

# Human $\alpha_2$ -macroglobulin is composed of multiple domains, as predicted by homology with complement component C3

Ninh DOAN and Peter G. W. GETTINS<sup>1</sup>

Department of Biochemistry and Molecular Genetics, University of Illinois at Chicago, Chicago, IL 60607, U.S.A.

Human  $\alpha_2$ M ( $\alpha_2$ -macroglobulin) and the complement components C3 and C4 are thiol ester-containing proteins that evolved from the same ancestral gene. The recent structure determination of human C3 has allowed a detailed prediction of the location of domains within human  $\alpha_2$ M to be made. We describe here the expression and characterization of three  $\alpha_2$ M domains predicted to be involved in the stabilization of the thiol ester in native  $\alpha_2$ M and in its activation upon bait region proteolysis. The three newly expressed domains are MG2 (macroglobulin domain 2), TED (thiol ester-containing domain) and CUB (complement protein subcomponents C1r/C1s, urchin embryonic growth factor and bone morphogenetic protein 1) domain. Together with the previously characterized RBD (receptor-binding domain), they represent approx. 42% of the  $\alpha_2$ M polypeptide. Their expression as folded domains strongly supports the predicted domain organization of  $\alpha_2$ M. An X-ray crystal structure of MG2

shows it to have a fibronectin type-3 fold analogous to MG1–MG8 of C3. TED is, as predicted, an  $\alpha$ -helical domain. CUB is a spliced domain composed of two stretches of polypeptide that flank TED in the primary structure. In intact C3 TED interacts with RBD, where it is in direct contact with the thiol ester, and with MG2 and CUB on opposite, flanking sides. In contrast, these  $\alpha_2$ M domains, as isolated species, show negligible interaction with one another, suggesting that the native conformation of  $\alpha_2$ M, and the consequent thiol ester-stabilizing domain–domain interactions, result from additional restraints imposed by the physical linkage of these domains or by additional domains in the protein.

**Key words:** complement component C3, crystal structure, domain organization,  $\alpha_2$ -macroglobulin ( $\alpha_2$ M), receptor-binding domain, thiol ester.

## INTRODUCTION

Human  $\alpha_2$ M ( $\alpha_2$ -macroglobulin) is a homotetrameric plasma glycoprotein composed of polypeptide chains of 1451 residues [1]. It is found in all individuals, suggesting that it is essential for life, and it is highly abundant (1–2 g/l). In addition to being a pan-proteinase inhibitor capable of inhibiting proteinases of all mechanistic families by a remarkable physical sequestration mechanism [2] that has been likened to the closing of a Venus fly trap around its prey,  $\alpha_2$ M can tightly bind a number of growth factors in a non-covalent, reversible manner [3,4]. The massive conformational changes that result in proteinase entrapment are initiated by proteolytic cleavage of the bait region by the prey proteinase. The bait region is a stretch of approx. 30 residues located roughly in the middle of the polypeptide chain; it is a flexible region [5] containing many types of residues [6,7] that give it a wide spectrum of cleavage sites.

In addition to trapping the proteinase, the conformation changes result in activation of the thiol ester that is present between the side chains of residues Cys<sup>949</sup> and Glx<sup>952</sup> and in exposure of the four previously cryptic RBDs (receptor-binding domains), each of which is formed by the C-terminal 138 residues of the monomer (Figure 1) [8]. The activated thiol ester reacts rapidly with any available nucleophile [9]. If this is provided by a side chain from the attacking proteinase (usually a lysine  $\epsilon$ -amino group), it results in covalent linkage between the proteinase and  $\alpha_2$ M [10]. Exposure of the RBD results in tight binding to the receptor LRP (low-density-lipoprotein-receptor-related protein)

[11] and consequent internalization and, in some circumstances, signalling [12].

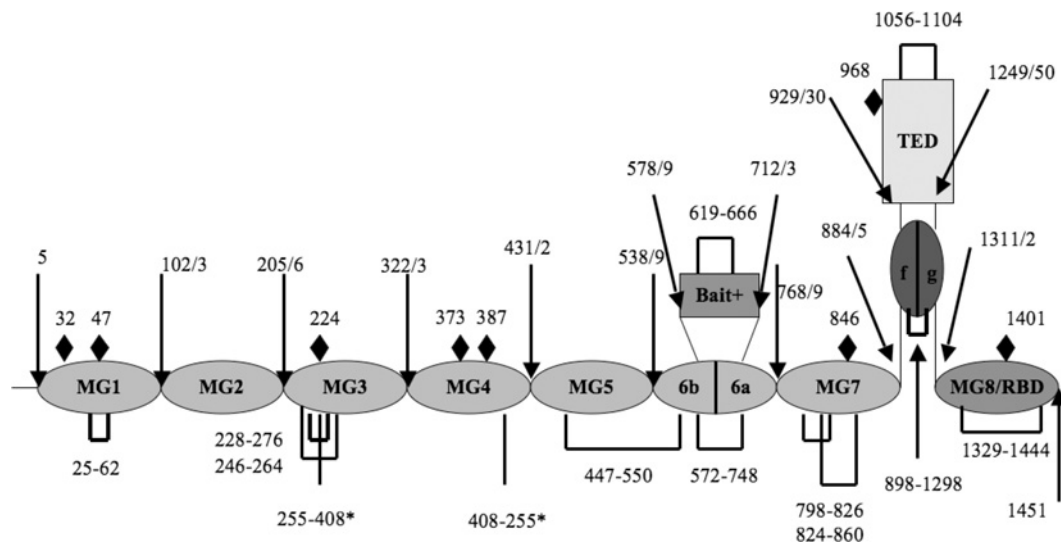
It was recognized many years ago that  $\alpha_2$ M is closely related to the complement proteins C3 and C4, with all probably arising from the same ancestral gene, though with features unique to each protein [13]. Thus C3 and C4 are monomers that contain a furin cleavage site that is absent from  $\alpha_2$ M, an excisable anaphylotoxin domain in C3, as well as a C-terminal domain designated C345 that lies beyond the RBD. Conversely,  $\alpha_2$ M contains a bait region domain that is located in a position equivalent to the C3a anaphylotoxin domain. The similarities in the remainder of the chain, however, are high enough to suggest that the remaining domains present in C3 are also present in  $\alpha_2$ M. Thus the recent determinations of the structures of C3, C3b and C3c by X-ray crystallography [14–17] provide a solid basis for predicting the location and structural type of the domains in  $\alpha_2$ M (Figure 1) and thus a means of examining domain–domain interactions in  $\alpha_2$ M that might be expected to exist in the native state.

We describe here the expression and characterization of three  $\alpha_2$ M domains that, together with RBD, are expected to interact intimately in native  $\alpha_2$ M and thereby play a role in the maintenance of the native conformation and of the integrity of the thiol ester. The ability to express each as a ‘well-behaved’ protein, especially for the spliced CUB (complement protein subcomponents C1r/C1s, urchin embryonic growth factor and bone morphogenetic protein 1) domain, supports the predicted domain organization. However, the absence of strong interactions between any of them suggests that the conformation of one or

Abbreviations used:  $\alpha_2$ M,  $\alpha_2$ -macroglobulin; CUB, complement protein subcomponents C1r/C1s, urchin embryonic growth factor and bone morphogenetic protein 1; DSC, differential scanning calorimetry; HSQC, heteronuclear single-quantum coherence; IAF, 5-iodoacetamidofluorescein; IPTG, isopropyl  $\beta$ -D-thiogalactoside; ITC, isothermal titration calorimetry; LRP, low-density-lipoprotein-receptor-related protein; MG1, macroglobulin domain 1; Ni-NTA, Ni<sup>2+</sup>-nitrilotriacetate; PEG, poly(ethylene glycol); RBD, receptor-binding domain; rmsd, root mean square deviation; TED, thiol ester-containing domain.

<sup>1</sup> To whom correspondence should be addressed (email pgettins@uic.edu).

The structural co-ordinates reported for  $\alpha_2$ -macroglobulin will appear in the Protein Data Bank under accession code 2P9R.



**Figure 1** Schematic representation of the domain organization of the  $\alpha_2M$  monomer, based on alignment with the structure of C3 [17]

MG designates a macroglobulin domain; CUB, a domain found in a number of developmentally regulated proteins [32]; 'bait', bait region domain. Intramolecular disulfide bonds are shown as solid black rectangular loops. The two inter-chain disulfides (Cys<sup>255</sup> and Cys<sup>408</sup>) [33,34] are shown as straight black lines and asterisked. The position of possible N-linked glycosylation sites are marked with black diamonds. Note that both MG6 and CUB are split domains whose primary structures are interrupted by other domains, the bait domain and TED respectively.

more as isolated domains, most likely the thiol ester domain, differs from native  $\alpha_2M$  and that constraints imposed by being part of the intact  $\alpha_2M$  tetramer result in a 'stressed' native state that is critical for maintaining these interactions.

## MATERIALS AND METHODS

### Expression of MG2 (macroglobulin domain 2)

cDNA encoding MG2 (residues 103–205) of human  $\alpha_2M$  was cloned into pQE-30 (Qiagen) and transformed into *Escherichia coli* SG13009 cells. Cells were grown in 2YT medium [1.6% (w/v) tryptone/1% (w/v) yeast extract/0.5% (w/v) NaCl] to a  $D_{600}$  (attenuance) of 0.6–0.8 and induced for 5 h at 37°C with 1 mM IPTG (isopropyl  $\beta$ -D-thiogalactoside). Cell lysate, obtained after sonication, was then loaded on to an Ni-NTA (Ni<sup>2+</sup>-nitrilotriacetate) Superflow (Qiagen) column, washed with 50 mM sodium phosphate buffer, 300 mM NaCl and 10 mM imidazole (pH 7.4) and eluted with 50 mM NaPO<sub>4</sub>, 300 mM NaCl and 250 mM imidazole (pH 7.4). Fractions collected from the Ni-NTA column were then dialysed against 20 mM Bis-Tris (pH 6.0) overnight, loaded on to an SP-Sepharose (sulfopropyl-Sepharose) HP column equilibrated with 20 mM Bis-Tris and eluted with a 500 mM to 1 M NaCl gradient. Selenomethionine-labelled MG2 was expressed by growing the cells in M9 medium followed, at a  $D_{600}$  of 0.3–0.5, by the addition of L-selenomethionine to a final concentration of 60 mg/l and L-isoleucine, L-leucine, L-lysine, L-phenylalanine, L-threonine and L-valine to final concentrations of 100 mg/l each. Induction and purification were the same as for non-labelled MG2. MG2 was concentrated to 10 mg/ml, dialysed against 20 mM Bis-Tris (pH 6.0) and kept frozen at –20°C until crystallization.

### Expression of TED (thiol ester-containing domain)

Expression of TED in *E. coli* and refolding of the material obtained from inclusion bodies yielded non-glycosylated TED with low solubility. As an alternative, wild-type and variant TEDs were expressed in yeast. cDNA encoding wild-type and

variant TEDs (residues 935–1242) of human  $\alpha_2M$  was cloned into pPICZ $\alpha$ A (Invitrogen) and transformed into *Pichia pastoris*. Cells were grown in BGC (10 g of yeast extract, 20 g of peptone, 20 ml of 50% glycerol, 3.4 g of yeast nitrogen base, 10 g of ammonium sulfate, 40  $\mu$ g of biotin and 100 mM potassium phosphate, pH 6.0, in 1 litre) medium at 30°C overnight. The cells were then transferred into BMC (as BGC, but with 5 ml of 100% methanol instead of glycerol) medium containing methanol for induction. Medium was supplemented with 0.5% methanol every 24 h. After 96 h, cells were spun down and discarded, while medium containing proteins of interest was loaded on to an Ni-NTA Superflow (Qiagen) column, washed with 50 mM sodium phosphate buffer, 300 mM NaCl and 10 mM imidazole (pH 7.4) and eluted with 50 mM sodium phosphate buffer, 300 mM NaCl and 250 mM imidazole (pH 7.4). Fractions collected from Ni-NTA column were then dialysed against 20 mM Tris and 500 mM NaCl (pH 8.0) buffer overnight and then concentrated to 6 mg/ml and stored at –20°C.

### Expression and purification of CUB

cDNA encoding CUB (residues 885–930 and 1250–1311) of human  $\alpha_2M$  was cloned as a single transcript into pQE-30 (Qiagen) and transformed into *E. coli* Origami cells (Novagen). Cells were grown in 2YT medium at 37°C to a  $D_{600}$  of 0.6–0.8 and induced for 5 h at 20°C with 0.5 mM IPTG. The cell lysate obtained after sonication was then loaded on to an Ni-NTA Superflow (Qiagen) column, washed with 50 mM sodium phosphate buffer, 300 mM NaCl and 10 mM imidazole (pH 7.4) and eluted with 50 mM sodium phosphate buffer, 300 mM NaCl and 250 mM imidazole (pH 7.4). Fractions collected from the Ni-NTA column were then dialysed against 20 mM Tris (pH 8.0) buffer overnight, loaded on to a Q-Sepharose HP column equilibrated with 20 mM Tris (pH 8.0) and eluted with a 0–500 mM NaCl gradient. Fractions from this Q-Sepharose HP column were then concentrated to 2 ml and loaded on to a Superdex-75 gel-filtration column to isolate monomers. Purified monomers were then concentrated to 10 mg/ml in 20 mM Tris (pH 8.0).

### Expression, purification and refolding of RBD (MG8)

RBD was expressed, purified and refolded as previously described [18].

### Structure determination of MG2

MG2 was crystallized in sitting drops from mother liquor containing 200 mM di-ammonium hydrogen citrate (pH 6.6), 11% (v/v) PEG [poly(ethylene glycol)] 3350. Crystals grew to  $100\ \mu\text{m} \times 80\ \mu\text{m} \times 80\ \mu\text{m}$  within 2 weeks. Crystals were flash-frozen in liquid nitrogen after cryoprotection in oil and data were collected at 100 K from a single crystal at the Argonne SER-CAT 22-BM beamline synchrotron. Three datasets at three different wavelengths were collected for selenomethionine-labelled MG2. Crystals had the space group of  $P4_32_12$  ( $a = 68.81\ \text{\AA}$ ,  $b = 68.81\ \text{\AA}$  and  $c = 120.07\ \text{\AA}$ ;  $1\ \text{\AA} = 0.1\ \text{nm}$ ), contained two molecules per asymmetric unit and diffracted to  $2.3\ \text{\AA}$ . Diffraction data were processed using HKL2000. Selenium MAD (multiwavelength anomalous dispersion) was used to obtain initial phases. Phase calculation and density modification were done by SOLVE and RESOLVE respectively. RESOLVE outputted the initial model and the remaining portions for the model were built manually using O. CNS was used for refinement and the refined structure of MG2 has a  $R$  factor of 24.44% and a  $R_{\text{free}}$  of 28.11% at the resolution of  $2.3\ \text{\AA}$ .

### Crystallization of TED

Plate-like crystals of C949A mutant TED were obtained from hanging drops containing 20% PEG 3350, 200 mM ammonium formate and 50 mM  $\text{MnCl}_2$  after 5 days. There were many attempts to optimize the crystallization conditions to obtain three-dimensional crystals, but all failed to give crystals thicker than  $20\ \mu\text{m}$ . Crystals diffracted poorly even at the synchrotron (Argonne SER-CAT 22-BM beamline). SDS/PAGE of crystals confirmed the presence of TED protein. Only TED variants failed to yield crystals.

### CD spectroscopy

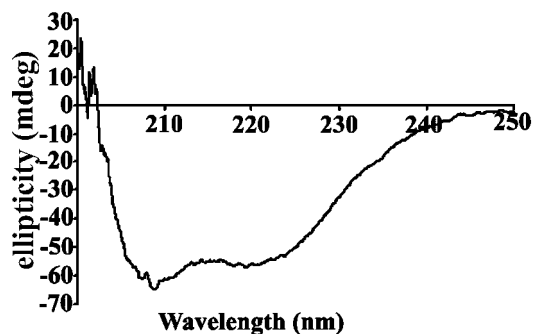
Proteins were dialysed against 20 mM potassium phosphate buffer and concentrated to  $1\ \mu\text{M}$  for data measurement using a Jasco J-710 spectropolarimeter.

### NMR spectroscopy

Variants of TED were dialysed against 50 mM sodium phosphate buffer and 100 mM NaCl, (pH 7.4) and concentrated to the specified concentrations. CUB and MG2 were dialysed against 20 mM sodium phosphate buffer (pH 7.4) and concentrated to specified concentrations. NMR data were recorded on a Bruker 900 MHz Avance NMR spectrometer equipped with a cryoprobe.

### Fluorescence spectroscopy

Fluorescence experiments were performed on a PTI Quantmaster instrument equipped with double monochromators on both the excitation and emission sides. For spectra of the TED species,  $1\ \mu\text{M}$  protein in 20 mM sodium phosphate buffer (pH 7.4) was excited at 280 nm and emission was recorded from 300 to 380 nm in 1 nm steps. For examining binding between C949A and Q952C TED and RBD, the absence of tryptophan residues in RBD was exploited to permit selective monitoring of the tryptophan residues in TED by excitation at 295 nm. To detect possible interactions between MG2 and C949A TED, C949A TED was excited at 280 nm while titrating with MG2, with correction made for the



**Figure 2** CD spectrum of TED, showing exclusive  $\alpha$ -helical secondary structure

The spectrum is of  $1\ \mu\text{M}$  TED in 20 mM potassium phosphate buffer (pH 7.4).

fluorescence of the MG2, which contains one tryptophan residue. The 'melting' of CUB ( $5\ \mu\text{M}$ ) was followed by change in tyrosine fluorescence as a function of temperature, with excitation at 280 nm and emission at 307 nm.

### ITC (isothermal titration calorimetry)

ITC measurements were performed with a VP-ITC calorimeter (Microcal). Portions of  $10\ \mu\text{M}$  C949A and Q952C TED were titrated with RBD or MG2 to a final concentration of  $400\ \mu\text{M}$ . Experiments were carried out in 20 mM sodium phosphate buffer (pH 7.4).

### DSC (differential scanning calorimetry)

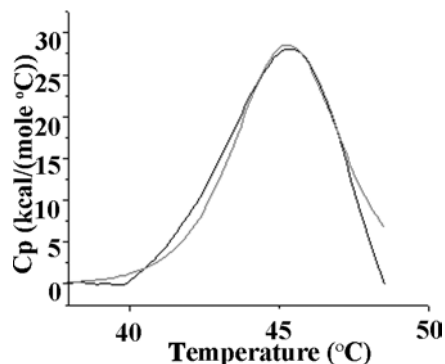
DSC measurements were performed using a VP-DSC calorimeter (Microcal). Melting temperatures were measured in 20 mM sodium phosphate buffer (pH 7.4) with proteins at 1 mg/ml for TED and 3 mg/ml for MG2. Data were analysed with Origin software (Microcal).

## RESULTS AND DISCUSSION

Prior to the present study, the only discrete domain definitely known to exist in human  $\alpha_2\text{M}$  was the 138-residue RBD at the extreme C-terminal end of the polypeptide chain [8, 19–21]. Using the revelation of the domain organization of C3 from its recent X-ray structure determination, together with the sequence alignment of  $\alpha_2\text{M}$  with C3 to give prediction of domain boundaries in  $\alpha_2\text{M}$ , we have now expressed three new domains from human  $\alpha_2\text{M}$  that represent an additional 513 residues out of the full-length chain of 1451 residues. Together with RBD, they represent approx. 42% of the whole polypeptide chain.

### Thiol ester domain

One of the consequences of proteolytic activation of C3 and C4 is the formation in each case of a discrete  $\sim 310$ -residue fragment, designated C3d and C4d respectively, that contains the thiol ester-forming residues CGEQ. X-ray structures of both C3d [22] and C4d [23] were determined even before those of the parent proteins C3 and C3b and they showed them to be all  $\alpha$ -helical domains composed of 12 helices, with the thiol ester-forming residues being located at the surface of the domain in a turn between the first and second helices. We expressed in *E. coli* the equivalent region of  $\alpha_2\text{M}$ , representing residues 935–1242 (designated TED). This was expressed in two forms: one in which the thiol ester-forming  $\text{Cys}^{949}$  had been mutated to alanine, thus precluding formation



**Figure 3** DSC-monitored unfolding of TED

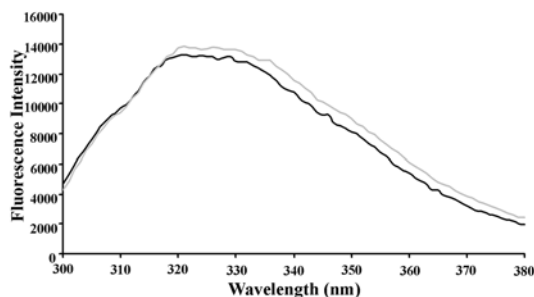
TED unfolds as a single domain with  $T_m$  of 45 °C (grey line). The fit to a single two-state unfolding is shown as a continuous black line. C949A TED (1 mg/ml) was in 20 mM sodium phosphate buffer (pH 7.4). 1 kcal = 4.184 kJ.

of a thiol ester, and one in which the Gln<sup>952</sup> had been mutated to cysteine, to permit formation of an approximately isosteric disulfide in place of the more labile thiol ester.

As expected from homology with C3d and C4d, as well as secondary-structure predictions of this region, the CD spectrum of this species was consistent with an all  $\alpha$ -helical domain (Figure 2). Analysis of the spectrum (<http://www.embl-heidelberg.de/~andrade/k2d/>) using reference sets of proteins and homopolymers of known structure gave 100%  $\alpha$ -helix [24,25]. DSC-monitored unfolding of TED gave a single peak (Figure 3), with a  $T_m$  (melting temperature) of 45 °C and a  $\Delta H_{\text{calorimetric}}/\Delta H_{\text{van't Hoff}}$  value of 1, confirming the presence of only a single domain [26]. The <sup>1</sup>H-NMR spectrum showed upfield-shifted methyl and aromatic resonances (results not shown) consistent with the species being a discrete folded domain. Also consistent with the homogeneity of the species was the ability to obtain plate-like crystals that were shown by SDS/PAGE to contain TED and to diffract, albeit poorly (results not shown).

The rationale for expressing one form of TED with C949A mutation and a second with the Q952C mutation was to be able to examine the effect on conformation of having no thiol ester (C949A variant) or a surrogate thiol ester (Q952C variant) that would be both resistant to facile cleavage by ambient nucleophiles and capable of being cleavable by reduction under controlled conditions. The presence of an intact disulfide bond between Cys<sup>949</sup> and Cys<sup>952</sup> in the unreduced Q952C variant (in addition to the buried natural disulfide) was confirmed by allowing the TED to react with the fluorescent cysteine-reactive compound IAF (5-iodoacetamidofluorescein). Overnight incubation of both native Q952C and C949A variants with IAF (200  $\mu$ M) resulted in no label incorporation, consistent with no free accessible cysteine residues. Following reaction of both variants with 0.5 mM 2-mercaptoethanesulfonate, chosen to preferentially reduce surface-accessible disulfides, the C949A variant still gave no label incorporation, whereas the Q952C gave two labels per molecule, supporting the presence of an intact surface-accessible disulfide in the unreduced form of the Q952C variant. Finally, it was confirmed by reaction with DTNB [5,5'-dithiobis-(2-nitrobenzoic acid)] that the denatured and fully reduced Q952C variant contained four free cysteine residues [27].

Since TED contains five tryptophan residues, two of which are positioned relatively close to the thiol ester-forming motif CGEQ in C3d and C4d and might therefore be very sensitive probes of conformational change in this region, we used tryptophan fluorescence to determine if there was a significant conformational



**Figure 4** Fluorescence spectra of TED species

Fluorescence emission spectra of C949A and Q952C (oxidized) TED in 20 mM sodium phosphate buffer (pH 7.4) showing very similar environments for the five tryptophan residues present.

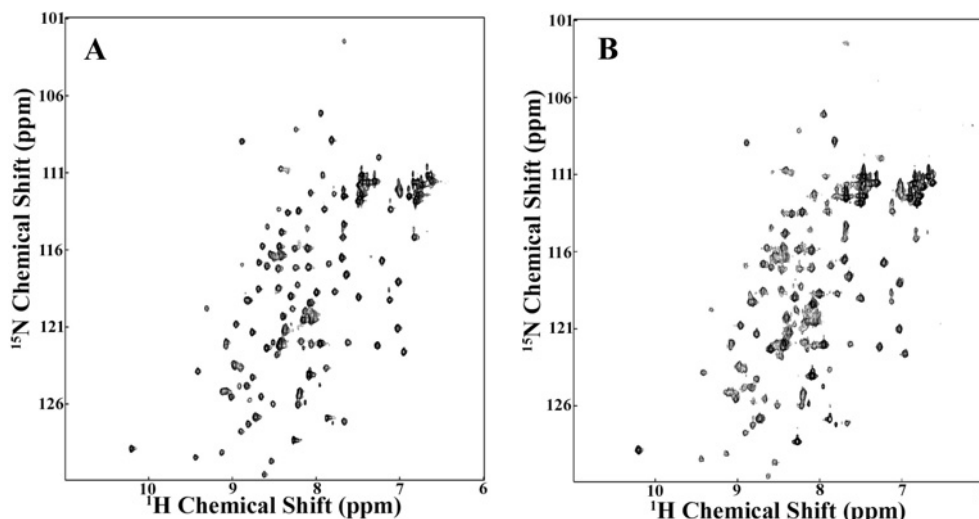
difference between the two TED species. The fluorescence spectrum of C949A TED had a broad maximum centred at approx. 322 nm, consistent with the expected buried location of most of the tryptophan residues (from homology with C3d and C4d) (Figure 4). Spectra of C949A TED and oxidized Q952C TED were nearly superimposable (Figure 4), while reduction of Q952C TED caused no significant perturbation. This suggested that the presence of a covalent linkage between residues 949 and 952 did not influence the conformation of TED in the context of the isolated domain.

### Structure of MG2

The various structures of C3 and C3 fragments all show the presence of eight domains with the fibronectin type-3 fold, each consisting of one three-strand and one four-strand antiparallel  $\beta$ -sheet stacked in an approximately parallel sandwich arrangement. These have been named MG (for macroglobulin) domains. The second macroglobulin domain (MG2) of human  $\alpha_2$ M is predicted to stretch from residue 103 to 205. We expressed a slightly longer version that had an N-terminal His tag. A <sup>15</sup>N-labelled sample gave a well-dispersed two-dimensional HSQC (heteronuclear single-quantum coherence) spectrum (Figure 5A), consistent both with a homogeneous folded domain and with the expected  $\beta$ -sheet secondary structure. This construct with His tag still attached gave crystals that diffracted to 2.3 Å. The structure was determined using anomalous dispersion from incorporated selenomethionine for phasing, to ensure a completely unbiased structure, and was found to be the expected  $\beta$ -sandwich (Figure 6A and Table 1) present in the MG domains of C3. Superpositioning of the present MG2 domain from  $\alpha_2$ M with that from MG2 in C3 showed that the structures are almost identical, with an rmsd (root mean square deviation) of 1.35 Å (Figure 6B).

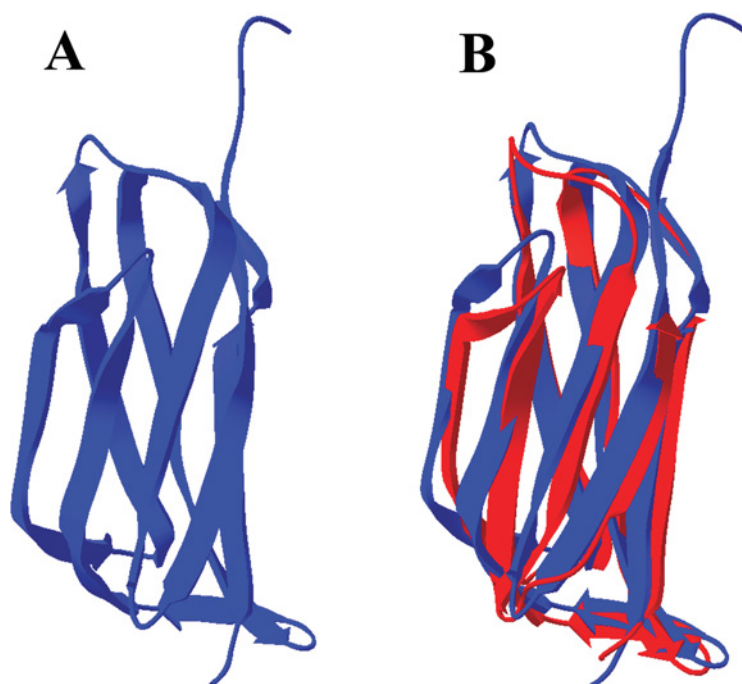
### Expression and properties of the CUB domain

The primary structure of the CUB domain in C3 (two stacked interdigitated four-stranded  $\beta$ -sheets) is interrupted by the entire TED (Figure 1). The ability to express the CUB domain of  $\alpha_2$ M as an isolated folded species by constructing an artificial single stretch of polypeptide composed of the two separate moieties would therefore be a rigorous test of the secondary-structure prediction for this region. In the tertiary structure of C3, the points of exit from CUB into TED (residue 962) and re-entry into CUB from the C-terminus of TED (residue 1269) lie close together in space (9.9 Å). Sequence alignment predicted a similar spliced arrangement for the CUB domain in human  $\alpha_2$ M [17], but with the difference that CUB from human  $\alpha_2$ M has an intra-domain



**Figure 5**  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectra of MG2

(A) Uniformly  $^{15}\text{N}$ -labelled MG2 ( $100\ \mu\text{M}$ ) alone; (B) uniformly  $^{15}\text{N}$ -labelled MG2 ( $100\ \mu\text{M}$ ) after addition of  $150\ \mu\text{M}$  of unlabelled TED. Samples were in 20 mM sodium phosphate buffer (pH 7.4). Spectra were recorded at 900 MHz.



**Figure 6** X-ray crystal structure of MG2 and comparison with MG2 from C3

(A) Ribbon diagram of MG2 from  $\alpha_2\text{M}$  determined in the present study (Protein Data Bank accession code 2P9R). (B) Super-positioning of MG2 from  $\alpha_2\text{M}$  (blue) with MG2 from bovine C3 (red).

disulfide, whereas there are no such disulfides in C3. The disulfide in  $\alpha_2\text{M}$  is between Cys<sup>898</sup> and Cys<sup>1298</sup> and thus covalently links the N- and C-terminal moieties of the domain. These residues align with residues 925 and 1317 in C3, which are close enough together to permit disulfide bond formation in  $\alpha_2\text{M}$  CUB without conformational change of the domain relative to that in C3. To permit expression of CUB from  $\alpha_2\text{M}$  as a single species, we constructed cDNA in which the encoded C-terminal residue of the stretch of CUB preceding TED (residue 929) was linked to the N-terminal residue following TED (residue 1250). This was

expressed in *E. coli* and refolded as described in the Materials and methods section.

The resulting soluble species gave a well-dispersed  $^1\text{H}$  NMR spectrum (Figure 7), with amide protons extending down to 10.0 p.p.m., and a number of shifted aliphatic resonances upfield from 0.7 p.p.m., consistent with a folded, structured domain. Also consistent with this was a single unfolding transition at 69 °C, monitored by change in tyrosine fluorescence as a function of temperature, and a CD spectrum with a single minimum at 215 nm consistent with  $\beta$  secondary structure (results not shown).

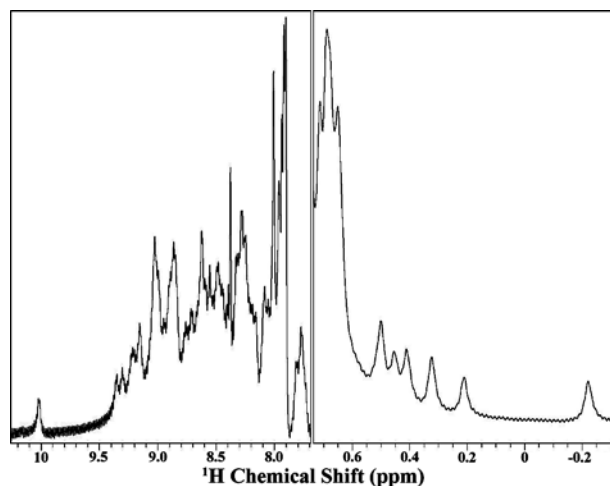
**Table 1** Data collection and refinement statistics for MG2

Parameter	Value
Data collection	
Space group	P4 <sub>3</sub> 2 <sub>1</sub> 2
Unit cell dimensions (Å)	<i>a</i> = 68.81, <i>b</i> = 68.81, <i>c</i> = 120.07
Wavelength (Å)	1.000
Resolution*	48.7–2.3 (2.38–2.3)
Measured reflections	310 398 (14 526)
Unique reflections	13 500 (1297)
<i>I</i> / $\sigma$ <i>I</i>	58.0 (4.7)
Redundancy	23 (11.2)
Completeness (%)	99.4 (98.6)
<i>R</i> <sub>sym</sub> †	6.1 (37)
Refinement	
Resolution	20–2.3
<i>R</i> ‡/ <i>R</i> <sub>free</sub> (%)	24.4/28.1
No. of atoms	
Protein	1612
Water	65
<i>B</i> factors (Å <sup>2</sup> )	
rmsd	39.8
Bond lengths (Å)	0.007
Bond angles (°)	1.35
Ramachandran plot	
Most favoured (%)	86.8
Additionally allowed (%)	9.6
Generously allowed (%)	2.4
Disallowed (%)	1.2

\* The numbers in parentheses are for the highest resolution shell.

†  $R_{\text{sym}} = \frac{\sum \sum |I_i - \langle I \rangle|}{\sum \langle I \rangle}$ , where  $\langle I \rangle$  is the mean intensity of the *N* reflections with intensities *I<sub>i</sub>* and common indices *hkl*.

‡  $R = \frac{\sum hkl ||F_{\text{obs}}| - k|F_{\text{calc}}||}{\sum hkl |F_{\text{obs}}|}$ , where *F<sub>obs</sub>* and *F<sub>calc</sub>* are the observed and calculated structure factors respectively. *R<sub>free</sub>* was calculated using 10% of all reflections, which were excluded from refinement.

**Figure 7** <sup>1</sup>H NMR spectrum of  $\alpha_2$ M CUB domain

Residues 885–930 were joined to residues 1251–1311 to generate a spliced, single-chain CUB domain. Low- and high-field regions of the one-dimensional spectrum of CUB are shown. Evidence for this spliced domain being well folded are the good spread of amide protons in the region from 10 to 8 p.p.m. and a number of upfield-shifted resonances between 0.7 and –0.2 p.p.m.

### Absence of domain–domain interactions

In the structure of C3, which can be considered to be equivalent to native  $\alpha_2$ M in terms of being the unactivated form, TED is in intimate contact with MG2, CUB and MG8 (Figure 8A). However, upon proteolytic activation of C3 to C3b, a series of major domain

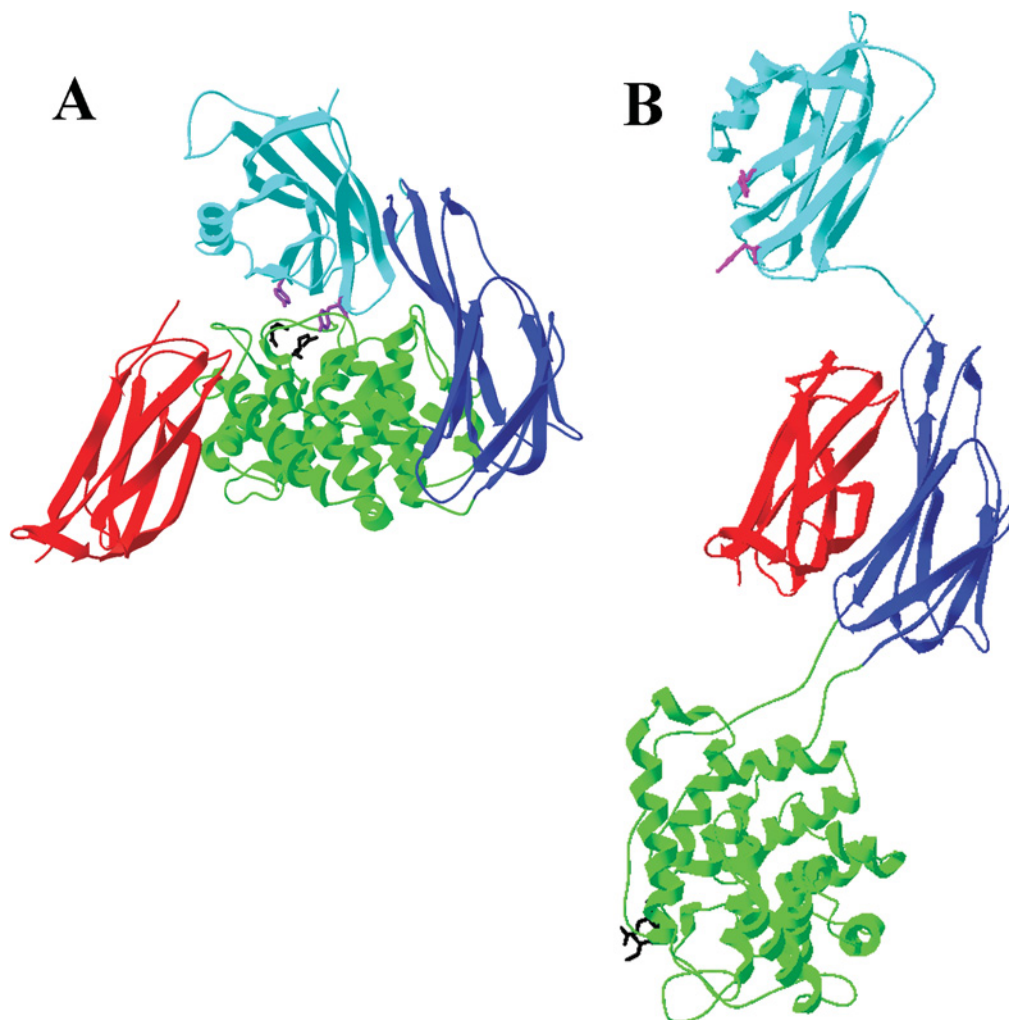
movements occurs that results, *inter alia*, in disruption of all of these contacts between TED and the domains MG2, CUB and MG8 (Figure 8B). In addition, the conformation of TED is significantly altered, such that it more closely resembles that of the fragment C3d. Whether the alteration in TED's conformation is a cause or consequence of the disruption of these interactions is not known. We were therefore interested in examining whether the equivalent domains from  $\alpha_2$ M could interact with one another as isolated species. For this purpose, we also expressed MG8, which is the RBD of  $\alpha_2$ M, which we have previously expressed and characterized structurally [18]. It should be noted that, although the TED used was glycosylated [approx. 3 kDa carbohydrate estimated from the MALDI–TOF (matrix-assisted laser-desorption ionization–time-of-flight) spectrum], the expected site of N-linked glycosylation is on the outer face of TED (bottom surface in Figure 8A), far from any predicted interfaces with MG2, CUB or MG8 and therefore very unlikely to interfere with domain–domain interaction.

As a qualitative means of detecting possible pairwise or higher-order interactions between these domains, native PAGE was used on the combinations TED–RBD, MG2–TED, RBD–CUB–TED–MG2, CUB–TED and CUB–RBD. In no instance was there evidence of a gel shift that might indicate a strong interaction. Subsequently, fluorescence spectroscopy and ITC were used to examine possible interactions between TED, RBD and MG2. Again, no evidence of binding interactions were observed, even when using a 20:1 molar excess of the titrant. As a last attempt to observe even weaker interactions, we used NMR spectroscopy at much higher protein concentrations to examine the interaction of TED with MG2. <sup>15</sup>N-labelled MG2 was titrated with unlabelled TED up to 1:1.5, without any significant perturbation of any of the backbone cross-peaks of MG2, or increases in line width suggestive of formation of a much-higher-molecular-mass species (Figure 5B). These negative findings suggest that the specific binding interactions that are present between these domains in intact  $\alpha_2$ M are greatly weakened as a result of the conformational changes that occur in TED subsequent to  $\alpha_2$ M activation.

Such a weakening, at least between TED and RBD, is not too surprising, since a principal function of the proteolytic activation of  $\alpha_2$ M and the subsequent conformational changes is to permit release of RBD from a tightly associated cryptic location within  $\alpha_2$ M so that it becomes accessible to its receptor, LRP. This contrasts with a demonstrated tightening of interactions elsewhere within  $\alpha_2$ M that makes it much harder to dissociate the non-covalently associated dimers following proteolysis within the bait region [28,29], as well as the finding of a shift in the dimer–tetramer equilibrium of the related human macroglobulin PZP (pregnancy zone protein) from mostly dimer to predominantly tetramer following bait region activation [30]. These observations, together with the transformation of tetrameric  $\alpha_2$ M into dimeric forms as a result of excision of residues within the 'bait domain' (Figure 1) that are immediately C-terminal to the exposed flexible bait region [31], suggest that the tightening involves strengthening of domain–domain interactions between the bait domains from monomers in opposing dimers. This may serve to help retain the proteinase within the  $\alpha_2$ M 'trap'.

### Conclusions

The ability to express three new domains from human  $\alpha_2$ M as folded, structured proteins, based on sequence alignment with complement C3, provides strong support for the predicted domain organization of  $\alpha_2$ M (Figure 1), both in terms of the existence of multiple discrete domains and in terms of their precise location and structure. However, the clear lack of a strong



**Figure 8 Relationship of MG2, TED, CUB and MG8 to one another in C3 and C3b**

(A) Domains in intact C3 [17]. (B) Domains following activation of C3 to C3b [15]. MG2 is in red, TED in green, CUB in blue and MG8(RBD) in cyan. The thiol ester-forming residues are shown in black and the conserved tyrosine pair in magenta.

interaction between TED and MG8(RBD), whether examined by fluorescence, ITC or gel shift, suggested that this critical interaction in native C3 and  $\alpha_2$ M must depend on additional constraints. Since no strong interaction was found between TED and either CUB or MG2, alone or together with MG8, this suggests that these restraints do not arise from strong binding interactions between TED and these domains, but rather from additional restraints that occur only in the context of native  $\alpha_2$ M or C3. Such an interpretation is consistent with a number of observations on both C3 and  $\alpha_2$ M, and is also more in keeping with the need for an essentially irreversible activation as a result of proteolysis.

We thank Dr Alex Dementiev for help with X-ray diffraction data collection and analysis, Dr Klavs Dolmer for help with protein expression and for Figure 1, and for helpful discussions, and Dr Steven Olson for comments on this paper. This work was supported by grant GM54414 from the NIH (National Institutes of Health). The ITC and DSC instruments were purchased with shared instrumentation grant S10 RR15958 from the NIH; the 900 MHz NMR spectrometer was purchased with funds from NIH grant P41 GM68944.

## REFERENCES

- 1 Sottrup-Jensen, L. (1989)  $\alpha$ -Macroglobulins. Structure, shape, and mechanism of proteinase complex formation. *J. Biol. Chem.* **264**, 11539–11542
- 2 Barrett, A. J. and Starkey, P. M. (1973) The interaction of  $\alpha_2$ -macroglobulin with proteinases. Characteristics and specificity of the reaction and a hypothesis concerning its molecular mechanism. *Biochem. J.* **133**, 709–724
- 3 Webb, D. J., Crookston, K. P., Hall, S. W. and Goniás, S. L. (1992) Binding of transforming growth factor- $\beta$ 1 to immobilized human  $\alpha_2$ -macroglobulin. *Arch. Biochem. Biophys.* **292**, 487–492
- 4 Ronne, H., Anundi, H., Rask, L. and Peterson, P. A. (1979) Nerve growth factor binds to serum  $\alpha_2$ -macroglobulin. *Biochem. Biophys. Res. Commun.* **87**, 330–336
- 5 Gettins, P. and Cunningham, L. W. (1986) Identification of  $^1$ H resonances from the bait region of human  $\alpha_2$ -macroglobulin and effects of proteases and methylamine. *Biochemistry* **25**, 5011–5017
- 6 Sottrup-Jensen, L., Lønblad, P. B., Stepanik, T. M., Petersen, T. E., Magnusson, S. and Jörnval, H. (1981) Primary structure of the 'bait' region for proteinases in  $\alpha_2$ -macroglobulin. *FEBS Lett.* **127**, 167–173
- 7 Sottrup-Jensen, L., Sand, O., Kristensen, L. and Fey, G. H. (1989) The  $\alpha$ -macroglobulin bait region. Sequence diversity and localization of cleavage sites for proteinases in five mammalian  $\alpha$ -macroglobulins. *J. Biol. Chem.* **264**, 15781–15789
- 8 Sottrup-Jensen, L., Gliemann, J. and Van Leuven, F. (1986) Domain structure of human  $\alpha_2$ -macroglobulin. Characterization of a receptor-binding domain obtained by digestion with papain. *FEBS Lett.* **205**, 20–24
- 9 Sottrup-Jensen, L., Petersen, T. E. and Magnusson, S. (1981) Trypsin-induced activation of the thiol esters in  $\alpha_2$ -macroglobulin generates a short-lived intermediate ('nascent'  $\alpha_2$ M) that can react rapidly to incorporate not only methylamine or putrescine but also proteins lacking proteinase activity. *FEBS Lett.* **128**, 123–126



- 10 Sottrup-Jensen, L. and Hansen, H. F. (1982) Nascent  $\alpha_2$ -macroglobulin-trypsin complex: incorporation of amines and water at the thiol esterified Glx residues of  $\alpha_2$ -macroglobulin during activation with trypsin. *Biochem. Biophys. Res. Commun.* **107**, 93–100
- 11 Kristensen, T., Moestrup, S. K., Gliemann, J., Bendtsen, L., Sand, O. and Sottrup-Jensen, L. (1990) Evidence that the newly cloned low density lipoprotein receptor-related protein (LRP) is the  $\alpha_2$ -macroglobulin receptor. *FEBS Lett.* **276**, 151–155
- 12 Bacskai, B. J., Xia, M. Q., Strickland, D. K., Rebeck, G. W. and Hyman, B. T. (2000) The endocytic receptor protein LRP also mediates neuronal calcium signaling via *N*-methyl-D-aspartate receptors. *Proc. Natl. Acad. Sci. U.S.A.* **97**, 11551–11556
- 13 Sottrup-Jensen, L., Stepanik, T. M., Kristensen, T., Lønblad, P. B., Jones, C. M., Wierzbicki, D. M., Magnusson, S., Domdey, P. B., Wetsel, R. A., Lundwall, Å. et al. (1985) Common evolutionary origin of  $\alpha_2$ -macroglobulin and complement components C3 and C4. *Proc. Natl. Acad. Sci. U.S.A.* **82**, 9–13
- 14 Wiesmann, C., Katschke, Jr, K. J., Yin, J. P., Helmy, K. Y., Steffek, M., Fairbrother, W. J., McCallum, S. A., Embuscado, L., DeForge, L., Hass, P. E. and van Lookeren Campagne, M. (2006) Structure of C3b in complex with CRIg gives insights into regulation of complement activation. *Nature* **444**, 217–220
- 15 Janssen, B. J. C., Christodoulou, A., McCarthy, A., Lambris, J. D. and Gros, P. (2006) Structure of C3b reveals conformational changes that underlie complement activity. *Nature* **444**, 213–216
- 16 Ajees, A., Gunasekaran, K., Narayana, S. V. L., Kotwal, G. J. and Murthy, H. M. K. (2006) The structure of complement C3b provides insights into complement activation and regulation. *Nature* **444**, 221–225
- 17 Janssen, B. J. C., Huizinga, E. G., Raaijmakers, H. C. A., Roos, A., Daha, M. R., Nilsson-Ekdahl, K., Nilsson, B. and Gros, P. (2005) Structures of complement component C3 provide insights into the function and evolution of immunity. *Nature* **437**, 505–511
- 18 Huang, W., Dolmer, K., Liao, X. B. and Gettins, P. G. W. (1998) Localization of basic residues required for receptor binding to the single  $\alpha$ -helix of the receptor binding domain of human  $\alpha_2$ -macroglobulin. *Protein Sci.* **7**, 2602–2612
- 19 Van Leuven, F., Marynen, P., Sottrup-Jensen, L., Cassiman, J.-J. and Van den Berghe, H. (1986) The receptor-binding domain of human  $\alpha_2$ -macroglobulin. Isolation after limited proteolysis with a bacterial proteinase. *J. Biol. Chem.* **261**, 11369–11373
- 20 Jenner, L., Husted, L., Thirup, S., Sottrup-Jensen, L. and Nyborg, J. (1998) Crystal structure of the receptor-binding domain of  $\alpha_2$ -macroglobulin. *Structure* **6**, 595–604
- 21 Huang, W., Dolmer, K., Liao, X. and Gettins, P. G. W. (2000) NMR solution structure of the receptor binding domain of human  $\alpha_2$ macroglobulin. *J. Biol. Chem.* **275**, 1089–1094
- 22 Nagar, B., Jones, R. G., Diefenbach, R. J., Isenman, D. E. and Rini, J. M. (1998) X-ray crystal structure of C3d: A C3 fragment and ligand for complement receptor 2. *Science* **280**, 1277–1281
- 23 van den Elsen, J. M. H., Martin, A., Wong, V., Clemenza, L., Rose, D. R. and Isenman, D. E. (2002) X-ray crystal structure of the C4d fragment of human complement component C4. *J. Mol. Biol.* **322**, 1103–1115
- 24 Yang, J. T., Wu, C.-S. C. and Martinez, H. M. (1986) Calculation of protein conformation from circular dichroism. *Methods Enzymol.* **130**, 208–269
- 25 Chang, C. T., Wu, C.-S. C. and Yang, J. T. (1978) Circular dichroic analysis of protein conformation: inclusion of the  $\beta$ -turns. *Anal. Biochem.* **91**, 13–31
- 26 Privalov, P. L. (1982) Stability of proteins. Proteins which do not present a single cooperative system. *Adv. Protein Chem.* **35**, 1–104
- 27 Ellman, G. L. (1959) Tissue sulfhydryl groups. *Arch. Biochem. Biophys.* **82**, 70–77
- 28 Pochon, F., Barray, M. and Delain, E. (1987) Functional  $\alpha_2$ -macroglobulin half-molecules induced by cadmium. *Biochem. Biophys. Res. Commun.* **149**, 488–492
- 29 Pochon, F., Barray, M. and Delain, E. (1989) Dissociation of  $\alpha_2$ -macroglobulin into functional half-molecules by mild acid treatment. *Biochim. Biophys. Acta* **996**, 132–138
- 30 Jensen, P. E. H. and Stigbrand, T. (1992) Differences in the proteinase inhibition mechanism of human  $\alpha_2$ -macroglobulin and pregnancy zone protein. *Eur. J. Biochem.* **210**, 1071–1077
- 31 Gettins, P. G. W., Hahn, K.-H. and Crews, B. C. (1995)  $\alpha_2$ Macroglobulin variants. A role for the bait region in tetramer formation. *J. Biol. Chem.* **270**, 14160–14167
- 32 Bork, P. and Beckmann, G. (1993) The CUB domain: a widespread module in developmentally regulated proteins. *J. Mol. Biol.* **231**, 539–545
- 33 Jensen, P. E. H. and Sottrup-Jensen, L. (1986) Primary structure of human  $\alpha_2$ -macroglobulin. Complete disulfide bridge assignment and localization of two intrachain bridges in the dimeric proteinase binding unit. *J. Biol. Chem.* **261**, 15863–15869
- 34 Oxvig, C., Gleich, G. J. and Sottrup-Jensen, L. (1994) Localization of disulfide bridges and free sulfhydryl groups in human eosinophil granule major basic protein. *FEBS Lett.* **341**, 213–217

Received 6 June 2007/22 June 2007; accepted 4 July 2007

Published as BJ Immediate Publication 4 July 2007, doi:10.1042/BJ20070764