ARTICLES

Structures of complement component C3 provide insights into the function and evolution of immunity

Bert J. C. Janssen[1], Eric G. Huizinga[1], Hans C. A. Raaijmakers[1], Anja Roos[2], Mohamed R. Daha[2], Kristina Nilsson-Ekdahl[3,4], Bo Nilsson[3] & Piet Gros[1]

The mammalian complement system is a phylogenetically ancient cascade system that has a major role in innate and adaptive immunity. Activation of component C3 (1,641 residues) is central to the three complement pathways and results in inflammation and elimination of self and non-self targets. Here we present crystal structures of native C3 and its final major proteolytic fragment C3c. The structures reveal thirteen domains, nine of which were unpredicted, and suggest that the proteins of the α2-macroglobulin family evolved from a core of eight homologous domains. A double mechanism prevents hydrolysis of the thioester group, essential for covalent attachment of activated C3 to target surfaces. Marked conformational changes in the α-chain, including movement of a critical interaction site through a ring formed by the domains of the β-chain, indicate an unprecedented, conformation-dependent mechanism of activation, regulation and biological function of C3.

The mammalian complement system mediates inflammation by generating anaphylatoxins to elicit chemotaxis and cell activation, and by promoting phagocytosis, degranulation and cell lysis. Its main functions include host defence against microorganisms, elimination of immune complexes and apoptotic cells, and facilitating adaptive immune responses[1,2]. The pivotal complement factor C3 is the convergence point for the classical, lectin and alternative pathways of complement activation[1,2].

C3 (187 kDa) emerged over 700 million years ago[3] and belongs to the α2-macroglobulin (α2M) family. These large (approximately 1,400–1,800 residue) proteins are characterized by homologous sequence features, including a unique thioester motif and a central, highly variable part that is functionally important[4]. Its members, such as the complement factors C3, C4 and C5, the proteinase inhibitor α2M and the insect and nematode thioester-containing proteins (TEPs), have important roles in the immune response in metazoans, considerably pre-dating the emergence of immunoglobulins[5].

C3 interacts with a large number of complement factors (for example, proteases, receptors and regulators) and non-complement proteins (for example, viral and bacterial proteins) via distinct binding sites. The function of C3 is regulated by conformational changes induced by sequential proteolytic cleavages (Fig. 1). Cleavage of mature C3 is mediated by enzyme complexes (that is, convertases), and generates the anaphylatoxin C3a (9 kDa) and the major fragment C3b (177 kDa)[6] (Fig. 1c). This step exposes a hidden thioester[7] and multiple cryptic binding sites in C3b for interacting complement proteins[8]. Nascent C3b is able to bind covalently to cell and other target surfaces via the exposed thioester[9]. Amplification of complement activity is achieved by association of surface-bound C3b and pro-enzyme factor B[10], yielding the short-lived ($t_{1/2} \approx 90$ s)[11] C3 convertase C3bBb of the alternative pathway. Additional cleavages in the α-chain of C3b (generating iC3b) mediated by factor I in association with soluble[12,13] or membrane-bound[13,14] co-factors prevent further convertase formation and profoundly alter the function of the protein[8]. The first two cleavages release C3f (2 kDa)[15] and the third cleavage in the remaining iC3b liberates C3c (135 kDa) from the target-bound C3dg (40 kDa) fragment[16].

Structure determination of C3c and C3

We determined the structures of human, native C3 (3.3 Å resolution) and its main proteolytic fragment C3c (2.4 Å resolution), which comprises 72% of C3. First, we determined the structure of C3c purified from outdated human plasma. The structure was solved by single-isomorphous replacement with anomalous scattering (SIRAS) and two-wavelength anomalous dispersion (MAD) phasing but required multi-crystal averaging[17] with nine partial masks to obtain an interpretable map. Second, we determined the structure of full-length C3 purified from fresh human plasma. We solved this structure by molecular replacement. Initial positioning of C3c, or any of its fragments, was unsuccessful. However, the α6–α6 helical structure of C3d[18] (18% of C3, in a protein otherwise rich in β-strands) was positioned correctly by Phaser[19]. Subsequently, domains of C3c were added gradually, yielding 85% of the model, which was completed by model building. The final refined models had R and R_{free} values of 22.3 and 27.9% (C3c) and 23.6 and 29.5% (C3); see Methods and Table 1; see also Supplementary Fig. 1 and Supplementary Table 1.

Domain organization

C3 consists of two chains—β (residues 1–645) and α (residues 650–1641) (ref. 20)—that together form 13 domains, whereas C3c consists of three chains (the β-chain and two fragments of the α-chain) that form ten domains (Fig. 1a–c; see also Supplementary Fig. 2). Surprisingly, one domain is formed by parts of both the β- and

[1]Crystal and Structural Chemistry, Bijvoet Center for Biomolecular Research, Faculty of Science, Utrecht University, Padualaan 8, 3584 CH Utrecht, The Netherlands. [2]Department of Nephrology, Leiden University Medical Center, 2300 RC Leiden, The Netherlands. [3]Department of Clinical Immunology, University Hospital, SE-751 85 Uppsala, Sweden. [4]Department of Chemistry and Biomedical Sciences, University of Kalmar, SE-391 82 Kalmar, Sweden.

α-chains. In addition, each chain forms six domains by itself. Eight domains (5.5 of β and 2.5 of α) exhibit a fibronectin-type-3-like core fold (Supplementary Fig. 2d, e); however, no sequence homology is apparent among these domains. In analogy to immunoglobulin domains, we refer to them as macroglobulin (MG) domains. Residues 1–534 form five MG domains, MG1–MG5 (Supplementary Fig. 3). Residues 535–577 form one half, denoted $MG6^{\beta}$, of the β/α intertwined MG6 domain. Together the chain of six MG domains forms 1.5 turns of helical coil as in a key ring (see domains of β-chain in Fig. 1c). Residues 578–645 of the β-chain exit from MG6. They loop downwards and through the ring, forming three helices in an extended configuration and one β-strand that aligns with the first strand of MG1. Three aromatic residues from the third helix and the β-strand of this linker domain, called LNK, form a small hydrophobic core. This structural element is wedged in between domains MG1, MG4 and MG5. The α-chain starts with the anaphylatoxin (ANA) domain (residues 650–726), which when cleaved off forms the C3a fragment. The structure of this domain is similar to the crystal structure of C3a (ref. 21), although the amino-terminal α-helical region was not resolved in that structure (Supplementary Fig. 3h). The subsequent scissile bond Arg 726–Ser 727 sits in a surface-exposed, disordered loop (residues 720–729). Following this, there is an extended loop (residues 730–745) that connects ANA to MG6. Loop 727–745 forms the N-terminal region of the cleaved α-chain, denoted α′, in C3b; we refer to this loop as α′NT. Residues 746–806 form $MG6^{\alpha}$, which complements $MG6^{\beta}$ in the formation of MG6. $MG6^{\alpha}$ is followed by a seventh MG domain, MG7 (residues 807–911). The polypeptide chain that is excised by factor I to yield fragments C3dg and C3f is located between MG7 and an eighth MG domain, MG8 (residues 1331–1474). MG8 corresponds to the receptor-binding domain in α2M; as expected, the structure of MG8 is homologous to this domain[22] (Supplementary Fig. 3p).

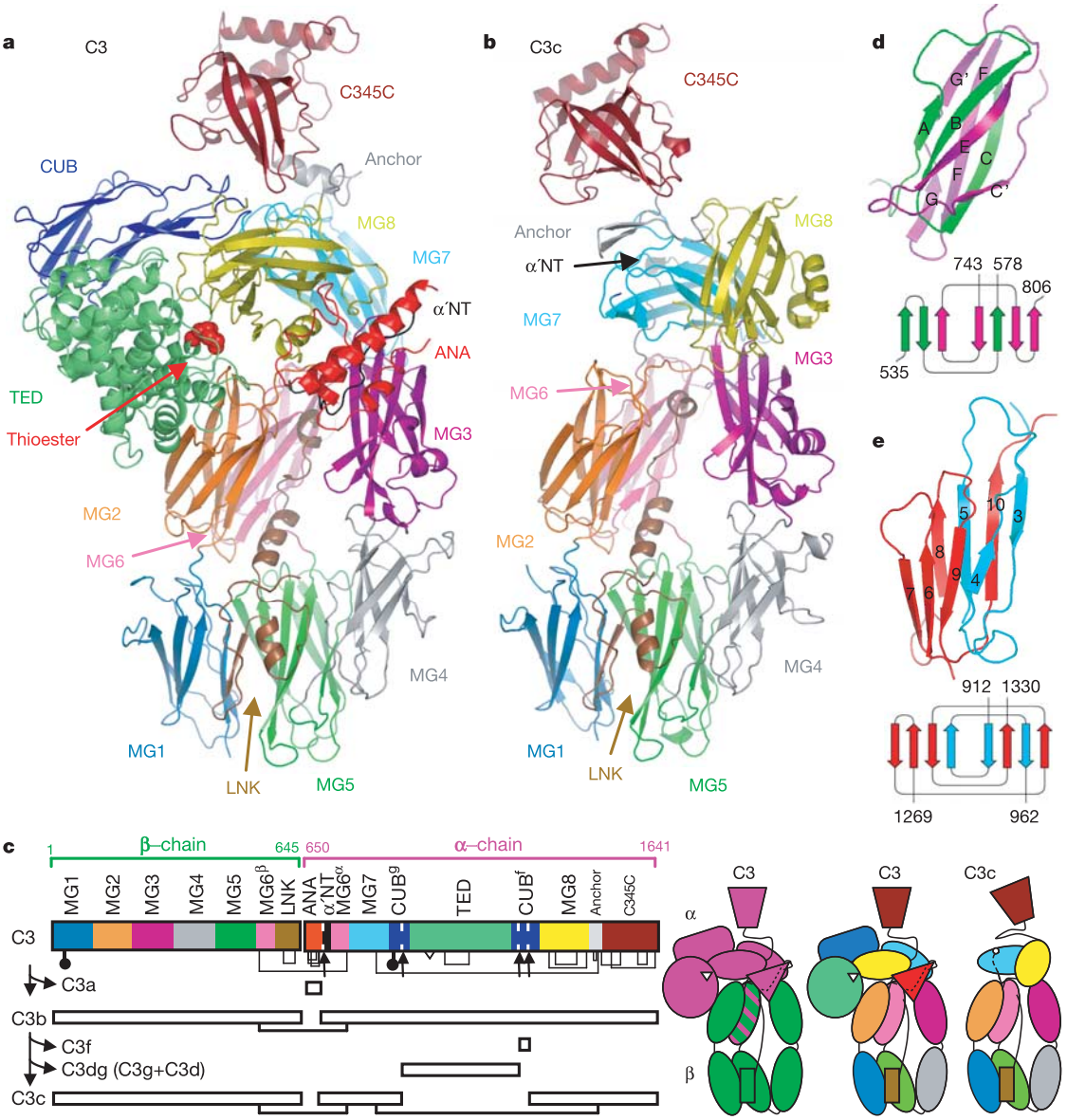

Figure 1 | Structures of human complement components C3 and C3c. **a**, **b**, Ribbon representation of native C3 (13 domains) and C3c (10 domains), respectively. Also shown are intact thioester (red spheres), anchor region (grey) and α′NT (black). **c**, Domain sequence and arrangements in C3 and C3c. The colour scheme matches that in **a**, **b**. Shown are the thioester site (white triangle), disulphide bridges, glycan positions (for details see Supplementary Figs 3 and 6) and cleavage sites. Sequential proteolysis from C3 to C3c is indicated. **d**, **e**, Intertwined domains. **d**, MG6 intertwines the β- and α-chain of mature C3. $MG6^{\beta}$, green; $MG6^{\alpha}$, pink. **e**, Intertwined CUB. CUB^{g}, cyan; CUB^{f}, red. Fibronectin-type 3 and CUB strand numbering are indicated.

Table 1 | Refinement statistics

	C3c	C3
Resolution (Å)	40–2.4	40–3.3
R_{work}/R_{free} (%)	22.3/27.9	23.6/29.5
Number of atoms		
Protein	17,686	12,859
Ligand/ion	90	-
Water	612	-
B factors (Å^2)		
Protein	46	93
Ligand/ion	67	-
Water	43	-
Root mean square deviations		
Bond lengths (Å)	0.012	0.002
Bond angles (degrees)	1.45	0.532

The cleavage sites yielding C3g, C3d and C3f[15,16] do not coincide fully with domain boundaries. Residues 912–962 and 1269–1330, 63% of C3g and all of C3f, together form one domain that displays a CUB fold; we refer to the two separate parts as CUB^g and CUB^f, respectively (Fig. 1e). The remaining part of C3g and C3d (residues 963–1268) together form a thioester-containing domain (TED). Despite a major rearrangement of the N-terminal region, the α6–α6 fold is conserved between TED and C3d[18] (Supplementary Fig. 3m). Finally, residues 1496–1641 form a carboxy-terminal C345C domain with a netrin-like fold. This domain is covalently linked to MG8 by the polypeptide chain and to MG7 by a disulphide bond (Cys 851–Cys 1491)[23] in what we refer to as an anchor region (residues 1475–1495). The overall structure of C345C is similar to the recent NMR-solution structure of the C345C domain of C5 (ref. 24; see also Supplementary Fig. 3r). In total, nine out of thirteen domains were unpredicted (MG1–MG7, LNK and CUB). All domains together form an irregularly shaped C3 and a disc-shaped C3c that lacks the ANA, CUB and TED domains, with strong dipolar surface-charge distributions (Supplementary Fig. 4).

The α-chain undergoes major domain rearrangements

There are marked conformational differences between C3 and C3c. The overall shape of the β-ring is well conserved; the only large change in MG1–MG6 and LNK is a 15° rotation and a 4.3 Å (centre of mass) shift of MG3 (Fig. 2; see also Supplementary Table 2). The domains of the α-chain, however, undergo large rearrangements when the molecule is processed from C3 into C3c. MG7 and MG8 almost swap places in the overall structure: they rotate and translate by 35° and 7.8 Å, and 62° and 24 Å, respectively. The C345C swivels 32° and moves 10.1 Å, resulting in a torque motion acting on the anchor region. The anchor region, with its internal Cys 1484–Cys 1489 disulphide bond, alters its conformation completely, from an α-helix in C3 to a β-hairpin in C3c (Fig. 2b; see also Supplementary Figs 2c and 3q). Overall, the structural differences between C3 and C3c suggest that the β-ring forms a relatively stable molecular platform for the structurally adaptable α-chain to undergo induced conformational changes.

Implications for evolutionary events

Intertwining of distant parts of the amino-acid sequence owing to inclusion of a domain into a loop is indicative of a gene insertion event. This situation occurs twice in C3 (Fig. 1d, e). MG6$^{\beta/\alpha}$ has an insert of 165 residues in loop βC–βC′ and CUB$^{g/f}$ has an insert of 307 residues in loop β5–β6. The MG6 insert forms the linker domain, the processing site 646-RRRR-649 (present in proC3), the anaphylatoxin domain, scissile bond Arg 726–Ser 727 and the α′NT, which contains four acidic residues important for factor B binding to C3b[25]. Because the MG6 insert includes the proC3 processing site, it follows that MG6 is formed by two separate chains in mature C3. The insert into the CUB domain is the TED domain with the critical thioester group. Moreover, we hypothesize that the CUB domain itself is a putative insert into the linker between MG7 and MG8 (and hence the obvious signal of a domain formed by distant parts of the sequence is lacking). Notably, critical C3 elements such as the thioester, the anaphylatoxin and the proteolytic cleavage sites lie within these putative inserts.

Comparison of C3 with other members of the α2M family (Fig. 3) shows that the MG6 insert encompasses the central variable region, which in α2M contains the bait region[26]. In phylogenetically distant (insect and nematode) TEPs this sequence is hypervariable[27]. In contrast, all sequences contain CUB^g-, TED- and CUB^f-like elements, the hallmark of α2M family members. Homologous sequences lacking these elements are not known. The occurrence of α6–α6 folds in enzymes[18] indicates perhaps that the TED fold existed separately before it was inserted and became a critical component in the proteins of the α2M family. The combination of two putative gene insertions and a gene extension of the C-terminal C345C domain[28] implies the existence of a hypothetical ancestral molecule comprised of eight MG domains, which possibly arose by gene duplication events. The sequences of the MG domains have diverged to an extent whereby there is no longer any detectable sequence

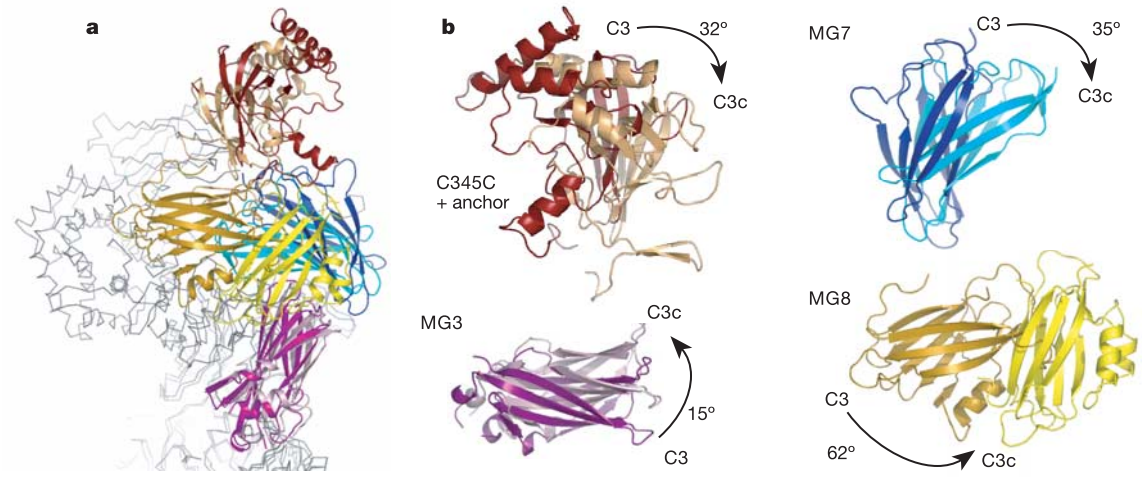

Figure 2 | Differences in domain arrangements between C3 and C3c. a, C3 and C3c superposed on the basis of MG1, MG2 and MG4–MG6 of the β-ring shown as Cα trace of C3 (grey) and C3c (light grey). Domains that undergo large rearrangements (MG3, MG7, MG8 and C345C) are shown in ribbon representation. The colour scheme matches that of Fig. 1, with dark colours for C3 and lighter colours for C3c. **b**, Rearrangements observed for domains MG3, MG7, MG8 and C345C; orientations are different compared with **a**, for clarity. MG8 and the anchor region of C345C differ in secondary structure between C3 and C3c (Supplementary Figs 2a and 3p, q).

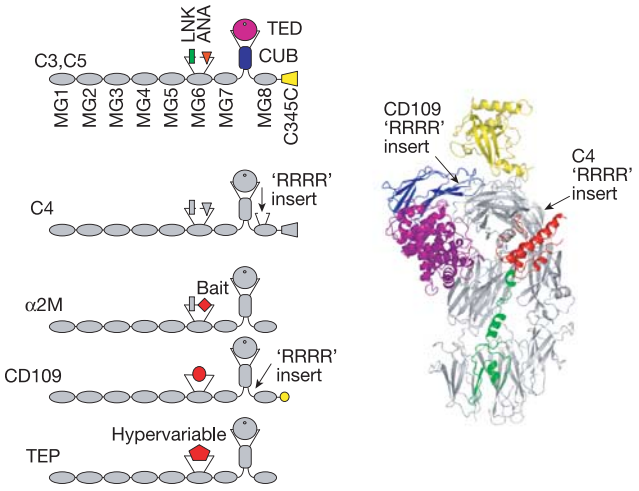

Figure 3 | Domain organization of the α2M family deduced from the C3 structure. Members of the α2M family are aligned schematically by domains with LNK, ANA, CUB and TED presented as inserts that potentially arose by gene-insertion events. Mature C4 has an additional tetra-arginine (RRRR) processing site that occurs in an insert of 46 residues in loop βA–βB of MG8. CD109 (ref. 50) has a tetra-arginine site in the linker between CUB and MG8. Proteins of the α2M family differ in their composition at the C terminus. α2M and TEPs lack the C345C domain, whereas CD109 has a GPI anchor here.

homology. Insertion of CUB^{g}/TED/CUB^{f} marks the emergence of the α2M family in host protection, and additional insertion of LNK/RRRR/ANA/α′NT and extension with C345C marks the emergence of the complement system in the innate immune defence more than 700 million years ago[29].

A double mechanism protects the thioester

The highly reactive thioester (formed by the side chains of Cys 988 and Gln 991)[23] is intact in the structure of C3. It is shielded from reacting with water ($t_{1/2} > 6$ days)[30] or other small nucleophiles[31,32] by a hydrophobic/aromatic pocket formed by residues Met 1378, Tyr 1425 and Tyr 1460 from MG8 and Phe 1047 from the TED (Fig. 4a). These residues are conserved in the α2M family, except for C5, which lacks the thioester moiety. Although fully buried, the thioester and its protective pocket are positioned at the TED–MG8 interface close to the protein surface.

Upon activation by the proteolytic cleavage of C3 into C3b, the protein reacts more readily with hydroxyl groups than amino groups[33]. The altered reactivity is due to a transformation of the thioester into a free thiolate anion, on Cys 988, and formation of an acyl-imidazole intermediate[34,35] by Gln 991 and His 1104 that is possibly stabilized by Glu 1106 (refs 18, 36). In the structure of C3 His 1104 and Glu 1106 (situated next to a disordered loop) are far from each other and far away from Gln 991 (11.7 Å distance between Gln 991 Cδ and His 1104 Nε; Fig. 4b), and hence the catalytic site for reaction with hydroxyls is not present. In the structure of C3d[18], expressed in *Escherichia coli* with a Cys988Ala mutation, His 1104 is only 4.1 Å from Gln 991 in the modelled thioester, and the side chains of His 1104 and Glu 1106 form a hydrogen bond, as would be the case when Glu 1106 stabilizes the acyl-imidazole intermediate (Fig. 4b). In native C3, movement of His 1104 and Glu 1106 towards the thioester is prevented by the MG8–TED interface, thereby blocking the formation of the free thiolate and acyl-imidazole intermediate. This provides a second, specific protection mechanism of the native protein against hydrolysis by water and explains why the reactive moiety in C3 is more resistant to water and less resistant to reaction with small amino nucleophiles[31,32].

Activation of C3

The orientation of the TED domain with the thioester pointing inwards seems to be essential for maintaining the protective TED–MG8 interactions in native C3. TED is buttressed in this position by interactions with MG2, MG8 and CUB (Fig. 4c). Cleavage of C3 at Ser 726–Arg 727 removes the ANA domain (which becomes the anaphylatoxin C3a) and yields an activated C3b with an exposed and reactive thioester. Therefore, the ANA domain has a critical role in protecting the thioester in native C3. However, in the structure of C3 no direct contacts are observed between the ANA and TED domains. Although loops of ANA and TED come within ~10 Å, there are no charged residues that could give long-range electrostatic

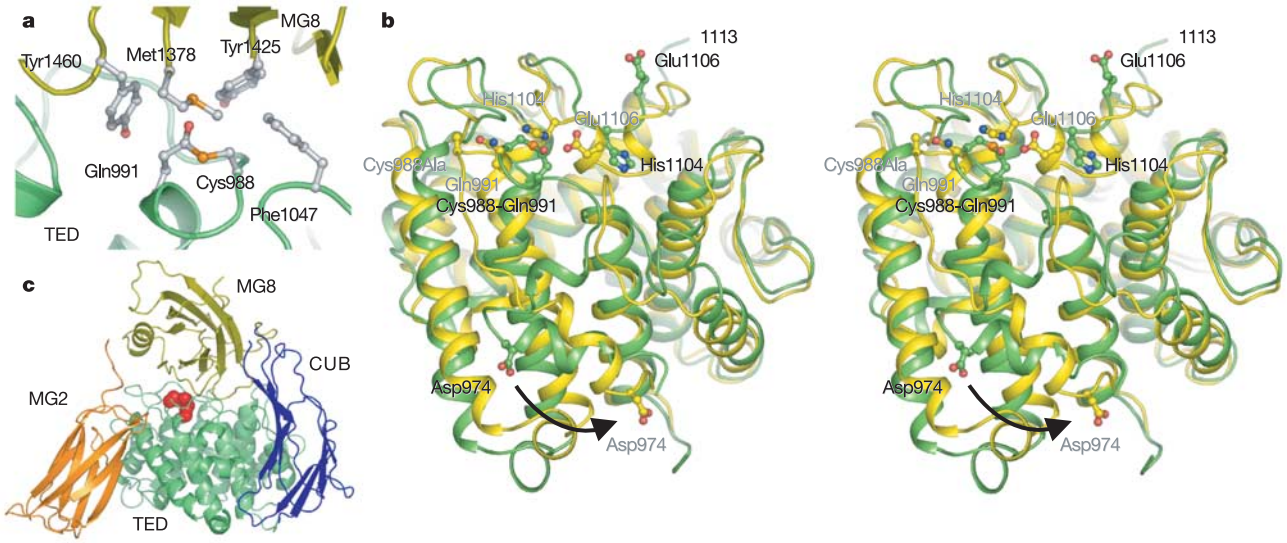

Figure 4 | Interactions of the thioester and the TED domain. a, The first thioester protection mechanism involves a hydrophobic/aromatic pocket formed by residues of TED and MG8. Shown are protective residues and thioester (Cys 988–Gln 991) (ball-and-stick representation) with TED (green ribbon) and MG8 (yellow ribbon). **b**, The second thioester protection mechanism blocks reaction with hydroxyls by placing His 1104 and Glu 1106 far from the thioester. Stereo diagram of superposed TED domain (green) and C3d[18] (yellow). Cys 988 was mutated to Ala in C3d[18]. Movements in the N-terminal α0–α1 region between TED and C3d are shown by 15 Å displacement of Asp 974. **c**, Domain–domain interactions stabilize the MG2–TED–CUB–MG8 arrangement. Domain colours match those used in Fig. 1.

attractions. Therefore ANA must protect the thioester in native C3 indirectly. ANA has extensive interactions with MG8. At the other side ANA interacts with MG3 (Fig. 5a) and thus bridges interactions between MG8 and MG3 of the stable β-ring. The ANA domain may serve to keep MG8 in a correct position for interaction with TED and, possibly, to induce a conformation of MG8 that enhances interactions with TED.

MG8 undergoes a conformational and a huge positional change from C3 to C3c (Fig. 2). The domain makes a dramatic swing from the interior of the C3 molecule to the exterior in C3c. This swing moves helix α1 and surrounding β-strands by 40–50 Å. In C3 these structural elements are hidden and interact with the ANA and TED domains, whereas in C3c they are fully exposed and have changed conformation into a β-strand and two α-helices. This large positional change corresponds well with what is expected for the equivalent receptor-binding domain in α2M, which alters from hidden to exposed upon encapsulating a proteinase in an α2M multimer[37]. In the exposed position its β-α-β motif is thought to bind the α2M receptor[22]. In native C3 the β-α-β configuration in MG8 is probably important; it contributes to interaction with TED (notably, residues Lys 1409 and the highly conserved Glu 1411). Furthermore, this MG8 region harbours a binding site for complement protein properdin[38]. The occluded position in C3 explains why properdin does not bind to native C3. Because properdin stabilizes the C3bBb convertase complex, it follows that MG8 exposes this binding site, possibly in a β-α-α configuration and a swung out position as in C3c.

The observed domain arrangement supports a simple model for exposing the thioester in the activated state of C3. The removal of ANA, yielding C3b, weakens the interactions between MG8 and TED, thereby allowing TED to swing out of its nestled position. Similar, although much slower, conformational changes occur after spontaneous hydrolysis of the thioester in C3, yielding C3(H_2O) (refs 31, 32). In this case, the conformational changes are retarded by the presence of the anaphylatoxin domain. Rotation of TED around its N- and C-terminal connections to CUB exposes TED with the intact thioester in C3b projected outwards. This simple rotation is supported by recent hydrogen–deuterium exchange data on C3 and C3(H_2O), which indicate that segments 956–968 and 1027–1036 buried in C3 become exposed in C3(H_2O) (ref. 39). Reduced exchange rates for segment 1108–1123 correlate with the disorder–order transition in loop 1107–1112 of C3 versus C3d. This loop lies next to the critical His 1104 and Glu 1106 residues. Possibly, the rotation of TED is linked to the formation of the free thiolate and acyl-imidazole intermediate that requires ~15 Å movements in the N-terminal α0–α1 region to establish an active conformation for attachment to a target surface (Fig. 4b).

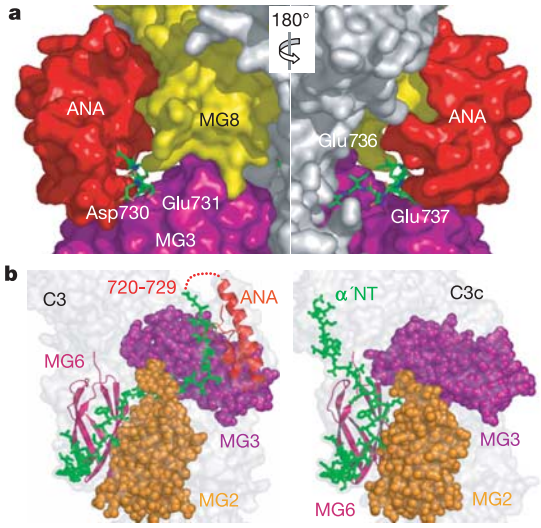

Figure 5 | The α′NT region (residues 727–744) slips through the β-ring. **a**, Two views, rotated by 180°, of the cone formed by ANA, MG3 and MG8 (surface representation), and residues 730–744 (stick representation). Residues important for factor B binding are labelled. **b**, The α′NT regions are on opposite sides of the molecule in C3 and C3c. In C3 this region is covalently linked to ANA; the scissile bond 726–727 is part of a disordered loop 720–729 (dashed line). Also shown are surface contours of C3 and C3c (transparent light grey), residues 727–767 (green sticks), ANA and MG6 (ribbons), and MG2 and MG3 of the β-ring (spheres).

Implications for convertase formation and regulation

Cleavage of C3 at 726–727 yields a novel N-terminal (that is, α′NT) region in C3b. Four acidic residues (Asp 730, Glu 731, Glu 736 and Glu 737) of the α′NT are important in the formation of the C3 convertase[25]. In addition, the extended segment 727–767 (that is, α′NT plus the beginning of $MG6^{\alpha}$) is important for binding of several regulators[25,40,41]. In native C3, acidic residues Asp 730, Glu 731, Glu 736 and Glu 737 are shielded by ANA (Fig. 5a), which may explain why factor B cannot bind to native C3. Most surprisingly, the α′NT region (residues 727–744) resides on opposite sides in C3 and C3c (Fig. 5b). From C3 to C3c the α′NT has moved from its position near ANA, through the β-ring and has emerged on the other side in C3c where it lies on the surface of MG7. Sequence conservation of the hinge 745-FPES-748 that connects α′NT to MG6, and of the bridging loop 205-YVLP-208 of the β-ring (Supplementary Fig. 5), suggests that this remarkable rearrangement may be part of a general mechanism of α2M protein family members.

The striking differences between C3 and C3c concerning the position of the α′NT region and the arrangement of the α-chain domains raise important questions with respect to formation and regulation of convertases. In the series of activating and deactivating steps (from C3, C3b, iC3b to C3c), when do these rearrangements occur? How do the rearrangements relate to the binding and functioning of the various regulatory factors and receptors? Possibly, most rearrangements occur upon activation, in which case C3b and iC3b will resemble C3c and differ primarily in unravelling of the CUB domain. Alternatively, the rearrangements occur in a gradual process yielding different arrangements of each molecule. The C3 and C3c structures facilitate further structural and mutational studies of the activation and regulation of this central component of the complement system, and create new opportunities for drug development, targeting a wide variety of inflammatory diseases that have been associated with complement activation.

METHODS

Protein purification. C3c was purified using methods as described previously for C3 (ref. 42), with slight modifications. C3c from outdated human plasma (stored for several weeks at 4 °C) was purified by polyethylene glycol (PEG) precipitation, anion-exchange chromatography (DEAE Sephacel), cation-exchange chromatography (CM-Sephadex C50) and size-exclusion chromatography (Sephacryl 300). Trace amounts of IgG and IgA were removed by immune absorption. C3c was concentrated to 20 mg ml^{-1} and dialysed against 10 mM Tris pH 7.4, 2 mM EDTA and 2 mM benzamidine. The glycan moiety on Asn 917 was cleaved off with *N*-glycosidase F (PNGase F) before crystallization. Both termini in the α-chain, generated by C3dg and C3f removal during plasma storage, displayed heterogeneity as observed by reduced gel and matrix-assisted laser desorption/ionization–time of flight (MALDI–TOF) analysis.

C3 was purified from frozen human plasma as described previously[42], with slight modifications. C3 was purified by PEG precipitation, plasminogen depletion (Sepharose 4B-L-lysine), anion-exchange chromatography (DEAE Sephacel) and size-exclusion chromatography (Sepharose CL-6B). Trace amounts of IgG, IgA, IgM, C5 and factor H were removed by immune absorption. C3 was concentrated to 8 mg ml^{-1} and precipitated by dialysis against 5 mM MES, pH 6.0 and stored at −80 °C until use to retain haemolytic activity. Before crystallization, C3 was solubilized by dialysis against 10 mM Tris, pH 7.4.

Crystallization and data collection. C3c was crystallized in hanging drops from mother liquor containing 18% w/v PEG-3000 and 200 mM $LiNO_3$ at 20 °C. Crystals grew to 100 × 100 × 100 μm within 2 weeks. For cryo-protection 20%

v/v glycerol was added to the mother liquor, and crystals were flash-cooled in liquid nitrogen. Crystals displayed space group $P2_12_12$ ($a = 126.9$, $b = 246.9$, $c = 87.4$ Å), contained two molecules per asymmetric unit and diffracted to 2.4 Å resolution at ESRF beamline ID14-EH4. Varying pseudo-B centring was observed between crystals with extinctions for $(h + l) =$ odd reflections ranging from 10% to 100% (at 100%, crystals exhibited space group $B22_12$). In addition, crystals showed significant variations in cell dimensions. Diffraction data were processed using MOSFLM/CCP4 (ref. 17) and Denzo/Scalepack[43]. Heavy-atom derivatives were prepared by soaking crystals in mother liquor containing 1–5 mM of heavy atom compounds for 24 h at 20 °C.

C3 was crystallized in hanging drops in 6% w/v PEG-monomethyl ether 550 and 100 mM sodium acetate at 20 °C. Crystals grew to 300 × 100 × 20 μm within 2 weeks. Similar to C3c, 20% v/v glycerol was added as a cryo-protectant; crystals were flash-cooled in liquid nitrogen. Crystals exhibited space group $I222$ ($a = 117.0$, $b = 156.3$, $c = 271.2$ Å) with one molecule per asymmetric unit and diffracted to 3.3 Å resolution at ESRF beamline ID14-EH4. Diffraction data were processed with XDS and XSCALE[44].

Structure determination. Twenty mercury sites were found and refined in SOLVE[45] for a $HgCl_2$ derivative with diffraction data from a MAD experiment at mercury-inflection and high-remote energy. Twenty-seven platinum sites were found and refined in SOLVE[45] for a K_2PtCl_4 derivative from data of a SIRAS experiment at the platinum-peak energy and a data set collected from a native crystal. Phase combination and extension by multi-crystal and non-crystallographic symmetry averaging were crucial to produce an interpretable electron-density map. In the end, we used nine partial masks created in RAVE[46] to envelope a single molecule for averaging over six copies of the protein molecule in DMMULTI[17]. The resulting electron density was of good quality. A partial model was built automatically by ARP/wARP[47]. The model of C3c was completed using O[48] and refined using REFMAC[17] to a final R-value of 22.3% (R_{free} of 27.9%).

C3 was solved by molecular replacement. Initial positioning of C3c, or any of its fragments, was unsuccessful. However, the α6–α6 barrel structure of C3d[18] (18% of C3) was positioned correctly by Phaser[19]. Subsequently domains of C3c (C345C, MG2 plus MG6, MG5, MG4, MG8, MG3, MG7 and MG1, respectively) were added step-by-step using both Phaser[19] and molecular graphics in O[48] with rigid-body refinement for positioning in Phaser. This yielded an 85% complete model that was completed by model building (of CUB, ANA and loops) and refined in CNS[49] and REFMAC[17]. The final refined model of C3 had R and R_{free} values of 23.6% and 29.5% (see Table 1; see also Supplementary Fig. 1 and Supplementary Table 1).

Received 8 June; accepted 5 July 2005.

1. Carroll, M. C. The complement system in regulation of adaptive immunity. *Nature Immunol.* **5,** 981–986 (2004).
2. Walport, M. J. Complement. First of two parts. *N. Engl. J. Med.* **344,** 1058–1066 (2001); Complement. Second of two parts. *N. Engl. J. Med.* **344**, 1140–1144 (2001).
3. Sunyer, J. O., Zarkadis, I. K. & Lambris, J. D. Complement diversity: a mechanism for generating immune diversity? *Immunol. Today* **19,** 519–523 (1998).
4. Levashina, E. A. *et al.* Conserved role of a complement-like protein in phagocytosis revealed by dsRNA knockout in cultured cells of the mosquito, *Anopheles gambiae*. *Cell* **104,** 709–718 (2001).
5. Budd, A., Blandin, S., Levashina, E. A. & Gibson, T. J. Bacterial α2-macroglobulins: colonization factors acquired by horizontal gene transfer from the metazoan genome? *Genome Biol.* **5,** R38 (2004).
6. Bokisch, V. A., Muller-Eberhard, H. J. & Cochrane, C. G. Isolation of a fragment (C3a) of the third component of human complement containing anaphylatoxin and chemotactic activity and description of an anaphylatoxin inactivator of human serum. *J. Exp. Med.* **129,** 1109–1130 (1969).
7. Tack, B. F., Harrison, R. A., Janatova, J., Thomas, M. L. & Prahl, J. W. Evidence for presence of an internal thiolester bond in third component of human complement. *Proc. Natl Acad. Sci. USA* **77,** 5764–5768 (1980).
8. Lambris, J. D. The multifunctional role of C3, the third component of complement. *Immunol. Today* **9,** 387–393 (1988).
9. Law, S. K., Lichtenberg, N. A. & Levine, R. P. Evidence for an ester linkage between the labile binding site of C3b and receptive surfaces. *J. Immunol.* **123,** 1388–1394 (1979).
10. Muller-Eberhard, H. J. & Gotze, O. C3 proactivator convertase and its mode of action. *J. Exp. Med.* **135,** 1003–1008 (1972).
11. Fishelson, Z., Pangburn, M. K. & Muller-Eberhard, H. J. Characterization of the initial C3 convertase of the alternative pathway of human complement. *J. Immunol.* **132,** 1430–1434 (1984).
12. Pangburn, M. K., Schreiber, R. D. & Muller-Eberhard, H. J. Human complement C3b inactivator: isolation, characterization, and demonstration of an absolute requirement for the serum protein β1H for cleavage of C3b and C4b in solution. *J. Exp. Med.* **146,** 257–270 (1977).
13. Ross, G. D., Lambris, J. D., Cain, J. A. & Newman, S. L. Generation of three different fragments of bound C3 with purified factor I or serum. I. Requirements for factor H vs CR1 cofactor activity. *J. Immunol.* **129,** 2051–2060 (1982).
14. Seya, T., Turner, J. R. & Atkinson, J. P. Purification and characterization of a membrane protein (gp45-70) that is a cofactor for cleavage of C3b and C4b. *J. Exp. Med.* **163,** 837–855 (1986).
15. Harrison, R. A. & Lachmann, P. J. Novel cleavage products of the third component of human complement. *Mol. Immunol.* **17,** 219–228 (1980).
16. Lachmann, P. J., Pangburn, M. K. & Oldroyd, R. G. Breakdown of C3 after complement activation. Identification of a new fragment C3g, using monoclonal antibodies. *J. Exp. Med.* **156,** 205–216 (1982).
17. Collaborative Computational Project, Number 4, The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50,** 760–763 (1994).
18. Nagar, B., Jones, R. G., Diefenbach, R. J., Isenman, D. E. & Rini, J. M. X-ray crystal structure of C3d: a C3 fragment and ligand for complement receptor 2. *Science* **280,** 1277–1281 (1998).
19. Storoni, L. C., McCoy, A. J. & Read, R. J. Likelihood-enhanced fast rotation functions. *Acta Crystallogr. D* **60,** 432–438 (2004).
20. de Bruijn, M. H. & Fey, G. H. Human complement component C3: cDNA coding sequence and derived primary structure. *Proc. Natl Acad. Sci. USA* **82,** 708–712 (1985).
21. Huber, R., Scholze, H., Paques, E. P. & Deisenhofer, J. Crystal structure analysis and molecular model of human C3a anaphylatoxin. *Hoppe-Seyler's Z. Physiol. Chem.* **361,** 1389–1399 (1980).
22. Jenner, L., Husted, L., Thirup, S., Sottrup-Jensen, L. & Nyborg, J. Crystal structure of the receptor-binding domain of α2-macroglobulin. *Structure* **6,** 595–604 (1998).
23. Thomas, M. L., Janatova, J., Gray, W. R. & Tack, B. F. Third component of human complement: localization of the internal thiolester bond. *Proc. Natl Acad. Sci. USA* **79,** 1054–1058 (1982).
24. Bramham, J. *et al.* Functional insights from the structure of the multifunctional C345C domain of C5 of complement. *J. Biol. Chem.* **280,** 10636–10645 (2005).
25. Taniguchi-Sidle, A. & Isenman, D. E. Interactions of human complement component C3 with factor B and with complement receptors type 1 (CR1, CD35) and type 3 (CR3, CD11b/CD18) involve an acidic sequence at the N-terminus of C3 alpha′-chain. *J. Immunol.* **153,** 5285–5302 (1994).
26. Sottrup-Jensen, L., Sand, O., Kristensen, L. & Fey, G. H. The α-macroglobulin bait region. Sequence diversity and localization of cleavage sites for proteinases in five mammalian α-macroglobulins. *J. Biol. Chem.* **264,** 15781–15789 (1989).
27. Lagueux, M., Perrodou, E., Levashina, E. A., Capovilla, M. & Hoffmann, J. A. Constitutive expression of a complement-like protein in toll and JAK gain-of-function mutants of *Drosophila*.. *Proc. Natl Acad. Sci. USA* **97,** 11427–11432 (2000).
28. Ishii, N., Wadsworth, W. G., Stern, B. D., Culotti, J. G. & Hedgecock, E. M. UNC-6, a laminin-related protein, guides cell and pioneer axon migrations in *C. elegans*. *Neuron* **9,** 873–881 (1992).
29. Sahu, A. & Lambris, J. D. Structure and biology of complement protein C3, a connecting link between innate and acquired immunity. *Immunol. Rev.* **180,** 35–48 (2001).
30. Pangburn, M. K., Schreiber, R. D. & Muller-Eberhard, H. J. Formation of the initial C3 convertase of the alternative complement pathway. Acquisition of C3b-like activities by spontaneous hydrolysis of the putative thioester in native C3. *J. Exp. Med.* **154,** 856–867 (1981).
31. Isenman, D. E., Kells, D. I., Cooper, N. R., Muller-Eberhard, H. J. & Pangburn, M. K. Nucleophilic modification of human complement protein C3: correlation of conformational changes with acquisition of C3b-like functional properties. *Biochemistry* **20,** 4458–4467 (1981).
32. Pangburn, M. K. Spontaneous reformation of the intramolecular thioester in complement protein C3 and low temperature capture of a conformational intermediate capable of reformation. *J. Biol. Chem.* **267,** 8584–8590 (1992).
33. Law, S. K. & Dodds, A. W. The internal thioester and the covalent binding properties of the complement proteins C3 and C4. *Protein Sci.* **6,** 263–274 (1997).
34. Gadjeva, M. *et al.* The covalent binding reaction of complement component C3. *J. Immunol.* **161,** 985–990 (1998).
35. Dodds, A. W., Ren, X. D., Willis, A. C. & Law, S. K. The reaction mechanism of the internal thioester in the human complement component C4. *Nature* **379,** 177–179 (1996).
36. van den Elsen, J. M. *et al.* X-ray crystal structure of the C4d fragment of human complement component C4. *J. Mol. Biol.* **322,** 1103–1115 (2002).
37. Qazi, U., Gettins, P. G. & Stoops, J. K. On the structural changes of native human α2-macroglobulin upon proteinase entrapment. Three-dimensional structure of the half-transformed molecule. *J. Biol. Chem.* **273,** 8987–8993 (1998).
38. Daoudaki, M. E., Becherer, J. D. & Lambris, J. D. A 34-amino acid peptide of the third component of complement mediates properdin binding. *J. Immunol.* **140,** 1577–1580 (1988).
39. Winters, M. S., Spellman, D. S. & Lambris, J. D. Solvent accessibility of native and hydrolyzed human complement protein 3 analyzed by hydrogen/deuterium

exchange and mass spectrometry. *J. Immunol.* **174,** 3469–3474 (2005).
40. Oran, A. E. & Isenman, D. E. Identification of residues within the 727–767 segment of human complement component C3 important for its interaction with factor H and with complement receptor 1 (CR1, CD35). *J. Biol. Chem.* **274,** 5120–5130 (1999).
41. Lambris, J. D. *et al.* Dissection of CR1, factor H, membrane cofactor protein, and factor B binding and functional sites in the third complement component. *J. Immunol.* **156,** 4821–4832 (1996).
42. Hammer, C. H., Wirtz, G. H., Renfer, L., Gresham, H. D. & Tack, B. F. Large scale isolation of functionally active components of the human complement system. *J. Biol. Chem.* **256,** 3995–4006 (1981).
43. Otwinowski, Z. & Minor, W. Processing X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276,** 307–326 (1997).
44. Kabsch, W. Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *J. Appl. Crystallogr.* **26,** 795–800 (1993).
45. Terwilliger, T. C. & Berendzen, J. Automated MAD and MIR structure solution. *Acta Crystallogr. D* **55,** 849–861 (1999).
46. Kleywegt, G. J. & Jones, T. A. Software for handling macromolecular envelopes. *Acta Crystallogr. D* **55,** 941–944 (1999).
47. Perrakis, A., Morris, R. & Lamzin, V. S. Automated protein model building combined with iterative structure refinement. *Nature Struct. Biol.* **6,** 458–463 (1999).
48. Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47,** 110–119 (1991).
49. Brunger, A. T. *et al.* Crystallography NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54,** 905–921 (1998).
50. Solomon, K. R., Sharma, P., Chan, M., Morrison, P. T. & Finberg, R. W. CD109 represents a novel branch of the α2-macroglobulin/complement gene family. *Gene* **327,** 171–183 (2004).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We acknowledge the help of beamline scientists of the EMBL/ESRF and in particular R. B. G. Ravelli for help in data collection, and A. Perrakis and D. Egan for help and use of their crystallization robots. We thank A. T. Brunger, M. Bowen, B. DeLaBarre, J. D. Lambris, C. W. Vogel, D. Fritzinger, T. Springer and T. K. Sixma for critically reading the manuscript. This work was supported by a 'Pionier' programme grant (P.G.) of the Council for Chemical Sciences of the Netherlands Organization for Scientific Research (NWO-CW), the Dutch Kidney Foundation (A.R.), the Swedish Research Council (B.N.) and faculty grants at the University of Kalmar (K.N.-E.). A.R. does not support the evolutionary implications.

Author Information Co-ordinates and structure factors have been deposited in the Protein Data Bank under accession numbers 2A73 (C3) and 2A74 (C3c). Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to P.G. (p.gros@chem.uu.nl).