VIP Molecular Biology

DOI: 10.1002/anie.201108015

The Crystal Structure of Human α_2-Macroglobulin Reveals a Unique Molecular Cage

*Aniebrys Marrero, Stephane Duquerroy, Stefano Trapani, Theodoros Goulas, Tibisay Guevara, Gregers R. Andersen, Jorge Navaza, Lars Sottrup-Jensen, and F. Xavier Gomis-Rüth**

Dedicated to Professors Wolfram Bode and Robert Huber on the occasion of their 70th and 75th birthdays

Protein inhibitors provide a major means for the regulation of proteolytic enzymes. Human α_2-macroglobulin (α_2M) is an abundant 720 kDa homotetrameric blood-plasma proteinase inhibitor of broad specificity synthesized in the liver.[1–3] It was first isolated by Cohn and co-workers in 1946,[4] and each of its four subunits is a multidomain 1451 residue glycosylated molecule. α_2M targets a large variety of proteinases of the four major classes, which form 1:1 or 2:1 complexes with tetrameric α_2M.[2,3] The α_2M tetramer operates through a unique irreversible "venus flytrap" mechanism (Supporting Information, Figure S1a), which "entraps" proteinases in the inhibitor tetramer that remain still active against small substrates or inhibitors.[1,2] Entrapping results from large conformational changes elicited upon cleavage by the proteinase within a multitarget "bait region"[1,2] and exposition of a buried cysteine–glutamine thioester bond located within a "thioester domain" (TED).[5] The thioester is then readily cleaved by surface lysine residues of the prey proteinase, which becomes covalently bound to α_2M (Supporting Information, Figure S1a). Independently of bait-region hydrolysis, the buried thioester can also be cleaved by small nucleophiles, such as methylamine (MA). In such forms of human α_2M, the bait regions are intact but the molecule no longer inhibits proteinases.[3,6,7] Induction by proteinases or MA results in similar undisclosed conformational transitions of the α_2M tetramer, as revealed to some extent by low-resolution electron microscopy[8] and electrophoretic mobility studies.[9] As a result, a slow (hereafter "native") and a fast (hereafter "induced") form of α_2M are distinguished. Induction further causes the C-terminal domain of each monomer, the "receptor binding domain" (RBD), to become exposed on the tetramer surface. Thereby, induced α_2M is recognized by specific cell-surface receptors, which do not recognize native α_2M. The complex is internalized by endocytosis and degraded in the lysosomes, thus ensuring clearance of the inhibitor and its cargo from the circulation within minutes of complex formation. Owing to its general inhibitory activity against proteinases, α_2M has universal physiological functions and imbalance of its activity contributes to several major human diseases, such as Alzheimer's disease, AIDS, inflammatory diseases, tumor growth and progression, and cardiovascular disease.[10–14]

To shed light on the molecular features of the intricate working mechanism of α_2M, we determined the crystal structure of human MA-induced α_2M (α_2M-MA) to 4.3 Å resolution (see the Supporting Information). The human α_2M-MA monomer consists of 11 domains (Figure 1a) and is shaped like a distorted equilateral triangular prism with a side of approximately 110 Å (Figure 1b,c). The monomer is tethered by 13 disulfide bonds, two of which contribute to intermolecular cross-linking to form dimers (see below), and contains eight N-linked solvent-exposed carbohydrate groups. The front surface is convex (Figure 1c, left; this view is taken as a reference hereafter) and the back is concave (Figure 1c, right). The first seven domains, termed macroglobulin-like (MG) domains, MG1–MG7,[15,16] are approximately 110-residue, seven-stranded antiparallel β-sandwiches comprising a three- and a four-stranded sheet (Figure 1d; see Supporting Information, Figure S1b for an approximate secondary-structure assignment), as previously described for recombinant α_2M MG2.[15] The first six MG domains are arranged as a compact ellipsoidal one-and-a-half-turn right-handed super-

[*] Dr. A. Marrero,[#] [+] Dr. T. Goulas, T. Guevara, Prof. Dr. F. X. Gomis-Rüth
Proteolysis Lab, Molecular Biology Institute of Barcelona, CSIC, Barcelona Science Park
c/Baldiri Reixac, 15-21, 08028 Barcelona (Spain)
E-mail: fxgr@ibmb.csic.es
Dr. S. Duquerroy[+]
Institut Pasteur, Unité de Virologie Structurale, Département de Virologie, 25 Rue du Dr. Roux, 75724 Paris Cedex 15 (France)
and
CNRS URA 3015
and
Université Paris-Sud, 91405 Orsay Cedex (France)
Dr. S. Trapani[+]
Université de Montpellier 2, Centre de Biochimie Structurale
29 Rue de Navacelles, 34090 Montpellier (France)
and
CNRS UMR5048
and
Université de Montpellier 1, INSERM UMR 1054 (France)
Dr. J. Navaza
CEA, Institut de Biologie Structurale "Jean-Pierre Ebel", UMR 5075
41 Rue Jules Horowitz, 38027 Grenoble Cedex 1 (France)
Prof. Dr. G. R. Andersen, Prof. Dr. L. Sottrup-Jensen
Department of Molecular Biology, Aarhus University
Gustav Wieds Vej 10C, 8000 Aarhus (Denmark)
[#] Present address: Laboratory of Organic Chemistry, HCI F330, ETH Hönggerberg
8093 Zürich (Switzerland)
[+] These authors contributed equally to this work.

Supporting information for this article (experimental procedures) is available on the WWW under http://dx.doi.org/10.1002/anie.201108015.

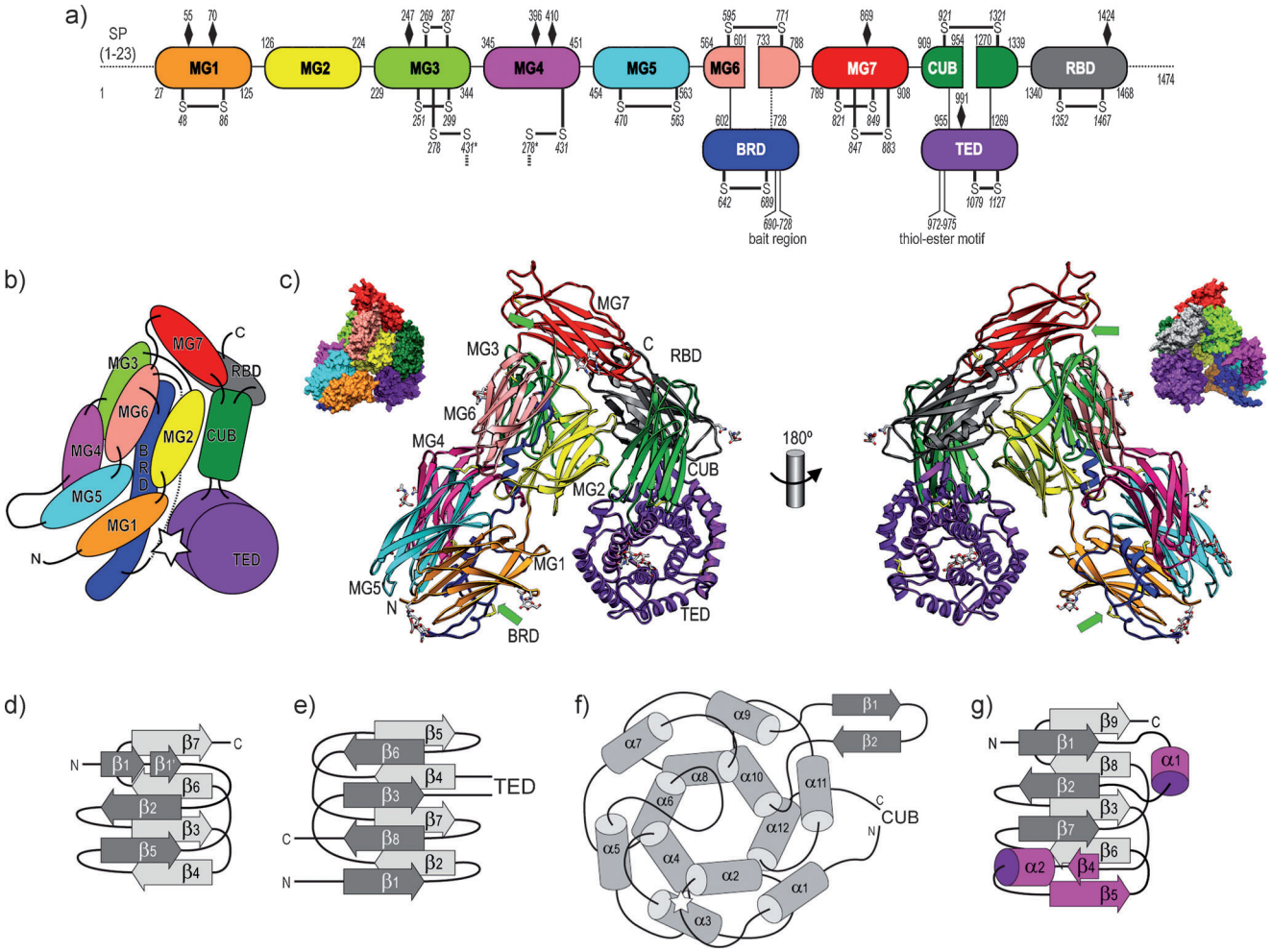

Figure 1. a) Domain organization of human α_2M depicting the flanking residues of each domain, the disulfide bonds, the N-linked glycosylation sites (as black rhombuses), and the insertion sites of the bait region and the thioester. b) Approximate arrangement of the domains constituting the α_2M-MA monomer in front view. The thioester site is shown as a white star. The flexible bait region is depicted as a dashed line. c) Connolly surface and Richardson-type plot of an α_2M-MA monomer in front view (convex face; left) and in back view (concave face; right). Green arrows pinpoint the anchor points of the flexible bait region. d–g) Topology scheme with secondary-structure element nomenclature of the MG domains (d), the CUB domain (e), the TED domain (f), and the RBD domain (g). The elements distinguishing RBD from MGs are shown in magenta.

helix (Figures 1 b,c and Supporting Information, Figure S2a). As a result of this arrangement, MG3–MG6 encircles a central ellipsoidal opening, hereafter "entrance 1" (Supporting Information, Figure S2d). Access to this opening from the bulk solvent is modulated by the glycan chain attached to N_{247} (residue numbering according to Uni*P*rot P01023, subtract 23 for numbering of mature α_2M). Downstream of MG6, MG7 is inserted, and it closes the superhelix like a plug and forms the upper limit of the molecule (Figure 1 b,c). Following MG7, a 116-residue all-β CUB domain is found, which consists of two four-stranded antiparallel β-sheets (Figure 1 c,e) and attaches laterally to domain MG2. Inserted between strands β3 and β4 of CUB, domain TED features a 315-residue helical domain with an α/α-toroid topology (Figure 1 f and Supporting Information, Figure S2b). It consists of six concentric α-hairpins arranged as a six-fold α-propeller around a central axis. This architecture gives rise to a thick disc with two parallel flat sides, the entry and the exit side, which are shaped by the N- and C-termini of the ring of inner helices, respectively. TED includes a β-hairpin between α9 and α10 (Figure 1 f), and it is placed below CUB (Figure 1 b,c). In this way, it lies adjacent to MG1 and MG2, thus contributing to the compact overall structure of the α_2M-MA monomer. In the native state of the α_2M monomer, a thioester formed between the side chains of the cysteine and the glutamine residues within TED segment C_{972}-G_{973}-E_{974}-Q_{975} is buried. In the present MA-induced state, this bond is cleaved and the resulting free residues, C_{972} and MA-modified Q_{975}, protrude from the entry surface of TED (Supporting Information, Figure 2 b,c). In accordance with the high-resolution structure of C3 fragment C3d (see Supporting Information, Table S1), these two amino acids flank a 1,4-tight turn, whose intervening residues, G_{973} and E_{974}, are widely conserved across thioester-containing proteins.[7,17] Overall, such a thioester-site architecture is likely present in other α_2M orthologs and paralogs (for discussion, see the Supporting Information).

After forming TED, the polypeptide chain rejoins CUB, which leads to the C-terminal 129-residue RBD (Figures 1 a–c,g)—also known as MG8.[15] This domain lies behind CUB, close to MG3, and contacts TED and CUB. RBD features the third type of β-sandwich architecture found within α_2M-MA, as previously described for the module isolated from bovine α_2M:[18] it is a variant of the MG fold, into which a β-α-β motif has been inserted (Figure 1 g), thus giving rise to a four-stranded and a five-stranded twisted sheet. In response to induction of α_2M, RBD protrudes from the molecular surface into the circulation. This, in turn, explains why RBD was only observed in one of the four monomers in the asymmetric unit of the crystal (chain D), presumably in a physiologically non-relevant position. In the remaining three monomers, RBD was solvent-exposed and flexible.

The ten aforementioned domains form a large cavity on the back face of the monomer, which shelters the "bait-region domain" (BRD; Q_{602}–T_{728}; see Figure 1 c, right and Supporting Information, Figure S2a). This is an extended flexible domain, which is inserted into domain MG6 and spans approximately 85 Å. It lines most of the vertical dimension of the cavity and interacts with MG1–MG3 and MG5. At T_{628}, the BRD adopts a more compact structure and includes a long loop (C_{642}–D_{665}), which protrudes from the monomer structure and is hereafter termed "tethering loop" (Supporting Information, Figure S2a). Segment T_{628}–C_{689} is held together by a disulfide bond, C_{642}–C_{689}, and the latter cysteine is the last residue defined by electron density for this domain. At this point, the polypeptide enters the 39-residue bait-region segment (P_{690}–T_{728};[19]), which is disordered. The bait region is intact in α_2M-MA—as it is in native α_2M—and the distance between C_{689} and the next residue defined by electron density, E_{729}, can be easily covered by the intervening 39 flexible residues. This intrinsic flexibility is in accordance with a function as a universal bait for endopeptidases that must be freely accessible and prepared to adapt to different types of active-site cleft.[5] In addition, C_{689} is in the lumen of the cavity, approximately at the same height as MG1 and MG5, and E_{729} is on the outer surface of the protomer, close to MG7 (Figure 1 c). This indicates that the flexible bait region lines the concave monomer surface and runs out across entrance 1, which is also likely to be accessible in native α_2M.

In the circulation, functional α_2M is a homotetramer, which is also the content of the asymmetric unit in our structure (Figure 2 a–c). The tetramer has overall dimensions of approximately 210 Å in length and approximately 140 Å in width and depth, and all the N-linked glycan chains of each monomer are surface-located and protrude into the bulk solvent, thus contributing to the high solubility of the protein in blood plasma. The particle can be presented in three

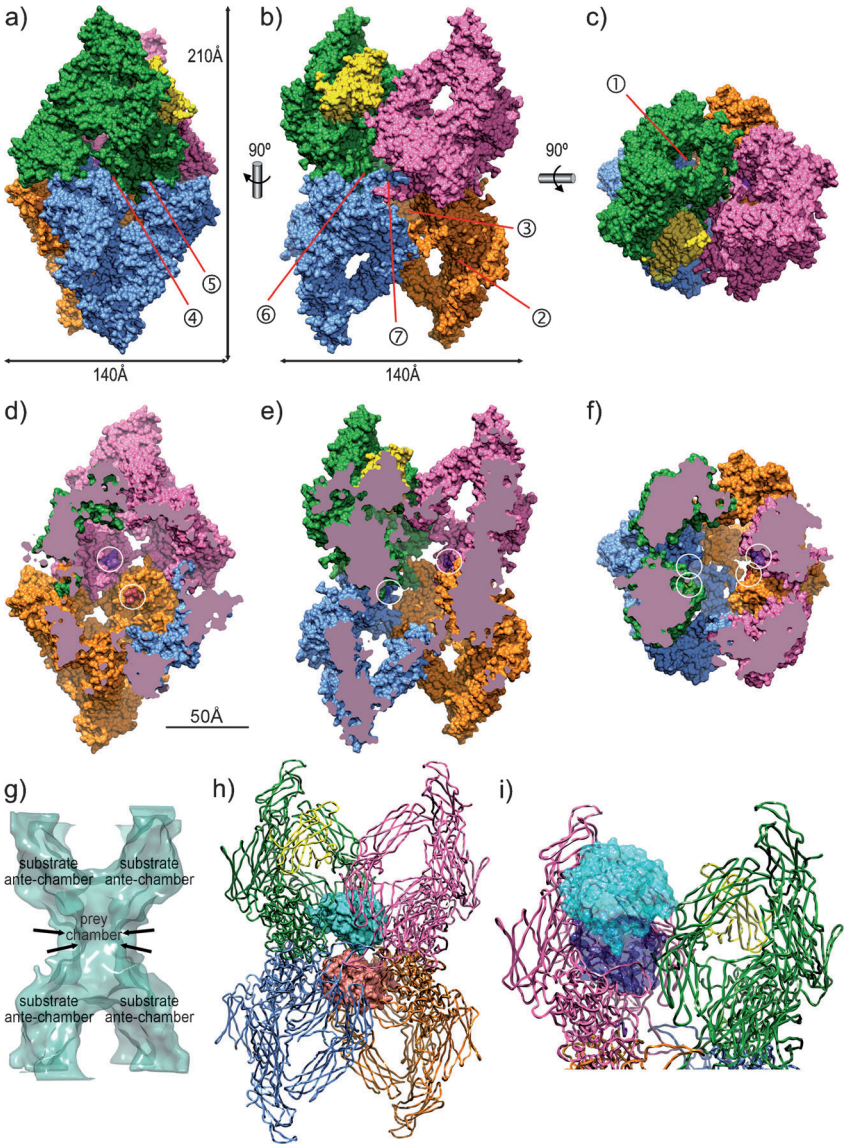

Figure 2. Surface plot of the α_2M-MA tetramer in a) X-view, b) H-view, and c) end view, with each monomer in a separate color. The RBD, only defined for monomer D, is further shown in yellow. The location of examples of entrances 1–3, as well as of regions engaged in tetramerization, is denoted by circled numbers, see text for details. d–f) Surface representation of the tetramer as in (a–c) after clipping off the frontal part to allow insight into the central cavity. The thioester sites are encircled. g) Smoothed semi-transparent representation of the inner-cavity surface in H-view. The localization of the central prey chamber and the four substrate ante-chambers is indicated. The four arrows point to the narrowest part of the central cavity. h) Possible representation of the tetramer in H-view as a coil (coloring as in (a)) with two trapped molecules of HIV-1 proteinase as cyan and salmon pink solid surfaces. i) Proposed interaction of soybean trypsin inhibitor with porcine trypsin, shown as their respective semi-transparent surfaces (in cyan: inhibitor; in blue: proteinase) across an entrance 2 of the upper disulfide-linked dimer. View as in (h) after a vertical rotation of 180°.

orientations (Figure 2a–c;[20]). These views confirm that the overall arrangement of α_2M-MA is very similar to proteinase- and MA-induced tetramers as revealed by electron-microscopy and low-resolution crystallography.[8,20] Notably, the protrusions of the tetramer in its H-view (see Figure 4b in Ref. [8]) do not correspond to the RBDs, which are flexible (see above), but rather to the MG7 domains.

The tetramer is the result of the noncovalent association of two dimers, a top and a bottom dimer, which are made up by two monomers covalently linked by two symmetric intermolecular disulfide bonds between residues C_{278} of MG3 and C_{431} of MG4 (green plus pink monomers and blue plus orange monomers, respectively, in Figure 2a–c). This arrangement entails that each molecule (e.g. the green one in Figure 2a–c) has a "vicinal" protomer (in blue), an "opposite" protomer (in orange), and a "disulfide-linked" protomer (in pink).

Within each disulfide-linked dimer, the monomers are joined through their respective back surfaces (Figure 1c, right). The interaction surface is symmetrically shaped by the MG3 and MG4 domains, which meet at the top center. Additional, likewise symmetric contacts are made by TED with MG4. Each of the four TEDs interacts with its vicinal TED (Figure 2a, ④), MG1 (Figure 2a, ⑤), and BRD (Figure 2b, ⑥). In addition, each BRD interacts with its opposite BRD through the tethering loop (Figure 2b, ⑦). Each dimer provides access to the central cavity through entrance 1 of each monomer (see above, Figure 2c, ①) and through two other equivalent openings, hereafter "entrances 2" (Supporting Information, Figure S2e and Figure 2b, ②). Entrances 2 are each framed by MG2, MG3, MG7, CUB, and TED of one monomer. The respective disulfide-linked monomer further contributes through MG4. Entrance 2 is blocked by RBD in chain D (Figure 2a–c). Furthermore, tetramerization entails that a third major access to the central tetramer cavity is created in the form of "entrance 3" (Figure 2b, ③ and Supporting Information, Figure S2f). It is framed by the TED of one monomer, the tethering loop of the BRD of its vicinal protomer, and MG4 and the BRD of the disulfide-linked partner of the former. Access to this entrance from the bulk solvent is modulated by the glycan chains attached to N_{396} and N_{410}. All together, the entire tetrameric particle has four major entrances of each type (1, 2, and 3), making a total of twelve. These match the maximal dimensions estimated to prevent noncovalently bound preys from escaping rapidly but to enable small protein substrates or inhibitors to enter the central cavity (25–30 Å[1]).

α_2M-proteinase complexes and α_2M-MA apparently have similar overall shapes and dimensions, but in α_2M-MA the cavity is empty. It is surrounded by the four monomers, and it has an irregular shape and a volume of approximately 167000 $Å^3$. It consists of a central part of around 60 Å in diameter, termed the "prey chamber", which is delimited at the top and bottom by the segments of MG3 and MG4 participating in disulfide-mediated dimerization (Figure 2d–g). The prey chamber is restricted at its center, at half particle height, to a diameter of approximately 30 Å by loops provided by the four TEDs, which thus act as a cavity-narrowing belt. This chamber is complemented by additional elongated volumes provided by the concave face of each of the four monomers, called "substrate ante-chambers" (Figure 2g). Overall, the prey chamber has space for up to two proteinase molecules of 20–30 kDa, one above and one below the narrowing belt, that is, one for each disulfide-linked dimer (Figure 2h). This is consistent with the maximal 2:1 stoichiometry of inhibition determined for proteinase binding by tetrameric α_2M.[2] Based on its size, the chamber could also accommodate a single, larger molecule of up to around 80–90 kDa if the loops of the cavity-narrowing belt were folded back towards the inner wall of the cavity.

Bound proteinases may process small substrates and inhibitors, and also small proteins of 6–9 kDa.[1] These would access the cavity lumen and thus the active-site of the bound proteinase through any of the twelve entrances and reach a substrate ante-chamber. The largest protein inhibitor reported to interact with a bound prey is soybean trypsin inhibitor, a 20 kDa molecule that slowly inhibits α_2M-bound trypsin.[21] This inhibitor could not pass through any of the entrances 1–3. However, computational modeling showed that binding of the active-site cleft of the proteinase by the reactive-site loop of the Kunitz-type inhibitor, as found in the crystal structure of the complex,[22] could occur through an entrance 2, while the main body of the inhibitor remained outside the α_2M cage (Figure 2i). This would require rearrangement of the segments framing the entrance, which would explain the long time required for the inhibitory reaction to reach equilibrium. Similarly, the finding that noncovalently bound proteinases, for example, methylated trypsin, may slowly dissociate from the α_2M-proteinase complex[23] would likewise imply extensive "breathing" motions of segments forming the walls of the transformed tetramer.

In conclusion, the structural features determined for α_2M provide the molecular basis of the function of a unique inhibitory snap trap. Structural comparisons enabled us to further construct a working model of native human α_2M and discuss its mechanistic implications (see Supporting Information).

Received: November 14, 2011
Published online: January 31, 2012

Keywords: blood plasma inhibitors · conformational change · macroglobulin · proteinase inhibitors · structure elucidation

[1] A. J. Barrett, P. M. Starkey, *Biochem. J.* **1973**, *133*, 709.
[2] L. Sottrup-Jensen, *J. Biol. Chem.* **1989**, *264*, 11539.
[3] J. Travis, G. S. Salvesen, *Annu. Rev. Biochem.* **1983**, *52*, 655.
[4] E. J. Cohn, L. E. Strong, W. L. Hughes, D. L. Mulford, J. N. Ashworth, M. Melin, H. L. Taylor, *J. Am. Chem. Soc.* **1946**, *68*, 459.
[5] L. Sottrup-Jensen, O. Sand, L. Kristensen, G. H. Fey, *J. Biol. Chem.* **1989**, *264*, 15781.
[6] L. Sottrup-Jensen, T. E. Petersen, S. Magnusson, *FEBS Lett.* **1980**, *121*, 275.
[7] P. B. Armstrong, J. P. Quigley, *Dev. Comp. Immunol.* **1999**, *23*, 375.
[8] S. J. Kolodziej, T. Wagenknecht, D. K. Strickland, J. K. Stoops, *J. Biol. Chem.* **2002**, *277*, 28031.

[9] F. van Leuven, J. J. Cassiman, H. Van den Berghe, *J. Biol. Chem.* **1981**, *256*, 9016.
[10] S. B. Athauda, E. Ido, H. Arakawa, M. Nishigai, H. Kyushiki, Y. Yoshinaka, T. Takahashi, A. Ikai, J. Tang, K. Takahashi, *J. Biochem.* **1993**, *113*, 742.
[11] J. F. Woessner, Jr., *Ann. N. Y. Acad. Sci.* **1999**, *878*, 388.
[12] J. Schaller, S. S. Gerber, *Cell. Mol. Life Sci.* **2011**, *68*, 785.
[13] A. J. Saunders, R. E. Tanzi, *Exp. Neurol.* **2003**, *184*, 50.
[14] C. J. Liu, *Nat. Clin. Pract. Rheumatol.* **2009**, *5*, 38.
[15] N. Doan, P. G. Gettins, *Biochem. J.* **2007**, *407*, 23.
[16] B. J. C. Janssen, E. G. Huizinga, H. C. A. Raaijmakers, A. Roos, M. R. Daha, K. Nilsson-Ekdahl, B. Nilsson, P. Gros, *Nature* **2005**, *437*, 505.
[17] L. Sottrup-Jensen in *The plasma proteins. Structure, function and genetic control*, *Vol. 5*, 2nd ed. (Ed.: F. W. Putnam), Academic Press, Orlando, FL, **1987**, p. 191.
[18] L. Jenner, L. Husted, S. Thirup, L. Sottrup-Jensen, J. Nyborg, *Structure* **1998**, *6*, 595.
[19] L. Sottrup-Jensen, P. B. Lønblad, T. M. Stepanik, T. E. Petersen, S. Magnusson, H. Jörnvall, *FEBS Lett.* **1981**, *127*, 167.
[20] G. R. Andersen, T. J. Koch, K. Dolmer, L. Sottrup-Jensen, J. Nyborg, *J. Biol. Chem.* **1995**, *270*, 25133.
[21] J. G. Bieth, M. Tourbez-Perrin, F. Pochon, *J. Biol. Chem.* **1981**, *256*, 7954.
[22] H. K. Song, S. W. Suh, *J. Mol. Biol.* **1998**, *275*, 347.
[23] D. Wang, K. Wu, R. D. Feinman, *Arch. Biochem. Biophys.* **1983**, *222*, 117.

