-linked peptidases in denaturing SDS-PAGE (SI Appendix, Fig. S1J). In addition, peptidase activity against cell-envelope extracts was inhibited by nECAM (see above) but not by MA-ECAM (SI Appendix, Fig. S3D).

Finally, binding of trypsin through ECAM was semi-quantitatively assessed by using fluorogenic methylcoumarin-labeled peptidase. This revealed that one peptidase molecule was covalently bound by between three and four iECAM molecules on average, i.e. thioester-mediated prey binding is relatively inefficient because a surface lysine needs to be close to the thioester bond to be bound when the bait region is cleaved. This contrasts with the high efficacy of bait-region cleavage (see above), and indicates that iECAM can be either bound or unbound to the peptidase.

Crystal structure of trypsin-induced iECAM — To shed light on the structural basis of the molecular mechanism of ECAM, we crystallized and solved the structure of trypsin-induced iECAM by four-wavelength anomalousdiffraction with a selenomethionine derivative of the protein and a dataset to high resolution from wild-type protein (see SI Appendix, SEP §1.10 and §1.11.
and Tables S3 and S4). Although the peptidase was inside the crystals—as revealed by fluorescence microscopy (SI Appendix, Fig. S4)—we could not localize it due to the low binding efficiency of iECAM (see previous section) and to intrinsic disorder in its overall conformation. The iECAM oligomeric structure in the crystals is a dimer formed by a crystallographic dyad (for a detailed description of the dimer, see SI Appendix, SRD §2.2 and Fig. S5). The iECAM monomeric structure includes fragment P166–P165, which is organized in 12 domains and a linker region (NIE to RBD; Fig. 1A,B). The molecule is arranged as an elliptical grommet with a ~105-Å major axis and a ~60-Å minor axis (Fig. 1C, center). A large hook protrudes ~80 Å from one of its major-axis vertices and is inclined ~30° towards the center of the ellipse. We distinguish between a front convex face (Fig. 1C, left; reference orientation hereafter) and a back concave face (Fig. 1C, right). The polypeptide starts at the bottom with domain NIE, which features one of the ellipse major-axis ends. Thereafter, six macroglobulin (MG) domains (MG1-MG6) are arranged as a one-and-a-half-turn superhelix (MG-superhelix) around a central lumen of ~20-Å diameter (“entrance 1”) in such a way that domains MG5 and MG6 are, respectively, aligned and in contact with MG1 and MG2. Perpendicularly attached to MG3 and MG6, domain MG7 features the opposite end of the ellipse and leads to the hook, which includes domains CUB, TED and RBD. Overall, iECAM includes six structurally different domain types: MGs, NIE, CUB, TED, RBD, and BRD (see SI Appendix, Fig. S6A-E).

MG domains are fibronectin-type-III-like β-sandwiches comprising a three- and a four-stranded antiparallel β-sheet, whose planes are rotated away by ~40° (SI Appendix, Fig. S6A; for assignment of secondary structure elements, see SI Appendix, Fig. S7). Into this basic scaffold, additional elements are inserted, which cause the eight MG domains (including MG0, see below) to span between 78 and 128 residues and vary in domain length (along the sheets; SI Appendix, Fig. S6A,F,G) between ~30 Å (MG1) and ~50 Å (MG7). Domain NIE is a variant of an MG domain, into which an extra short strand has been inserted between NIE-β6 and -β7, which interacts with NIE-β1 (SI Appendix, Fig. S6E). This entails that while the four-stranded β-sheet overlaps with that of MG1 (SI Appendix, Fig. S6I), the three-stranded back sheet is rotated and translated, thus causing the planes of the two NIE sheets to intersect at an angle of ~70° on the right lateral face, while the opposite lateral face opens. In addition, two helices are inserted in the segment connecting strands β3 and β4 of domain NIE (NIE-β3–β4).

The CUB domain is a β-sandwich of two parallel four-stranded antiparallel β-sheets (I and II), which is unrelated to the MG fold (Fig. S6D). A short helix is inserted at CUB-β6–β7, as is domain TED at CUB-β3–β4. The TED domain, in turn, is a six-fold α/α-toroid made up by six α-hairpins that resides on the outer surface of CUB sheet I and whose central axis is rotated ~45° away from the sheet planes of the CUB β-sandwich. The arrangement of the α-hairpins is clockwise when viewed from the entry surface of the toroid (Fig. S5C). The thioester segment is a fifteen-atom thiocysteine ring composed of four residues,
C118→L118, E1189→Q1190. It is located at the beginning of the first 
toroid helix, TED-α2, on the domain entry face and, compatible 
with an induced peptide-bound inhibitor, the thioster bond is 
broken (Fig. 1C, right). This segment is shielded by TED-α4−α5. 
However, while the side chain of C118 is surrounded by the side 
chains of E1189, L1242, and W1245, Q1190 points to the bulk solvent, 
which is consistent with a disordered trypsin molecule bound to 
side chain in the crystal structure.

The C-terminal domain of ECAM, RBD (name based on its 
structural similarity with hо-M RBD, see (8)), occupies a key 
position in iECAM and interacts with TED, CUB and MG7 (Fig. 
1C). In addition, it stabilizes the hook structure protruding from 
the MG-superhelix by interacting with MG2 and MG3. RBD has 
a complex topology (SI Appendix, Fig. S6B,11) and consists of 
a central MG2 core expanded to a six-stranded front and a five-
stranded back β-sheet, whose planes are rotated away by ∼40° as 
in MG domains.

The BRD is inserted at MG6−β3−β4, spans 66 residues (S901− 
N966) and is folded irregularly. It plays an important role not only 
in triggering the conformational rearrangement when cleaved 
but also in the stability of nECAM. A mutant, in which BRD 
was replaced by three glycines (protein ECAMΔBRD, see 
SI Appendix, Table S1) yielded poorly folded protein but was 
completely digested under conditions that only produced stable 
induced protein for the wild type. BRD is defined for S901−G929 (upstream 
of the cleavage site) and G940−N966 (downstream of the 
cleavage site) in the crystal structure due to trypsin cleavage 
after R946. The upstream segment of BRD is freely accessible: it 
lines part of the concave surface of the monomer and contains 
two helices. It interacts with MG2, the segment linking MG2 and 
MG3, MG6, and RBD. After the second helix, the BRD chain 
runs in extended conformation along the inner MG-superhelix 
surface and, between O921 and O931, the polypeptide is trapped 
between MG5, L, NIE, MG4, and MG1, with BRD segment A920− 
I931 performing a β-ribbon interaction with MG1−β1. The last 
upstream-segment residue defined in the structure, G949, emerges 
on the lower left outer surface of the monomer (U° in Fig. 1C).
The downstream segment of BRD, in turn, is defined from G939 onwards (D° in Fig. 1C), at the interface between MG2 and CUB. 
It encompasses a short helix, BRD-α3, and runs upwards, mainly 
interacting with MG2, MG7, CUB, RBD, and the MG7-CUB and 
CUB-RBD linkers, prior to joining MG6.

**Single-particle cryo-electron microscopy and homology modeling of native ECAM**—In order to complement the 
aforementioned crystal structure of iECAM, monomeric and 
dimeric trypsin-induced iECAM were further analyzed by 
three-dimensional cryo-EM reconstructions of single particles 
(Fig. 2A,B). Therefore, purified proteins were applied to carbon-
coated grids, and blotted and plunged into liquid ethane. Images 
were recorded on a CCD camera at low-dose conditions, with a 
200kV electron microscope equipped with a field emission gun.

Images were classified using a reference-free clustering approach 
to select homogeneous populations of 18,346 and 33,536 particles 
for monomeric iECAM and dimeric iECAM, respectively, which 
were used for reconstruction. The final resolution of the models 
was estimated by the Fourier shell correlation criterion between 
independent half-dataset maps applying a correlation limit of 
0.5 to be, respectively, 17 and 14 Å (SI Appendix, Fig. S8). The 
crystallographic coordinates of monomeric and dimeric iECAM 
(see SI Appendix, SRD §2.2 and Fig. S5) were adequately fitted 
to the corresponding cryo-EM maps, and the concordance 
of both structures is clear except for the slight deviation from 
the C2 symmetry of the dimeric iECAM cryo-EM map (Fig. 
2A,B). Although dimerization surfaces are probably flexible in 
vivo, such dimerization does not lead to major conformational 
rearrangement of a monomer once induced. Fostered by the 
agreement between the cryo-EM and X-ray structures, we 

determined a cryo-EM reconstruction for nECAM at 16 Å based on 
46,842 particles (Fig. 2C), as crystallization of the 
full-length protein produced only poorly diffraing crystals. To 
gain additional insight into the structure of nECAM at atomic 
resolution, we assayed several constructs and managed to solve 
the crystal structure of three of them, respectively spanning domains 
MG0-NIE-MG1, NIE-MG1, and MG7-CUB(TED)-RBD alias 
nECAMΔN. The first structure was solved by multi-wavelength 
adiabatic dispersion/multiple isomorphous replacement 
methodology and a MAD experiment (peak, inflection point and high-energy 
peak) to provide a high-resolution native crystal and a dataset obtained from wild-type protein. The other 
two structures were solved by single isomorphous replacement 
methodology and a MAD experiment (peak, inflection point and high-energy 
peak) to provide a high-resolution native crystal and a dataset obtained from wild-type protein. The other 
two structures were solved by single isomorphous replacement 
methodology and a MAD experiment (peak, inflection point and high-energy 
peak) to provide a high-resolution native crystal and a dataset obtained from wild-type protein. The other 
two structures were solved by single isomorphous replacement 
methodology and a MAD experiment (peak, inflection point and high-energy 
peak) to provide a high-resolution native crystal and a dataset obtained from wild-type protein. The other 
two structures were solved by single isomorphous replacement 
methodology and a MAD experiment (peak, inflection point and high-energy 
peak) to provide a high-resolution native crystal and a dataset obtained from wild-type protein. The other 
two structures were solved by single isomorphous replacement 
methodology and a MAD experiment (peak, inflection point and high-energy 
peak) to provide a high-resolution native crystal and a dataset obtained from wild-type protein. The other 
two structures were solved by single isomorphous replacement 
methodology and a MAD experiment (peak, inflection point and high-energy 
peak) to provide a high-resolution native crystal and a dataset obtained from wild-type protein. The other 
two structures were solved by single isomorphous replacement 
methodology and a MAD experiment (peak, inflection point and high-energy 
peak) to provide a high-resolution native crystal and a dataset obtained from wild-type protein. The other 
two structures were solved by single isomorphous replacement 
methodology and a MAD experiment (peak, inflection point and high-energy 
peak) to provide a high-resolution native crystal and a dataset obtained from wild-type protein. The other 
two structures were solved by single isomorphous replacement 
methodology and a MAD experiment (peak, inflection point and high-energy 
peak) to provide a high-resolution native crystal and a dataset obtained from wild-type protein. The other 
two structures were solved by single isomorphous replacement 
methodology and a MAD experiment (peak, inflection point and high-energy 
peak) to provide a high-resolution native crystal and a dataset obtained from wild-type protein. The other 
two structures were solved by single isomorphous replacement 
methodology and a MAD experiment (peak, inflection point and high-energy 
peak) to provide a high-resolution native crystal and a dataset obtained from wild-type protein. The other 
two structures were solved by single isomorphous replacement 
methodology and a MAD experiment (peak, inflection point and high-energy 
peak) to provide a high-resolution native crystal and a dataset obtained from wild-type protein. The other 
two structures were solved by single isomorphous replacement 
methodology and a MAD experiment (peak, inflection point and high-energy 
peak) to provide a high-resolution native crystal and a dataset obtained from wild-type protein. The other 
two structures were solved by single isomorphous replacement 
methodology and a MAD experiment (peak, inflection point and high-energy 
peak) to provide a high-resolution native crystal and a dataset obtained from wild-type protein. The other 
two structures were solved by single isomorphous replacement 
methodology and a MAD experiment (peak, inflection point and high-energy 
peak) to provide a high-resolution native crystal and a dataset obtained from wild-type protein. The other 
two structures were solved by single isomorphous replacement 
methodology and a MAD experiment (peak, inflection point and high-energy 
peak) to provide a high-resolution native crystal and a dataset obtained from wild-type protein. The other 
two structures were solved by single isomorphous replacement 
methodology and a MAD experiment (peak, inflection point and high-energy 
peak) to provide a high-resolution native crystal and a dataset obtained from wild-type protein. The other 
two structures were solved by single isomorphous replacement 
methodology and a MAD experiment (peak, inflection point and high-energy 
peak) to provide a high-resolution native crystal and a dataset obtained from wild-type protein. The other 
two structures were solved by single isomorphous replacement 
methodology and a MAD experiment (peak, inflection point and high-energy 
peak) to provide a high-resolution native crystal and a dataset obtained from wild-type protein. The other 
two structures were solved by single isomorphous replacement 
methodology and a MAD experiment (peak, inflection point and high-energy 
peak) to provide a high-resolution native crystal and a dataset obtained from wild-type protein. The other 
two structures were solved by single isomorphous replacement 
methodology and a MAD experiment (peak, inflection point and high-energy 
peak) to provide a high-resolution native crystal and a dataset obtained from wild-type protein. The other 

integrity. This role differs from metazoan αMs, where RBD targets cell-surface receptors prior to endocytosis (11).

Structure-derived snap-trap mechanism of induction — Comparison of iECAM and nECAM reveals the detailed mechanism of ECAM induction mediated by cleavage in the bait region (SI Appendix, Movie S1). This process yields a more compact structure (Fig. 3D), which is consistent with higher electrophoretic mobility—similarly to what happens with h20M (18)—and to the aforementioned differences in biophysical assays. Superposition shows that the structures only coincide on the bilayered side of the MG-superhelix—MG1-MG2 and MG5-MG6—and, partially, at BRD (up to Y152 and from H164 onwards). Upon induction, MG3 and MG4 are flipped outward towards MG6 as a rigid body due to a ~90°-rotation around the anchor point of MG3 with MG2 and a concomitant translation downward of up to ~50Å (for MG4; see Fig. 3D for spatial orientation hereafter). The new position of MG4 forces NIE to be moved outward along the outer surface of the four-stranded sheet of MG1. This movement traps the segment of the bait region upstream of the cleavage site following a ~180°-rotation downward around G11 (see above). The bait region is undefined from Q180 to G116 in iECAM and the distance between the flanking residues is too great to be covered by the ten missing residues (66Å). In contrast, in h20M the corresponding distance easily accommodates the missing residues (see SI Appendix, SRE §2.3 and Fig. S10 for a detailed comparison between iECAM and h20M). This explains why in ECAM cleavage in the bait region must occur to yield the induced form, whereas in h20M, the induced form is compatible with an intact bait region, and thus, can be obtained by MA-treatment (8). The displacement of MG3 is also concomitant with MG7 and RBD becoming rotated as a rigid body by ~25° downward, so RBD is displaced by ~25Å towards newly positioned MG3. Rearrangement of MG7 and RBD trypsin by soybean trypsin inhibitor.

Materials and Methods A detailed description of the Experimental procedures is provided under Supplementary Information. The latter also includes four supplementary tables, ten supplementary figures, the legend to a supplementary movie, the Acknowledgements and Supplementary Results and Discussion.

10. Strickland DK, et al. (1990) Sequence identity between the α1-α2-macroglobulin receptor and low density lipoprotein receptor-related protein suggests that this molecule is a multifunctional receptor. J. Biol. Chem. 265(29):17401-17404.
11. P170 (Fig. 3E-G). Rotating away CUB causes loops TED-a3→a4 (G1520,D1526) and TED-b2→a5 (E1256,E1260) to be displaced to the right, which further causes TED-a1 and α1→α2 (P1169,L1186) to undergo major rearrangement. In particular, TED-a1 (I1173, A1186) becomes unwound for its last six residues in iECAM. This causes displacement of segment Y1182-G1196, which acts as a protective lid of the thioester bond in nECAM. In this way, the thiostere becomes exposed and solvent accessible in iECAM, so it can be targeted by prey surface lysines (Fig. 3G). Most noteworthy, the initial movement of the mechanism, that of MG3 relative to MG2, is blocked in nECAM by the BRD segment after the bait region, which passes above the MG2-MG3 linker (Fig. 3B). Upon cleavage in the bait region this constraint is released, and the segment downstream of the cleavage site becomes rotated by ~50° around N863 towards and above MG2, and approaches the outer surface of CUB in its induced position.

Conclusions — Taken together, these results, together with the functional characterization in vitro and—partially—in vivo, indicate that ECAM works as an irreversible monomeric snap trap. This snap trap definitively differs from the tetramereric Venus-flytrap of mammalian αMs. Monomeric nECAM represents the baited and set trap, with a spring-loaded bar (the hidden thiostere) and a trip (BRD segment after the bait region) to release it. When the bait region is cleaved, induction occurs under large conformational rearrangement and exposure of a hidden thiostere loop, which is analogous to setting off the trap through the rapid swing-down of the spring-loaded bar. However, only if the thiostere bond is targeted by a surface lysine of the prey peptidase to yield a covalent bond, is the prey trapped by the released bar. In any case, the trap would remain irreversibly inactivated, either with or without a trapped peptidase. In contrast to a true snap trap, however, the prey peptidase is not disabled by ECAM but merely restricted in its radius of action and substrate size.
SUPPLEMENTARY INFORMATION

Table of Contents

§ 1. Experimental Procedures 2
   § 1.1. Protein production and purification 2
   § 1.2. Proteolytic inhibition assays 2
   § 1.3. Cell-envelope preparation and cleavage assays 3
   § 1.4. Oligomerization studies in vitro 3
   § 1.5. Protein binding studies in vitro 4
   § 1.6. Thermal shift assays and circular dichroism spectroscopy 4
   § 1.7. Protein localization assays 4
   § 1.8. Phenotypic analyses 4
   § 1.9. Western-blot analyses 5
   § 1.10. Crystallization and data collection 5
   § 1.11. Structure solution and refinement 6
   § 1.12. Single-particle cryo-electron microscopy and image processing 7
   § 1.13. Homology modeling of native ECAM 7
   § 1.14. Miscellaneous 8

§ 2. Supplemental Results and Discussion 9
   § 2.1. ECAM is a membrane-anchored cell-wall protector 9
   § 2.2. Dimeric arrangement of iECAM 9
   § 2.3. Comparison of iECAM with induced hα2M 9

§ 3. Acknowledgments 11

§ 4. Supplemental References 11

§ 5. Supplemental Tables 14
   Supplemental Table 1 14
   Supplemental Table 2 15
   Supplemental Table 3 16
   Supplemental Table 4 17

§ 6. Supplemental Figures 18
   Supplemental Figure 1 18
   Supplemental Figure 2 19
   Supplemental Figure 3 20
   Supplemental Figure 4 21
   Supplemental Figure 5 22
   Supplemental Figure 6 23
   Supplemental Figure 7 24
   Supplemental Figure 8 25
   Supplemental Figure 9 26
   Supplemental Figure 10 27

§ 7. Legends to Supplemental Movies 28
§ 1. EXPERIMENTAL PROCEDURES

§ 1.1. Protein production and purification — Constructs spanning different fragments of the genes coding for *Escherichia coli* strain K12 α2-macroglobulin (ECAM alias YfhM; 1,653 residues, UniProt [UP] sequence database accession code P76578) and penicillin-binding protein 1C (PBPI1C; 770 residues, UP P76577) were amplified with primers, which introduced restriction sites for directional cloning into vectors pET28a (Novagen), pCRI-7b, pCRI-8a and pCRI-8b (1). For constructs, plasmids, vectors and primers, see Suppl. Table 1. Polymerase chain reaction (PCR) primers and DNA modifying enzymes were purchased from Sigma-Aldrich and Thermo-Scientific, respectively. PCR was performed using Phusion High-Fidelity DNA Polymerase (Thermo-Scientific) according to the manufacturer’s instructions and following a standard optimization step by thermal gradient in each reaction. Mutants following were obtained a modified version of the procedure described by (2) as described elsewhere (1). DNA was purified with OMEGA-Biotek purification kits and all constructs were verified by DNA sequencing. Chemically competent *E. coli* DH5α and BL21 (DE3) cells (Novagen) were prepared and transformed according to Hanahan (3).

All constructs were overexpressed in *E. coli* BL21 (DE3) cells, which were grown at 37°C in lysogeny broth supplemented with 30µg/mL kanamycin. Selenomethione-derivatized (Se-Met) protein variants were obtained under the same conditions but using M9-based minimal medium (SelenoMethionine Medium; Molecular Dimensions) supplemented with selenomethionine instead of methionine. Cultures were induced at A600=0.6 with 0.4mM isopropyl-β-D-thiogalactopyranoside (IPTG) and incubated overnight at 18°C. Cells were collected by centrifugation at 6,000xg for 30min at 4°C, washed in buffer A (50mM Tris-HCl, 250mM sodium chloride, pH7.5) and resuspended in the same buffer plus 20mM imidazole, Complete EDTA-free Peptidase Inhibitor Cocktail tablets, and DNase I (both from Roche Diagnostics). Cells were lysed using a cell disrupter (Constant Systems) at a pressure of 1.35kbar, and the cell debris was removed by centrifugation at 30,000xg for 1h at 4°C. Supernatants were incubated with nickel-nitrilotriacetic acid (Ni-NTA) resin (Invitrogen) for 2h at 4°C, and bound proteins were subsequently loaded onto an open column for batch purification (Bio-Rad), washed extensively, and eluted with buffer A plus 350mM imidazole. Subsequently, samples were dialyzed overnight at 4°C against buffer A plus 1mM 1,4-dithio-DL-threitol (DTT) in the presence of hexahistidine(His6)-tagged tobacco-etch virus (TEV) peptidase at a peptidase:protein weight ratio of 1:100. The resulting cleavage left three additional residues at the N-terminus of target proteins due to the cloning strategy (see Suppl. Table 1). Digested samples were passed several times through Ni-NTA resin previously equilibrated with buffer A plus 20mM imidazole to remove His6-containing molecules. Whenever necessary for purification, the flow-through was collected, dialyzed overnight against buffer B (20mM Tris-HCl, 20mM sodium chloride, pH7.5) and further purified by ionic-exchange chromatography on a TSKgel DEAE-2SW column (TOSOH Bioscience) equilibrated with buffer B. A gradient of 2-to-30% buffer C (20mM Tris-HCl, 1M sodium chloride, pH7.5) was applied over 24mL and samples were collected and pooled. Subsequently, each pool was concentrated by ultrafiltration and subjected to final size-exclusion chromatography (SEC) on Superdex 75 or 200, 10/300 or 16/60 columns (GE Healthcare Life Sciences) in buffer D (20mM Tris-HCl, 150mM sodium chloride, pH7.5). Specifically, during the purification of a variant of ECAM lacking the receptor-binding domain (see Suppl. Table 1, ECAMΔRBD), the peptidase-independent dimerization that was observed during SEC was prevented by the addition of 1mM DTT to the running buffer.

Protein identity and purity were assessed by 10-15% Tricine-SDS-PAGE (4) stained with Coomassie blue, peptide-mass fingerprinting of tryptic protein digests (PMF), N-terminal sequencing through Edman degradation and mass spectrometry (MS). The latter three were carried out at the Protein Chemistry Service and the Proteomics Facilities of Centro de Investigaciones Biológicas (Madrid, Spain). Ultrafiltration steps were performed with Vivaspin 15 and Vivaspin 500 filter devices of 10-50kDa cut-off (Sartorius Stedim Biotech). Protein concentrations were estimated by measuring A280 in a spectrophotometer (NanoDrop) and applying the respective theoretical extinction coefficients. Concentrations were also measured by the BCA Protein Assay Kit (Thermo Scientific) with bovine serum albumin as a standard. Free sulfhydryl groups were determined by reaction of protein samples (at 5mg/mL) in buffer D at room temperature for 15min with Ellman’s reagent (5,5’-dithiobis-2-nitrobenzoic acid, DTNB; (5)), following the change in A412 in a microplate spectrophotometer (Power-Wave XS, Biotek). Methylamine(MA)-modified ECAM (MA-ECAM) was obtained by treating native ECAM (nECAM) in buffer D with 200mM methylamine hydrochloride for 1h at room temperature and subsequently buffer-exchanged to the same buffer.

§ 1.2. Proteolytic inhibition assays – nECAM produced from plasmid pECAM2 was used to study proteolytic inhibition towards peptide and protein substrates after previous incubation for different time spans
in buffer E supplemented with 10mM calcium chloride used. Reactions were incubated at room temperature at a trypsin:substrate ratio of 1:10,000 in complex with trypsin were both adjusted to be the equivalent to that of a free peptidase sample on the basis of their activity towards a small fluorogenic peptidic substrate (FRET4; see below) in order to provide comparable amounts of active trypsin for the inhibition assays. These samples were used to measure their inhibition of the aforementioned substrates at room temperature at a trypsin:substrate ratio of 1:5. Inhibition was measured after preincubation of the peptidases for 15min with excess nECAM (up to 20-fold molar excess). In all cases, the reactions were performed at room temperature over a period of 2h and cleavage was assessed by 10–15% Tricine-SDS-PAGE after stopping the reaction by treatment for 15min with small-molecule inhibitors (4mM Pefabloc SC [Roche Life Science] for serine peptidases and 10mM EDTA for metallopeptidases) or by precipitation with 2,2,2-trichloroacetic acid (TCA; Sigma-Aldrich).

In addition, cleavage-inhibition activity against peptides was determined in a microplate fluorimeter (Infinite M200, TECAN) with the fluorescence-based EnzCheck Assay Kit containing either BODIPY FL-casein (λex=505nm and λem=513nm), DQ-gelatine (λex=495nm and λem=515nm) or DQ-bovine serum albumin (λex=590nm and λem=620nm) as fluorescence conjugates (Invitrogen) at 10µg/mL in 200µL-reaction volumes. Further protein substrates assayed included bovine milk α-casein (35kDa), aldolase from rabbit muscle (160kDa), fumarase from porcine heart (200kDa), and bovine thyroglobulin (660kDa; all from Sigma-Aldrich and at 1mg/mL). Specifically for the results depicted in Suppl. Fig. 3A, the concentrations of purified monomeric and dimeric trypsin-induced iECAM (see also §1.4) in complex with trypsin were both adjusted to be the equivalent to that of a free peptidase sample with 0.3% SDS without boiling. 7.5% Tricine-SDS-PAGE with 0.05% casein in the resolving gel was pre-run for 1h at 125V and then loaded with the protein samples. Subsequently, SDS was removed from the gel by two washes in 2.5% Triton X-100 for 20min. The zymogram was incubated for 3h at 37°C in buffer E and finally stained with Coomassie blue.

§ 1.3. Cell-envelope preparation and cleavage assays — E. coli K12 cells were grown in lysogeny broth at 37°C until A580=0.6, collected by centrifugation at 6,000xg for 30min at 4°C, and resuspended in buffer E supplemented with 10mM EDTA. Cells were lyzed as described above at 2.4kbar, and intact cells were removed by centrifugation at 2,000xg for 1h at 4°C. Subsequently, cell-envelope fractions were collected by ultracentrifugation at 150,000xg for 2h at 4°C in a Beckman Optima L-90K using a 50.2 Ti rotor (Beckman) and 26.3-mL polycarbonate bottles with cap assembly (Beckman). Collected cell envelopes were homogenized using a glass potter in buffer E and the total protein concentration was adjusted to 2mg/mL as measured with the BCA Protein Assay Kit (see §1.4). Proteolytic activity of trypsin (1.1µg), chymotrypsin (1.5µg), subtilisin (0.5µg), thermolysin (2µg), pancreatic elastase (2.5µg), and ulilysin (2.5µg) against 200µL of purified cell envelope was tested at room temperature by measuring the decrease of A278 (13) in a microplate spectrophotometer (PowerWave XS, Biotek) in the presence or absence of freshly-purified nECAM or MA-ECAM at peptidase:ECAM ratios between 1:2 and 1:20.

§ 1.4. Oligomerization studies in vitro — Oligomerization assays of nECAM and mutants ECAM-R162G, ECAMABRD, and TEV-ECAM (all at 1mg/mL) were performed by treatment with peptidases (see §1.2) at a peptidase:inhibitor molar ratio of 2:1 for all reactions except for TEV-ECAM, where a molar ratio of 5:1 was used. Reactions were incubated at room temperature for 2h (Suppl. Figs. 1G and 2) or up to 24h (Suppl. Fig. 3B) in buffer E supplemented with 10mM calcium chloride (for reactions with thermolysin and ulilysin) or 1mM DTT.
(for reactions with TEV peptidase). Oligomerization was monitored by SEC and 5-10% Tris-glycine native-PAGE (14). Dimeric trypsin-induced MA-ECAM at 1mg/mL was tested for stability after treatment for 30min at room temperature with either 500mM sodium chloride, 1% n-octyl-β-D-glucopyranoside, or 1mM DTT and subsequently analyzed by SEC in buffer D supplemented with the respective above components.

§ 1.5. Protein binding studies in vitro — To estimate the amount of peptidase covalently bound to iECAM, trypsin was first labeled with fluorogenic sulfosuccinimidyl-7-amino-4-methylcoumarin-3-acetate (Sulfo-NHS-AMCA; Thermo Scientific) according to the manufacturer’s instructions at a reagent:peptidase molar ratio of 4:1. Subsequently, Tris-HCl, pH7.5 was added (100mM final concentration) to stop the reaction, and the protein was extensively dialyzed against buffer D. Labeled trypsin was first checked for activity and then reacted with nECAM at a peptidase:inhibitor molar ratio of 4:1 and the complex was purified by SEC. The amount of peptidase bound to ECAM was calculated based on a calibration curve with known concentrations of unbound labeled trypsin. Other peptidases listed under §1.2 lost activity or self degraded upon labeling and were not further studied.

Binding to TEV peptidase by TEV-ECAM, with or without methylation of TEV-peptidase lysines, was measured by 10% SDS-PAGE and 10% native PAGE. Methylation of TEV peptidase was previously performed with the JBS Methylation Kit (Jena Biosciences) according to the manufacturer’s instructions. Methylated protein was dialyzed against buffer D and the remaining activity was measured against fluorogenic substrate FRET8 (see §1.2).

§ 1.6. Thermal shift assays and circular dichroism spectroscopy — For thermal shift assays, aliquots were prepared by mixing 7.5µL of 300x Sypro Orange dye (Molecular Probes) and 42.5µL of protein solution (at 1mg/mL) in buffer E. Samples were analyzed in an iCycler iQ Real Time PCR Detection System (BioRad) by mixing 7.5µL of 300x Sypro Orange dye (Molecular Probes) and 42.5µL of protein solution (at 1mg/mL) in buffer D. Measurements were carried out in a Jasco J-810 spectrometer at 25°C with a high precision cell (Quartz SUPRASIL; Hellma Analytics) of 1mm path length. Both techniques were applied to study nECAM, MA-ECAM, and monomeric and dimeric trypsin-induced iECAM.

§ 1.7. Protein localization assays — E. coli strain BL21 (DE3) was used to co-express full-length ECAM fused to Strep-tag and PBPI (plasmid pECAM7; Suppl. Table 1) by auto-induction in lysogeny broth supplemented with 0.05% glucose at 37°C for 5h (15). Cells were collected by centrifugation, washed in buffer D, and subsequently fractionated as described elsewhere (16). Briefly, collected cells were resuspended in 200mM Tris-HCl, 500mM sucrose, 1mM EDTA, pH8.0 supplemented with Complete EDTA-free Peptidase Inhibitor Cocktail tablets and DNase I and incubated on ice for 30min, and the mixture was centrifuged at 15,000xg for 30min at 4°C. The supernatant included the periplasmic and outer-membrane fractions and the precipitant contained the cytoplasmic and inner-membrane fractions. The precipitant was further lyzed in a cell disruptor as described above and the extract was centrifuged at 15,000xg for 30min to remove unbroken cells, inclusion bodies and cell debris. Subsequently, supernatants from both steps were further ultracentrifuged at 150,000xg for 30min to separate cell-envelope fractions from soluble proteins. Samples from the inner-periplasmic, periplasmic, and outer-membrane fractions were analyzed by anti-Strep-tag Western-blot analysis (see also §1.9).

§ 1.8. Phenotypic analyses — E. coli K12 strains were obtained from the National BioResource Project KEIO collection (NIG, Japan). The wild-type strain (strain BW25113A) and the mutant lacking ECAM (strain JW2504), in which the gene encoding ECAM is replaced by a gene coding for kanamycin resistance, were grown in tryptic soy broth (TSB) supplemented with 30µg/mL kanamycin (only mutant strain) at 37°C under anaerobic conditions (H2:CO2:N2=10%:10%:80%). Cells were collected during exponential growth at A600=0.6-0.8, washed twice with phosphate buffered saline (PBS; 137mM sodium chloride, 2.7mM potassium chloride, 4.3mM disodium hydrogen phosphate 1.47mM sodium dihydrogen phosphate, pH 7.5) and resuspended in 10mM sodium phosphate, pH 7.0, supplemented with 5% [v/v] TSB medium. Cells were subsequently treated with trypsin at final concentrations of 100 and 500µg/mL for 5hours at 37°C under anaerobic conditions. Serial dilutions were immediately spread on plates and the number of colony forming units was determined after overnight incubation. In addition, the ECAM knockout was further transformed with plasmid pECAM9, which expresses ECAM and PBPI and is controlled by an IPTG-inducible Ptac promoter (see Suppl. Table 1), or the same plasmid expressing green fluorescence protein. Viability of transformants was assessed as mentioned above except for the addition of 20µg/mL zeocin and 0.1mM IPTG for ECAM and PBPI induction during the exponential growth
phase in TSB medium. Statistical significance of the results was determined with a Student’s paired t-test between the survivability of the wild-type strain and the mutants lacking ECAM.

§ 1.9. Western-blot analyses — Protein samples were separated in 10-15% SDS-PAGE, transferred to Hybond ECL membranes (GE Healthcare Life Sciences), and finally blocked overnight at room temperature with 20mL of blocking solution, which was PBS (see §1.8) further containing 0.05% Tween 20 and 1.5% bovine serum albumin. His$_{6}$-tag fused SEC fractions of peptidase-induced iECAM (see Suppl. Fig. 1G,H) were detected by immunoblot analysis using monoclonal antibodies diluted 1:1,000 and a secondary antibody (goat anti-rabbit IgG (HL) peroxidase-conjugated antibody; both from Santa Cruz Biotechnology) diluted 1:5,000 (both in blocking solution). Strep-tag fused ECAM (see Suppl. Fig. 2K) was detected with Strep-Tactin® HRP conjugate (IBA-Solutions for Life Sciences) diluted 1:2,000. Complexes were detected using an enhanced chemiluminescence system (Super Signal West Pico Chemiluminescent; Pierce) according to the manufacturer’s instructions. Membranes were exposed to Hyperfilm ECL films (GE Healthcare Life Sciences).

§ 1.10. Crystallization and data collection — Crystallization screenings were performed at the joint IBMB/IRB Crystallography Platform (PAC, Barcelona) by the sitting-drop vapor diffusion method using 96x2-well MRC plates (Innovadyne). A TECAN Freedom EVO robot was used to prepare reservoir solutions, and a Cartesian Microsys 4000 XL (Genomic Solutions) robot or a Phoenix/RE (Art Robbins) robot was used for 100-nL nanodrop dispensing. Crystallization plates were stored in Bruker steady-temperature crystal farms at 4°C or 20°C. Successful hits were scaled up to the microfilter range with 24-well Cryscrm crystallization dishes (Hampton Research).

Native protein spanning the residues from domain MG7 to the C-terminus (sequence $G^3+H^2+M^{14}+A^{1018}$, p$^{1653}$; see Fig. 1A and Suppl. Fig. 7 for ECAM domains) as produced with plasmid pECAM3, hereafter protein nECAMAN, was incubated for 30 min at room temperature with 5mM DTT and subsequently 30 min more with 25mM 2-iodoacetamide. Excess of the latter reagent was removed by SEC. This treatment prevents dimer formation when thioester bonds are open and, thus, cysteines are free, but it does not affect intact thioester bonds (17). The resulting protein was crystallized from drops containing 1µL of protein solution (at 10mg/mL in 10mM Tris-HCl, 75mM sodium chloride, pH7.4) and 1µL reservoir solution (20% [w/v] PEG4,000, 10% isopropanol, 100mM Tris-HCl, pH9.0). Trapezoidal bar-shaped crystals appeared within 3d at 20°C. Se-Met protein was crystallized similarly in drops containing 1µL of protein solution (at 20mg/mL in 10mM Tris-HCl, 75mM sodium chloride, pH7.4) and 1µL of reservoir solution (14% [w/v] PEG4,000, 20% isopropanol, 100mM Tris-HCl, pH9.0).

Native ECAM fragment spanning domains NIE-MG1 ($G^3+A^{2}+M^{14}+D^{163}-L^{368}$; produced with plasmid pECAM4) was crystallized from drops containing 1µL of protein solution (at 20mg/mL in 10mM Tris-HCl, 75mM sodium chloride, pH7.4) and 1µL reservoir solution (25% [w/v] PEG3,350, 200mM ammonium acetate, 100mM sodium phosphate, pH5.0). Se-Met protein was crystallized in the same reservoir conditions as the native protein but using a 10mg/mL protein solution. Long, flat bars appeared within 3-4d at 20°C in both cases.

Native ECAM fragment spanning domains MG0-NIE-MG1 ($G^3+H^2+M^{14}+A^{40}+S^{385}$; produced with plasmid pECAM5) was crystallized from drops containing 1µL of protein solution (at 10mg/mL in 10mM Tris-HCl, 75mM sodium chloride, pH7.4) and 1µL of reservoir solution (30% [w/v] PEG4,000, 100mM Tris-HCl, pH9.0-10.5). Se-Met protein crystallized in the same conditions. Octahedral crystals appeared within 2d at 20°C.

Trypsin-induced iECAM for crystallization was obtained by incubation of nECAM ($G^3+P^2+M^{14}+A^{40}$, p$^{1653}$; produced with plasmid pECAM2) with trypsin at a molar ratio of 1:2 for 3h at room temperature, treated with 1mg/mL Pefabloc to block trypsin activity, and purified by SEC. The protein crystallized in drops containing 1µL of protein solution (at 10-20mg/mL in 10mM Tris-HCl, 75mM sodium chloride, pH7.4) and 1µL of reservoir solution (12-16% [w/v] PEG8,000, 100mM Tris-HCl, pH8.5). The same reservoir solution was used to crystallize the Se-Met protein at 5-10mg/mL in 10mM Tris-HCl, 75mM sodium chloride, pH7.4. Bar-shaped crystals appeared within 1d at 20°C. Analysis by fluorescence microscopy in a Nikon E600 (filter UV-2A) apparatus operated at the Advanced Fluorescence Microscopy Unit (IBMB-PCB) of carefully washed crystals of iECAM obtained with trypsin previously treated with Sulfo-NHS-AMCA (see §1.5) revealed that the crystals contained trypsin in addition to iECAM (Suppl. Fig. 4). This presence was further confirmed by PMF of carefully washed dissolved crystals. Trypsin-treated MA-ECAM yielded crystals that were indistinguishable from those of iECAM. However, these crystals showed poor diffraction and were not further considered.

All crystals were cryo-protected by rapid passage through drops containing increasing concentrations of glycerol (up to 20% [v/v]). Complete diffraction datasets were collected at 100K from liquid-N$_2$ flash cryo-cooled crystals (Oxford Cryosystems 700 series cryostream) on a Pilatus 6M pixel detector (from Dectris) at beam line
XALOC of ALBA synchrotron (Barcelona, Spain; datasets Se-Met nECAMΔN, native and Se-Met NIE-MG1, and native and Se-Met MG0-NIE-MG1). Further data were collected on the same detector type at beam lines ID29 (native nECAMΔN) and ID23-1 (native and Se-Met iECAM) of ESRF synchrotron (Grenoble, France) within the Block Allocation Group “BAG Barcelona.” All crystals contained one protein molecule per asymmetric unit (solvent contents: nECAMΔN=67%; NIE-MG1=40%; MG0-NIE-MG1=49%; and iECAM=59%; calculated for the respective native datasets, see Suppl. Table 3). Diffraction data were integrated, scaled, merged, and reduced with programs XDS (18) and XSCALE (19), and transformed with XDSCONV to formats suitable for SHELX (20, 21) and the CCP4 suite of programs (22)(see Suppl. Table 3 for data processing statistics).

§ 1.11. Structure solution and refinement — The structure of nECAMΔN was solved by single isomorphous replacement including anomalous signal (SIRAS) using the program pipeline XPREP-SHELXD-SHELXE (20, 21) with a dataset of a Se-Met crystal collected at the selenium absorption peak and a native dataset to higher resolution. SHELXD identified 17 sites with the peak-wavelength dataset, which resulted from the 13 methionines of the chemical sequence of the construct (two in alternate positions) plus two nickel cations. Phasing with these sites and autotracing with SHELXE was followed by a density modification and model extension step with the AUTOBUILD protocol of PHENIX (23), which produced an improved Fourier map and a partial model. Subsequent manual model building with the COOT program (24) alternated with crystallographic refinement with PHENIX (25) and BUSTER/TNT (26), which included TLS refinement, until the final refined model was obtained. This consisted of residues G^{1015}\cdot P^{1635}; two nickels, possibly resulting from the purification strategy (see 814) and involved in functionally irrelevant crystal contacts; five glycerols; and 96 solvent molecules. The average thermal displacement parameters for the constituting domains were the following: MG7, 77.1Å²; CUB, 83.7Å²; TED, 81.2Å²; and RBD, 79.1Å² (Overall: 80.4Å²). See Suppl. Table 4 for final refinement and model quality statistics.

The structure of NIE-MG1 was likewise solved by SIRAS, with a dataset from a Se-Met crystal collected at the selenium absorption peak and a native dataset to higher resolution. Four selenium sites were found by SHELXD with the peak-wavelength dataset: three from the substituted methionines of the chemical sequence of the protein and one from the extra selenomethionine at the N-terminus due to the cloning strategy (see Suppl. Table 1). Initial phasing with these sites was followed by a density modification and model extension step with program ARP/wARP (27), which produced an improved Fourier map and a partial model. Thereafter, model completion and refinement proceeded as for nECAMΔN. The final model of NIE-MG1 comprised protein residues A^2\cdot M^4\cdot D^{163}\cdot L^{368}, one acetate, one glycerol, and 200 solvent molecules. The average thermal displacement parameters for the constituting domains were the following: NIE, 34.0Å² and MG1, 26.0Å² (Overall: 31.5Å²).

The structure of MG0-NIE-MG1 was solved by multi-wavelength anomalous dispersion/multiple isomorphous replacement including anomalous signal (MAD/MIRAS) with a native dataset and data from a three-wavelength MAD experiment (peak, inflection point and high-energy remote) performed with a Se-Met protein crystal. Four sites of the heavy-atom substructure—corresponding to the three methionines of the construct, with one in alternate position—were found by SHELXD with the peak-wavelength dataset. Subsequent initial phasing with SHELXE identified P4_{1}2_{1}2 as the correct Sohnke group. Thereafter, more accurate phasing with these positions was carried out with program SHARP (28), and subsequent density modification was performed with programs SOLOMON (29) and DM (30). The resulting map was automatically interpreted with the AUTOBUILD protocol of PHENIX to yield an improved Fourier map and a partial model. Thereafter, model completion and refinement proceeded as stated above. The final model of MG0-NIE-MG1 was refined against the merged Se-Met crystal data, i.e. it contained selenomethionines instead of methionines, and it comprised protein residues K^{57}\cdot K^{367} (except N^{218}\cdot Q^{224}), one molecule of di-, tri-, tetra-, and pentaethylene glycol, five glycerols, and 40 solvent molecules. The average thermal displacement parameters for the constituting domains were the following: MG0, 81.7Å²; NIE, 79.5Å² and MG1, 65.7Å² (Overall: 77.2Å²).

Finally, the structure of iECAM was likewise solved by MAD/MIRAS with a native dataset and data from a four-wavelength MAD experiment (peak, inflection point, second inflection point and high-energy remote) of a Se-Met crystal. Phasing was carried out with SHARP based on 21 of the 23 possible selenium sites of the construct. This was followed by a density modification and model extension step with the AUTOBUILD protocol of PHENIX, which produced an improved Fourier map and a partial model. The density for the first two domains of iECAM, NIE and MG1, was very weak due to intrinsic flexibility of the molecule, although N-terminal sequencing of carefully washed crystals unambiguously revealed their presence (data not shown). This weakness explained, in turn, why the first two selenium sites of the sequence (M^{168} and M^{231}) could not be identified during
phasing, and it motivated preparation and crystallization of constructs MG0-NIE-MG1 and NIE-MG1, which yielded good experimental models (see above). These, in turn, were used for careful model building of NIE and MG1 in iECAM. The final model of iECAM comprised protein residues A\textsuperscript{156}-P\textsuperscript{1653}, except for segments N\textsuperscript{947}, Y\textsuperscript{676}-D\textsuperscript{691}, D\textsuperscript{834}-A\textsuperscript{839}, N\textsuperscript{1372}-R\textsuperscript{1376} due to secondary trypsin cleavage (see Suppl. Fig. 1f), and residues upstream (Q\textsuperscript{939}-R\textsuperscript{946}) and downstream (P\textsuperscript{947}-G\textsuperscript{948}) of the primary trypsin cleavage site in the bait region (R\textsuperscript{946}-F\textsuperscript{947}). Three chlorides, one glycerol, and 105 solvent molecules completed the final model of iECAM. The high average thermal displacement parameter of all protein atoms (111.1Å\textsuperscript{2}; see Suppl. Table 4) supported the flexible character of the molecule, in particular of its N-terminal part, and the fact that the trypsin molecule trapped inside the iECAM crystals (Suppl. Fig. 4) was disordered and could not be modeled. The intrinsic flexibility of some (sub-)domains of the structure was revealed by the substantial differences in average thermal displacement parameters: NIE, 164.7Å\textsuperscript{2}; MG1, 165.8Å\textsuperscript{2}; L, 125.9Å\textsuperscript{2}; MG2, 107.2Å\textsuperscript{2}; MG3, 88.5Å\textsuperscript{2}; MG4, 136.8Å\textsuperscript{2}; MG5, 157.9Å\textsuperscript{2}; MG6, 126.8Å\textsuperscript{2}; BRD upstream of cleavage site, 112.9Å\textsuperscript{2}; BRD downstream of cleavage site, 85.0Å\textsuperscript{2}; MG7, 107.1Å\textsuperscript{2}; CUB, 112.0Å\textsuperscript{2}; TED, 75.6Å\textsuperscript{2}; and RBD, 69.9Å\textsuperscript{2}.

§ 1.12. Single-particle cryo-electron microscopy and image processing — Feasibility of electron microscopy (EM)-based studies of ECAM was initially assessed by conventional negative-straining EM. 5μL samples of purified nECAM and dimeric trypsin-induced iECAM (in buffer D) were applied to glow-discharged carbon-coated grids for 2min and negatively stained with 2% aqueous uranyl acetate. For single-particle cryo-EM, 5μL samples of purified nECAM, and monomeric and dimeric trypsin-induced iECAM were applied to glow-discharged carbon-coated Cu/Rh 300 mesh Quantifoil R 1.2/1.3μm grids, blotted and plunged into liquid ethane with a Leica EM CPC cryo-fixation unit. Images were recorded on a FEI Eagle 4k CCD camera at low-dose conditions, with a FEI Tecnai G2 electron microscope operated at 200kV and equipped with a field emission gun. Images were recorded at a nominal calibrated magnification of 84,269x, and an underfocus ranging 1.5-3.5μm.

General image processing operations were performed using XMIPP software (31) and pictures were drawn with the CHIMERA program (32). Images were CTF-corrected and down-sampled to a factor of 2, with a final sampling ratio of 3.56Å/pixel. The XMIPP automatic picking routine was used to select 74,099, 40,302 and 47,712 particles of nECAM, monomeric iECAM and dimeric iECAM, respectively. Images were classified using a reference-free clustering approach with program CL2D (33) to select homogeneous populations of 46,842, 18,346 and 33,536 particles for each of the aforementioned species. For each sample, a Gaussian blob and an artificial noise model were used as starting reference models for parallel iterative angular refinement using program EMAN (34). When both strategies converged to a similar solution, one of the resulting models was selected and refined using the XMIPP iterative projection matching routine (35). After independent refinement processes, 90% of particles were included in the final three-dimensional reconstruction, and the final resolution of the models was determined by the Fourier shell correlation (FSC) criterion between independent half-dataset maps applying a correlation limit of 0.5 (see Suppl. Fig. 8).

The CHIMERA fitting routine was used to dock atomic models in the cryo-EM maps after initial manual placement. For nECAM, an initial composite homology model (see §1.13) was iteratively fitted into the cryo-EM density as two independent rigid bodies: domains MG0-NIE and domains MG1-to-RBD. For monomeric iECAM, one subunit from the dimeric iECAM crystallographic structure (see §1.11) was fitted in the cryo-EM map. For dimeric iECAM, each protomer of the crystallographic dimer was fitted in the cryo-EM map as an independent rigid body.

§ 1.13. Homology modeling of native ECAM — Based on the experimental coordinates of native ECAM fragments MG0-NIE-MG1, NIE-MG1, and nECAMΔN (domains MG7-CUB-TED-RBD), and induced iECAM (domains MG2-MG3-MG4-MG5-MG6(RBD)); the cryo-EM reconstruction of nECAM to 16Å (see §1.12) and the coordinates of native Salmonella enterica serovar typhimurium αM (SEAM; PDB 4U48; (36); 82% sequence identity with ECAM), a composite homology model was constructed for the entire native structure of nECAM (K\textsuperscript{57}, P\textsuperscript{1653}). For this, the distinct ECAM domains were superposed onto the corresponding SEAM domains to yield an initial model, which, in several cycles, was: (i) fitted to the cryo-EM volume, first manually and thereafter with program CHIMERA; (ii) domain-connected with linkers of reasonable geometry; and (iii) geometry-minimized with programs COOT and PHENIX. In particular, the experimental structure nECAMΔN—spanning MG7-CUB(TED)-RBD—fitted onto the corresponding part of native SEAM (rmsd of 1.17Å), thus confirming that this ECAM fragment was in native conformation. The final nECAM model was validated with MOLPROBITY (37) [poor rotamers: 3 (0.2%); Ramachandran outliers/favored/all residues: 9 (0.6%)/1513 (94.9%)/1595 (100%); Cβ deviations/residues with poor bonds/angles: 0/0/0; rmsd bonds/angles: 0.003Å/0.648°; MOLPROBITY score: 2.03 (74th percentile)].
§ 1.14. Miscellaneous — Figures and Suppl. Movie were prepared with the CHIMERA program. Those in stereo are shown in cross-eye stereo. Repetitive secondary-structure element prediction was performed with JPRED3 (38). Pairwise structural superpositions were performed with the SSM routine (39) within COOT and with DALI (40). Model validation was performed with MOLPROBITY and analysis of protein interfaces, including theoretical calculation of thermodynamic parameters, was done with PISA (41). The final experimental coordinates of nECAMΔN, NIE-MG1, Se-Met MG0-NIE-MG1, and iECAM are deposited with the PDB (access codes 4ZIU, 4ZJH, 4ZJG, and 4ZIQ, respectively). The cryo-electron microscopy reconstructions of nECAM—with its associated composite atomic homology model—, trypsin-induced dimeric iECAM, and trypsin-induced monomeric iECAM have been deposited with the EMDataBank (http://www.emdatabank.org) with access codes EMD-3016, EMD-3017 and EMD-3018, respectively.
§ 2. SUPPLEMENTAL RESULTS AND DISCUSSION

§ 2.1. ECAM is a membrane-anchored cell-wall protector

ECAM was associated to the inner membrane of E. coli in a cell-fractionation experiment followed by Western-blot analysis but not to the periplasm or the outer membrane (see §1.7 and §1.9, and Suppl. Fig. 2E). These data, together with previous reports (42-44), are consistent with ECAM being anchored to the periplasmic side of the inner membrane through a “lipobox”-mediated lipidic linkage of the N-terminal cysteine residue of the secreted protein (C\(^\text{\text{3}}\)) through post-translational modification.

To shed light on the function of ECAM in vivo, wild-type E. coli strain K12 and the ECAM knockout mutant were analyzed phenotypically. Exposure of anaerobically growing cells to trypsin resulted in concentration-dependent decrease in cell survival that was significantly larger in the mutant (§1.8 and Suppl. Fig. 3E). Consistently, transformation of the ECAM mutant strain with a plasmid encoding ECAM and PBPlC (see §1.8 and Suppl. Table 1) rescued the wild-type phenotype but not the same plasmid encoding green fluorescence protein instead (Suppl. Fig. 3E). This agrees with the finding that the protein is an acute-phase protein that is over three-fold up-regulated under anaerobic conditions when compared with aerobic conditions (45). Together with the assays on cell-envelope extracts mentioned in the main text, these results indicate that ECAM protects essential cell-envelope components against digestive peptidases, possibly when the outer membrane is disrupted.

§ 2.2. Dimeric arrangement of iECAM

The iECAM oligomeric structure in the crystals is a dimer (Suppl. Fig. 5A-E). Dimerization buries an area of 6,010Å\(^2\), which contributes to a theoretical solvation free energy gain upon complex formation (\(\Delta G^{\text{int}}\)) of -68.3kcal/mol and a free energy of assembly dissociation (\(\Delta G^{\text{diss}}\)) of 38kcal/mol. This oligomerization is consistent with trypsin eventually causing iECAM to dimerize after induction and removal of the N-terminal fragment preceding NIE (see above), and with the induced form of ECAM adopting an equivalent conformation in monomers and dimers (see above and cryo-EM studies below). The dimer is a large elongated particle of maximal dimensions ~180Å (length; according to Suppl. Fig. 5A) x ~90Å (width) x ~80Å (depth), which encircles a central prey chamber of ~40Å in diameter (Suppl. Fig. 5E-F). Dimerization involves two regions. One symmetrically includes the NIEs of each protomer (NIE-α2, -β6 and -β6'). The other region includes TED (TED-α4, -β1, -β2, -α5, and -α6→α7) of one monomer and MG4-β1’→β1’’ and -3→β4 plus the outer surface of the three-stranded β-sheet and helix α1 of RBD of the other protomer. The dimeric particle is characterized by five large openings: two intramolecular “entrances 1” (see above and Suppl. Fig. 5B), two very large intermolecular “entrances 2” at the dimerization interface (~60Å x ~30Å; Suppl. Fig. 5C), and one “entrance 3” (~20Å in diameter, symmetrically created by TED and RBD of each protomer; Suppl. Fig. 5D). The thioster segments are on the inner surface of the particle and modeling indicates that a single trypsin molecule could be placed in the central prey chamber and thus be accessible to substrates through any of the five entrances (Suppl. Fig. 5A). However, the thiosteres are also close to an entrance 2, so the bound prey could likewise be placed outside the cage while remaining covalently linked to the particle. Both situations would be compatible with crystal packing and would explain the activity of bound peptidases except against very large substrates. In addition, like other such cage structures, the dimer may undergo “breathing” motions of segments surrounding the trapped prey as suggested for hα\(_2\)M (46) and by our cryo-EM reconstruction of the iECAM dimer (see below). Here it should be pointed out, however, that dimerization is not a universal mechanism for ECAM but rather restricted to specific peptidases such as trypsin and ulilysin (see above), and that the induced monomer is the most abundant functional species.

§ 2.3. Comparison of iECAM with induced hα\(_2\)M

Superposition of the monomers of iECAM and our previous MA-induced hα\(_2\)M structure (PDB 4ACQ: (46, 47)) reveals that they share the overall shape grasso modo. This is consistent with both structures corresponding to induced species. However, detailed inspection shows that only a core consisting of domains MG2, MG3, MG6 and MG7 can be reasonably well matched (Suppl. Fig. 10A-C). MG4 is rotated around its anchor point with MG3 ~80° outward respective to hα\(_2\)M in iECAM (Suppl. Fig. 10B,C). This entails that the MG1+MG5 tandem preceding MG4 is likewise rotated outward by ~70° around its hinge with the MG2+MG6
tandem. Tandem CUB+TED of hα2M, in turn, matches the respective part of iECAM reasonably well separately, i.e. the relative disposition and orientation of CUB and TED is comparable in both structures. Comparison of the respective TED domains alone reveals substantial differences in the loops linking the α-helices of the toroid. In particular, ECAM segment connecting helices α6 and α7 of domain TED (TED-α6→α7) is shorter than in hα2M (18 vs. 24 residues). This segment covers the central shaft of the toroid on its entry face and approaches the thioester segment through N²⁰⁸⁸ (hα₂M residues in italics, see UP 01023) in the human protein. In the more distantly related complement proteins, this residue is replaced by a histidine, which plays a crucial role in thioester binding of sugar hydroxyls instead of lysine Nζ-amines only as in α₂Ms (48). In ECAM, TED-α6→α7 does not even come close to the thioester region. Other largely deviating loops are TED-α8→α9 and TED-α10→α11, which are likewise shorter in the bacterial enzyme. Taken together, these features make the bacterial thioester more accessible than the mammalian one, which is consistent with covalent prey linkage being an absolute requirement for inhibition in the former, while it is just optional in the latter.

In the context of the entire protomer, iECAM CUB+TED is rotated by ~30° with respect to hα₂M toward the MG-superhelix around its anchor point with MG7 (Suppl. Fig. 10A,C), with a maximal displacement of ~35Å for the distal part of TED. Most interestingly, domain RBD, which was present in hα₂M in only one of the protomers of the tetramer, occupies a similar position to hα₂M in iECAM—at the interface with TED, CUB, MG7 and MG3—just being relatively translated by ~10Å. This suggests that, contrary to what we had postulated previously (46), RBD may very well occupy a functionally relevant position in hα₂M, i.e. poised to be recognized by its cognate cell-surface receptor. As this domain is not exposed and, thus, not recognized by its specific receptors in native hα₂M, it must adopt a different arrangement in the latter. Finally, hα₂M does not have a counterpart for MG0 and NIE, which are placed below MG1 in iECAM and nECAM. This explains the differences in chain length between bacterial (~1600-1700 residues) and metazoan α₂Ms (~1400-1500 residues) due to the former being membrane-anchored proteins and the latter being soluble serum proteins.

Within the respective oligomers, the central prey chambers likewise diverge: in dimeric iECAM, it merely spans ~40Å in diameter, which may accommodate a single small-to-medium sized endopeptidase at best; in tetrameric hα₂M, it spans up to ~60Å, which is consistent with up to two trypsin-sized peptidase molecules being trapped (46, 49). In hα₂M, the distance between the two anchor points of the flexible bait region (C⁶₈₉⁹ and E²⁷²⁹, 90Å) clearly suffices to accommodate the missing 41 residues (46). In contrast, in iECAM 66Å separate G⁹³⁸ from G⁹⁴⁹, which is too far to be spanned by the twelve missing residues. This explains why in ECAM cleavage in the bait region must occur to yield the induced form, whereas in human α₂M the induced form is compatible with an intact bait region, so it can be obtained by MA-treatment. In addition, the flexible bait region runs freely through the inner cavity of the tetramer in hα₂M. In iECAM, this is also the case for the upstream segment but not for the downstream segment, which is trapped (see main text). While the upstream anchor point of the bait region is found in similar regions of the protomer in both structures (22Å apart; (“UP” in Suppl. Fig. 10A), the downstream anchor point is 49Å away and found in disparate regions of the protomer: on the outer surface at the MG6-MG7 interface in hα₂M and between MG2 and CUB in iECAM (“DOWN” in Suppl. Fig. 10A). Overall, this correlates with the entire BRD domain spanning 126 residues in hα₂M (bait region P²⁹⁹⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻˓
§ 3. ACKNOWLEDGMENTS

We are indebted to Tibisay Guevara for her outstanding dedication during crystallization experiments and assistance throughout the project; to Robin Rycroft for helpful contributions to the manuscript; and to Jan J. Enghild for discussions on monomeric α₂-macroglobulins. We are also grateful to Joan Pous, Xandra Kreplin, and Sonia Rodriguez Puente from the joint IBMB/IRB Crystallography Platform (PAC) at Barcelona Science. David Waugh (Frederick, MD; USA) and Ulrich Baumann (Cologne; Germany) are further acknowledged for providing the expression plasmids and purification protocols of, respectively, TEV peptidase and aeruginolysin from Pseudomonas aeruginosa. Clara Rettenmaier (Vienna, Austria) is thanked for laboratory assistance during her training. This study was supported in part by grants from public and private agencies from the European Community, Spain, and Catalonia (FP7-PEOPLE-2011-ITN-290246 “RAPID,” FP7-HEALTH-2012-306029-2 “TRIGGER;” BFU2011-25902, BFU2012-32862, BIO2013-49320-EXP, MDM-2014-0435; and 2014SGR9) and an FPU Ph.D. fellowship (AP2010-3799) from the Spanish Ministry for Education, Culture and Sport to I.G.-F. The Department of Structural Biology of IBMB is a “María de Maeztu” Unit of Excellence from the Ministry of Economy and Competitiveness. We acknowledge the help provided by ESRF and ALBA synchrotron local contacts. Funding for travelling and synchrotron data collection was provided in part by ESRF.

§ 4. SUPPLEMENTAL REFERENCES

### § 5. SUPPLEMENTAL TABLES

<table>
<thead>
<tr>
<th>Plasmid</th>
<th>Construct</th>
<th>Parental DNA/plasmid</th>
<th>Forward-primer</th>
<th>Reverse-primer</th>
<th>Sequence$^c$</th>
<th>Fused tag</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>pECAM1</td>
<td>Secreted ECAM $^a$</td>
<td>E. coli K12 genomic DNA</td>
<td>CAATCATATGGGACACACAGAT AAGGCCCAAC</td>
<td>CAATCTCGAGGTTACGGCTCT GACAATCAGCAG</td>
<td>GHM-D$^{19}$,P$^{1653}$</td>
<td>N-t. His-tag</td>
<td>Cloned into pCRI-8b$^b$ at NdeI and XhoI restriction sites</td>
</tr>
<tr>
<td>pECAM2</td>
<td>ECAM $^c$</td>
<td>pECAM1</td>
<td>CAATCATATGGGCTCTTCAAGAA AAGGCCGACGT</td>
<td>CAATCTCGAGGTTACGGCTCT GACAATCAGCAG</td>
<td>GPM-A$^{40}$,P$^{1653}$</td>
<td>N-t. His-tag</td>
<td>Cloned into pCRI-8a at NcoI and XhoI restriction sites</td>
</tr>
<tr>
<td>pECAM3</td>
<td>ECAMΔN $^d$</td>
<td>pECAM1</td>
<td>CAATCATATGGGCGACCCGGTG ATGGCTGAACCTG</td>
<td>CAATCTCGAGGTTACGGCTCT GACAATCAGCAG</td>
<td>GHM-A$^{101}$,P$^{1653}$</td>
<td>N-t. His-tag</td>
<td>Cloned into pCRI-8b at NdeI and XhoI restriction sites</td>
</tr>
<tr>
<td>pECAM4</td>
<td>NIE-MG1</td>
<td>pECAM1</td>
<td>CAATCATATGGGCTCATACCAACCA AGGTC</td>
<td>CAATCTCGAGGTTAAATGGT TAAATGAGTAC</td>
<td>GAM-D$^{163}$,L$^{36}$</td>
<td>N-t. His-tag</td>
<td>Cloned into pCRI-8a at NcoI and XhoI restriction sites</td>
</tr>
<tr>
<td>pECAM5</td>
<td>MG9-NIE-MG1</td>
<td>pECAM1</td>
<td>CAATCATATGGGCTCTTCAAGAA AAGGCCGACGT</td>
<td>CAATCTCGAGGTTACGGCTCT GACAATCAGCAG</td>
<td>GHM-A$^{20}$,S$^{165}$</td>
<td>N-t. His-tag</td>
<td>Cloned into pCRI-8b at NdeI and XhoI restriction sites</td>
</tr>
<tr>
<td>pECAM6</td>
<td>Full-length ECAM+PBPI $^e$</td>
<td>E. coli K12 genomic DNA</td>
<td>CAATCATATGGGCGACCCGGTG ATGGCTGAACCTG</td>
<td>CAATCTCGAGGTTACGGCTCT GACAATCAGCAG</td>
<td>—</td>
<td>—</td>
<td>Cloned into pCRI-7b at NdeI and XhoI restriction sites</td>
</tr>
<tr>
<td>pECAM7</td>
<td>Full-length ECAM(strep)+PBPI $^e$</td>
<td>pECAM6</td>
<td>CAATTCCGAAAAATTGCTGTCACT CTGGTAC</td>
<td>TGGATGGCTCCAACCTCCA CTTCCGGCTGAC</td>
<td>M$^1$-$^3$P$^{1653}$ (ECAM) + M$^2$-$^7$Q$^{1653}$ (PBPI)</td>
<td>C-t. Strep-tag</td>
<td>Strep-tag at the C-terminus of ECAM</td>
</tr>
<tr>
<td>pECAM8</td>
<td>Full-length ECAM-PBPI (zeo/PTac)</td>
<td>pECAM7</td>
<td>GAATTAATTCATGAGGCTATAC ATATGGTGG</td>
<td>TACGAGCTGGCAAGCTACAG</td>
<td>M$^1$-$^3$P$^{1653}$ (ECAM) + M$^2$-$^7$Q$^{1653}$ (PBPI)</td>
<td>C-t. Strep-tag</td>
<td>For amplification of gene ble conferring zeocin resistance.</td>
</tr>
<tr>
<td>pECAM9</td>
<td>ECAM pre-TEV</td>
<td>pECAM1</td>
<td>CAATCATATGGGCTCTTCAAGAA AAGGCCGACGT</td>
<td>CAATCTCGAGGTTACGGCTCT GACAATCAGCAG</td>
<td>GSH-A$^{40}$,P$^{1653}$</td>
<td>N-t. His-tag</td>
<td>Cloned into pET28-a at NdeI and XhoI restriction sites</td>
</tr>
<tr>
<td>pECAM10</td>
<td>TEV-ECAM</td>
<td>pECAM10</td>
<td>CAATGTCGACAGCTGAAACCT GGGT</td>
<td>CAATGTCAGGCTCGCTTACCT GAAATGCAATTTGCTGAA</td>
<td>GSH-A$^{40}$,P$^{1653}$</td>
<td>N-t. His-tag</td>
<td>pECAM10 with TEV cleavage site (ENLYFQGVD) between L$^{42}$ and D$^{952}$</td>
</tr>
<tr>
<td>pECAM12</td>
<td>ECAM-R$^{162}$G</td>
<td>pECAM2</td>
<td>GTGACATCACAACCCAGCT</td>
<td>GACACATGGAAGAGAGAT</td>
<td>GSH-A$^{40}$,P$^{1653}$</td>
<td>N-t. His-tag</td>
<td>pECAM2 mutant R$^{162}$G</td>
</tr>
<tr>
<td>pECAM13</td>
<td>ECAMABRD $^f$</td>
<td>pECAM2</td>
<td>CAATGTCGACAGCAGCAGACACAGGT GTGAAAGCCGCTACATACCACG</td>
<td>CAATGTCGACAGCAGCAGACACAGGT GTGAAAGCCGCTACATACCACG</td>
<td>GPM-A$^{40}$,P$^{1653}$</td>
<td>N-t. His-tag</td>
<td>pECAM2 with replacement of BRD segment S$^{491}$-R$^{586}$ with GGG</td>
</tr>
<tr>
<td>pECAM14</td>
<td>ECAMABRD $^g$</td>
<td>pECAM2</td>
<td>CAATGTCGACAGCAGCAGACACAGGT GTGAAAGCCGCTACATACCACG</td>
<td>CAATGTCGACAGCAGCAGACACAGGT GTGAAAGCCGCTACATACCACG</td>
<td>GPM-A$^{40}$,G$^{1516}$</td>
<td>N-t. His-tag</td>
<td>pECAM2 without RBD</td>
</tr>
</tbody>
</table>

$^a$ Peptide sequence of the expressed protein after fusion-tag removal. $^b$ Construct without the 17-residue signal peptide and the N-terminal lipoxin-cysteine (C$^{17}$; see UP76578) of mature ECAM. $^c$ For plasmids of the pCRI System, see (1). $^d$ Construct lacking the first 22 amino acids of secreted ECAM, which are predicted to be disorderd. $^e$ Construct spanning domains from MG7 to the C-terminus. $^f$ Construct with ECAM and PBPI in tandem amplified directly from genomic DNA. $^g$ ECAM lacking the bait-region domain (BRD). $^h$ ECAM lacking the receptor-binding domain (RBD).

Restriction-site sequences are underlined; italics indicate inserted or mutated nucleotide sequences; TEV, tobacco-etch virus peptidase.
**Suppl. Table 2** — Free cysteine content.

<table>
<thead>
<tr>
<th>ECAM species</th>
<th>$A_{412}$*</th>
</tr>
</thead>
<tbody>
<tr>
<td>nECAM</td>
<td>0.030</td>
</tr>
<tr>
<td>MA-ECAM</td>
<td>0.432</td>
</tr>
<tr>
<td>Trpsin-induced iECAM monomer</td>
<td>0.457</td>
</tr>
</tbody>
</table>

*Absorbance measured after reaction of 5,5'-dithiobis-(2-nitrobenzoic acid) with free cysteine thiol groups arising from broken thioesters. The only cysteine of the ECAM construct studied is the thioester cysteine.
### Suppl. Table 3 — Crystallographic data processing.

<table>
<thead>
<tr>
<th>Protein</th>
<th>iECAMAN (wild-type)</th>
<th>iECAMAN (Se-Met)</th>
<th>NIE-MG1 (wild-type)</th>
<th>NIE-MG1 (Se-Met)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Space group / wavelength (Å)</td>
<td>122.2 / 0.9763</td>
<td>122.2 / 0.9788 (peak)</td>
<td>P2₁ / 0.9793</td>
<td>P2₁ / 0.9687 (peak)</td>
</tr>
<tr>
<td>Cell constants (a, b, c, in Å)</td>
<td>86.7, 136.2, 172.8</td>
<td>85.8, 136.0, 171.5</td>
<td>50.49, 32.86, 58.32, β=102.23°</td>
<td>50.47, 32.75, 58.34, β=102.64°</td>
</tr>
<tr>
<td>No. of measurements / unique reflections</td>
<td>181,985 / 28,350</td>
<td>297,188 / 42,709</td>
<td>161,151 / 24,951</td>
<td>132,405 / 39,869</td>
</tr>
<tr>
<td>Resolution range (Å) (outermost shell)</td>
<td>48.0 – 2.70 (2.77 – 2.70)</td>
<td>47.6 – 2.90 (3.10 – 2.90)</td>
<td>57.0 – 1.60 (1.69 – 1.60)</td>
<td>42.1 – 1.70 (1.80 – 1.70)</td>
</tr>
<tr>
<td>Completeness (%) / anom. Correl.</td>
<td>99.4 (95.3) / -</td>
<td>99.3 (99.4) / 0.49</td>
<td>99.4 (97.6) / -</td>
<td>99.2 (97.0) / 0.36</td>
</tr>
<tr>
<td>R&lt;sub&gt;merge&lt;/sub&gt;</td>
<td>0.048 (0.654)</td>
<td>0.072 (0.834)</td>
<td>0.040 (0.367)</td>
<td>0.062 (0.438)</td>
</tr>
<tr>
<td>R&lt;sub&gt;free&lt;/sub&gt; / CC(1/2)</td>
<td>0.100 (1.85)</td>
<td>0.097 (1.03)</td>
<td>0.078 (0.839)</td>
<td>0.057 (0.471)</td>
</tr>
<tr>
<td>Average intensity / SigAno</td>
<td>0.278 (0.78)</td>
<td>0.278 (0.78)</td>
<td>0.278 (0.78)</td>
<td>0.278 (0.78)</td>
</tr>
<tr>
<td>B-Factor (Wilson) (Å²) / Average multiplicity</td>
<td>45.4 / 6.4 (6.6)</td>
<td>45.4 / 6.4 (6.6)</td>
<td>45.4 / 6.4 (6.6)</td>
<td>45.4 / 6.4 (6.6)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Protein</th>
<th>MG0-NIE-MG1 (wild-type)</th>
<th>MG0-NIE-MG1 (Se-Met)</th>
<th>MG0-NIE-MG1 (Se-Met)</th>
<th>MG0-NIE-MG1 (Se-Met)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Space group / wavelength (Å)</td>
<td>122.2 / 0.9763</td>
<td>122.2 / 0.9788 (peak)</td>
<td>P2₁ / 0.9793 (infection)</td>
<td>P2₁ / 0.9688 (remote)</td>
</tr>
<tr>
<td>Cell constants (a, b, c, in Å)</td>
<td>86.7, 136.2, 172.8</td>
<td>85.8, 136.0, 171.5</td>
<td>50.49, 32.86, 58.32, β=102.23°</td>
<td>50.47, 32.75, 58.34, β=102.64°</td>
</tr>
<tr>
<td>No. of measurements / unique reflections</td>
<td>181,985 / 28,350</td>
<td>297,188 / 42,709</td>
<td>161,151 / 24,951</td>
<td>132,405 / 39,869</td>
</tr>
<tr>
<td>Resolution range (Å) (outermost shell)</td>
<td>48.0 – 2.70 (2.77 – 2.70)</td>
<td>47.6 – 2.90 (3.10 – 2.90)</td>
<td>57.0 – 1.60 (1.69 – 1.60)</td>
<td>42.1 – 1.70 (1.80 – 1.70)</td>
</tr>
<tr>
<td>Completeness (%) / anom. Correl.</td>
<td>99.4 (95.3) / -</td>
<td>99.3 (99.4) / 0.49</td>
<td>99.4 (97.6) / -</td>
<td>99.2 (97.0) / 0.36</td>
</tr>
<tr>
<td>R&lt;sub&gt;merge&lt;/sub&gt;</td>
<td>0.048 (0.654)</td>
<td>0.072 (0.834)</td>
<td>0.040 (0.367)</td>
<td>0.062 (0.438)</td>
</tr>
<tr>
<td>R&lt;sub&gt;free&lt;/sub&gt; / CC(1/2)</td>
<td>0.0114 (1.099) / 0.999 (0.970)</td>
<td>0.086 (1.106) / 1.000 (0.981)</td>
<td>0.081 (1.110) / 1.000 (0.958)</td>
<td>0.088 (1.332) / 1.000 (0.942)</td>
</tr>
<tr>
<td>Average intensity / SigAno</td>
<td>0.278 (0.78)</td>
<td>0.278 (0.78)</td>
<td>0.278 (0.78)</td>
<td>0.278 (0.78)</td>
</tr>
<tr>
<td>B-Factor (Wilson) (Å²) / Average multiplicity</td>
<td>45.4 / 6.4 (6.6)</td>
<td>45.4 / 6.4 (6.6)</td>
<td>45.4 / 6.4 (6.6)</td>
<td>45.4 / 6.4 (6.6)</td>
</tr>
</tbody>
</table>

### Notes:
- a Values in parentheses refer to the outermost resolution shell.
- b For definitions, see Table 1 in (50).
- c According to Diederichs & Karplus (51).
- d According to Karplus & Diederichs (52).
- e According to program XDS/XSCALE. Average intensity is <i>d</i> / σ<i>d</i>. Friedel mates were kept separately.
## Suppl. Table 4 — Crystallographic refinement.

<table>
<thead>
<tr>
<th>Protein</th>
<th>wild-type nECAM (MG7-CUB(TED)-RBD)</th>
<th>wild-type NIE-MG1</th>
<th>Se-Met MG0-NIE-MG1</th>
<th>wild-type iECAM (NIE-MG1-to-MG5-MG6(BRD)-MG7-CUB(TED)-RBD + Trypsin)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Resolution range used for refinement (Å)</td>
<td>48.0 – 2.70</td>
<td>49.3 – 1.60</td>
<td>63.8 – 2.30</td>
<td>47.7 – 2.55</td>
</tr>
<tr>
<td>No. of reflections used (among them, test set)</td>
<td>28,349 (747)</td>
<td>24,939 (730)</td>
<td>17,298 (712)</td>
<td>65,374 (1,083)</td>
</tr>
<tr>
<td>Crystallographic R\textsubscript{free} (free R\textsubscript{free}) (^a)</td>
<td>0.190 (0.240)</td>
<td>0.182 (0.217)</td>
<td>0.213 (0.256)</td>
<td>0.190 (0.221)</td>
</tr>
<tr>
<td>No. of protein atoms / ions</td>
<td>4,920 / 2 Ni(^{2+})</td>
<td>1,650 / -</td>
<td>2,375 / -</td>
<td>11,224 / 3 Cl(^-) (occ. 0.5)</td>
</tr>
<tr>
<td>solvent molecules / ligands</td>
<td>96 / 5 GOL</td>
<td>200 / 1 GOL, 1ACT</td>
<td>40 / 5 GOL, 1 PG4, 1 PGE, 1 PEG, 1 PEG</td>
<td>105 / 1 GOL</td>
</tr>
<tr>
<td>Rmsd from target values</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bonds (Å) / angles (°)</td>
<td>0.010 / 1.16</td>
<td>0.010 / 1.08</td>
<td>0.010 / 1.23</td>
<td>0.009 / 1.05</td>
</tr>
<tr>
<td>B-Factor (Wilson) (Å(^2)) from data processing</td>
<td>74.5</td>
<td>30.4</td>
<td>61.2</td>
<td>81.0</td>
</tr>
<tr>
<td>Aver. B-factors protein atoms (Å(^2))</td>
<td>80.4</td>
<td>31.5</td>
<td>77.2</td>
<td>111.1</td>
</tr>
<tr>
<td>Residue main-chain conformational angle and side-chain rotamer analysis (^b)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>favored regions / outliers / all residues /</td>
<td>622 / 1 / 637</td>
<td>203 / 0 / 206</td>
<td>290 / 1 / 300 /</td>
<td>1,380 / 6 / 1,438</td>
</tr>
<tr>
<td>bad rotamers (%) / MolProbity score</td>
<td>5.6 / 2.09 (97(^\text{th}) percent.)</td>
<td>4.0 / 1.71 (74(^\text{th}) percent.)</td>
<td>3.4 / 2.04 (89(^\text{th}) percent.)</td>
<td>4.3 / 2.17 (93(^\text{th}) percent.)</td>
</tr>
</tbody>
</table>

\(^a\) For definitions, see Table 1 in (50). \(^b\) According to MOLPROBITY (53). GOL, glycerol; PG4, tetraethylene glycol; PGE, triethylene glycol; 1PE, pentaethylene glycol; PEG, diethylene glycol; and ACT, acetate.
§ 6. SUPPLEMENTAL FIGURES

Suppl. Fig. 1 — Functional and biophysical studies in vitro (All the figures are representatives of more than three independent experiments). (A) Size-exclusion chromatography (SEC). Elution volumes of ~11mL correspond to monomeric species. iECAM was obtained by trypsin induction and shows an incipient fraction of dimeric inhibitor (~10mL). Note the significant difference in elution volumes between induced (green chromatogram) and native (blue and red chromatogram) monomers. (B) Native PAGE of nECAM (lane 1), MA-ECAM (lane 2), monomeric (lane 3) and dimeric (lane 4) trypsin-induced monomeric (black arrow) and dimeric (grey arrow) iECAM. (C) Thermal-shift assays. Light purple, yellow and pink background strips indicate different thermal transitions. (D) Circular dichroism spectra. (E) Native PAGE of nECAM (lane 1), and iECAM induced by thermolysin (lane 2), chymotrypsin (lane 3) and subtilisin (lane 4). Induced species (black arrow) migrate faster than native species (grey arrow; see also (B)). (F) SDS-PAGE and N-terminal sequence analysis of trypsin-induced iECAM crystals (see §1.10). (G) SEC analysis of peptidase-induced iECAM (by pancreatic elastase and thermolysin) and trypsin-induced iECAM-R\textsuperscript{162}G. Each chromatogram constitutes an independent experiment, which explains the slight differences in elution volumes. The yellow strip corresponds to monomeric iECAM cleaved at the MG0-NIE junction, the pink strip to dimeric iECAM, and the purple strip to domain MG0. (H) Samples from (G) analyzed by Western blot with an anti-His\textsubscript{6}-antibody. Lane 1 and lane 2 correspond to the fractions over yellow and purple strips, respectively. The corresponding fractions from (A) were used for the experiment with trypsin on wild-type ECAM (leftmost lanes). (I) SEC analysis of trypsin-induced iECAM dimerization over time. Pink and yellow strips indicate dimeric and monomeric species, respectively. (J) SDS-PAGE and N-terminal sequence analysis of trypsin-induced MA-ECAM (see §1.10).
Suppl. Fig. 2 — Functional and biophysical studies in vitro (All the figures are representatives of more than three independent experiments). (A) SDS-PAGE of intact (lane 1) and TEV peptidase-induced TEV-iECAM (lane 2). Arrows indicate intact protein (black), complex of the C-terminal fragment with TEV peptidase (grey), and the N-terminal fragment (white) of TEV-iECAM as shown by peptide-mass fingerprinting. (B) Size-exclusion chromatography analyses of native TEV-ECAM induced with TEV-peptidase to yield TEV-iECAM. Elution volumes of ~11mL correspond to monomeric species. Note the significant difference in elution volumes between the induced (red chromatogram) and the native (blue chromatogram) monomer owing to large conformational rearrangement upon induction. (C) Analysis of TEV peptidase-induced TEV-iECAM by native-PAGE. Native protein (lane 1) was incubated with the peptidase (lane 2) or its lysine-methylated variant (lane 3). Arrows indicate native protein (black), complex of TEV-ECAM with peptidase (grey), and non-complexed peptidase-induced TEV-ECAM (white). Intact TEV-peptidase is bound by TEV-iECAM while lysine-methylated peptidase is not. (D) Zymogram of purified monomeric (lane 1) and dimeric (lane 2) trypsin-induced iECAM showing equivalent ECAM fragments with covalently-bound trypsin activity. (E) Western-blot analysis of Strep-tagged ECAM from E. coli inner membranes (lane 1), periplasm (lane 2) and outer membranes (lane 3). (F) SDS-PAGE of nECAM (lane 1), and iECAM induced by ulilysin (lane 2), trypsin (lane 3) and pancreatic elastase (lane 4).
Suppl. Fig. 3 — Inhibitory studies in vitro and biological function (All the figures are representatives of more than three independent experiments except otherwise stated). (A) Inhibition of trypsin activity against aldolase and thyroglobulin. Lane 1, substrate; lane 2, substrate plus peptidase; substrate plus peptidase and purified monomeric (lane 3) or dimeric (lane 4) previously trypsin-induced iECAM. Arrows indicate intact substrates. (B) Inhibition of subtilisin activity against fumarase and thyroglobulin. Lane 1, substrate; lane 2, substrate plus peptidase; substrate plus peptidase and 10× (lane 3) or 20× (lane 4) molar excess of directly added nECAM, respectively; substrate plus peptidase and 5× molar excess of nECAM (lane 5) or MA-ECAM (lane 6). (C) Inhibition of chymotrypsin activity against thyroglobulin. Lane 1, substrate; lane 2, substrate plus peptidase; substrate plus peptidase and 5× molar excess of nECAM (lane 3) and MA-ECAM (lane 4). A black arrow indicates the intact substrate. (D) Inhibition analysis of trypsin, subtilisin and chymotrypsin through ECAM and MA-ECAM in cleavage assays of E. coli envelope extracts. Inhibition was monitored through the decrease in A₅₇₈. See §1.2 and §1.3 for details of the respective experimental setups.(E) Concentration-dependent effect of trypsin on the survival of E. coli cells in vivo assessed by colony forming units for the wild type (WT), the ECAM deletion mutant (ΔECAM), the rescued deletion mutant (ΔECAM(+) and a negative control of the latter expressing green fluorescence protein instead of ECAM (ΔECAM(-)). Values are represented as means and standard deviations of three independent experiments with three replicates each. Statistical significance was determined with a Student’s paired t-test between the survivability of the WT and the deletion mutants (ΔECAM, ΔECAM (+), and ΔECAM (-)). Star (*) indicates statistical significance with p<0.05.
**Suppl. Fig. 4** — Fluorescence microscopy of trypsin-induced iECAM crystals. nECAM was reacted and crystallized with (A)(B) unlabeled trypsin (negative controls) and (C)(D) trypsin previously labeled with sulfosuccinimidyl-7-amino-4-methylcoumarin-3-acetate (see §1.5) and subsequently analyzed by fluorescence microscopy. Fluorescence in (D) reveals that trypsin is present in the crystals.
Suppl. Fig. 5 — Dimeric iECAM structure. (A) iECAM dimer with monomers in green and orange. The depicted front view is perpendicular to the crystallographic twofold relating both monomers. A trypsin molecule (magenta ribbon) has been placed in the central prey chamber for reference. (B) Top view of the dimer as surface model in stereo, with the domains of the upper monomer colored as in Fig. 1A, the lower monomer is in white. Entrance 1 is encircled. (C) Front and (D) lateral views of the dimer. Entrances 2 and 3 are encircled. In (D), the view is along the crystallographic dyad, which passes through entrance 3. (E) and (F), two orthogonal views of a cut-through surface model of the dimer in stereo colored as in (A), giving insight into the central prey chamber. Thioester residues in purple, bait-region domains in dark blue.
**Suppl. Fig. 6 — Domain topologies and structures.** Topology scheme depicting strands and helices of domains (A) MG (see (54)), (B) RBD, (C) TED, (D) CUB (see (55)) and (E) NIE. Extra elements distinguishing RBD and NIE from MG are shown in magenta. TED is shown along the pseudo-six-fold axis with the view on the entry face of the toroid, i.e. the one at which the inner helices enter the toroid. A red arrow pinpoints the thioester loop at the beginning of helix TED-α2. For extent of each of the secondary-structure elements, see Suppl. Fig. 7. (F) Superposition in stereo of MG0, MG2 and MG3 onto MG1; domain coloring as in Fig. 1A. (G) Superposition of MG4, MG5, MG6, and MG7 onto MG1; domain coloring as in Fig. 1A. (H) Superposition in stereo of RBD (gray) onto MG1 (orange). (I) Superposition in stereo of NIE (gray) onto MG1 (orange).
Suppl. Fig. 7 — Repetitive secondary structure elements of ECAM. Helices and strands are displayed as rods and arrows, respectively. The best structure available for each domain (coloring as in Fig. 1A) was taken for the assignment (see Suppl. Table 4: for domain MG0, structure MG0-NIE-MG1; for domains NIE and MG1, structure NIE-MG1; for domains MG2-MG6[BRD], structure iECAM; for domains MG7-CUB[TED]-RBD, structure nECAMΔN, except for TED, for which iECAM was also considered). The 17-residue signal peptide (SP) and the first 22 residues of the mature enzyme (part of A; see Suppl. Table 1, plasmid pECAM2) were absent in the studied proteins. The bait region and the thioester bond are pinpointed. Flexible undefined regions in the iECAM structure are shown with double strikethrough in red. Strand length of MG4-β3 and -β4 was assigned based on SEAM (36). Nomenclature of the seven canonical strands of MG domains was based on MG1 (see also Suppl. Fig. 6A). Split strands and additional strands are characterized by primed and double-primed numbers.
Suppl. Fig. 8 — Cryo-electron microscopy studies. Plot of Fourier space correlation against resolution (in 1/Å) for the cryo-EM 3D reconstructions of nECAM, monomeric iECAM and dimeric iECAM (see §1.12). The final calculated resolutions are, respectively, 16.3Å, 17.4Å, and 14.5Å.
Suppl. Fig. 9 — Ancillary crystal structures. Ribbon-type plot in two orthogonal orientations of the experimental structures of (A) nECAMΔN showing MG7 (red), CUB (dark green), TED (purple; the side chains of thioester segment C$^{1187}$-Q$^{1190}$ are shown for reference), and RBD (dark grey); and (B) MG0-NIE-MG1 (sienna, light grey and orange) superposed onto NIE-MG1 (yellow and chartreuse). The two structures reveal a virtually identical relative arrangement of domains NIE and MG1 despite completely different crystal environments. For crystallographic data on these structures, see Suppl. Tables 3 and 4 and §1.10 and §1.11.
Suppl. Fig. 10 — Superposition of iECAM onto hα2M. (A) Front, (B) lateral, and (C) back views in stereo. Domain colors of, respectively, hα2M and iECAM: MG1+MG5, medium purple and sandy brown; MG2+MG3+MG6+MG7, pink and yellow; CUB-TED, purple and tan; MG4, magenta and brown; BRD, dark blue and sea green; and RBD, orange and red. iECAM domain NIE is shown in white. The respective domains are labeled. The BRD anchor points upstream (“UP”) and downstream (“DOWN”) of the missing bait-region residues are marked by blue (hα2M) and green (iECAM) arrows. The rotation of iECAM CUB-TED around its anchor point with MG7 (purple straight arrow) required to match the corresponding hα2M domains is indicated by a purple curved arrow. The relative matching rotation of MG4 is shown by a cyan curved arrow and that of MG1+MG5 by a yellow curved arrow. The relative translation observed for RBD is depicted as an orange straight arrow. hα2M lacks A, MG0 and NIE, which accounts for shorter protein length in metazoan α2Ms.
7. LEGENDS TO SUPPLEMENTAL MOVIES

**Suppl. Movie 1** – Movie to illustrate a possible transition between native and induced ECAM. nECAM, shown in the colors of Fig. 1A and the orientation of Fig. 3A, undergoes major conformational rearrangement for all its constituting domains except MG1-L-MG2 and MG5-MG6, which leads to a more compact structure, in particular when the first N-terminal domain, MG0, is removed. The thioester residues are depicted for their side chains and highlighted by green glow.