Structural Basis for the Function of Complement Component C4 within the Classical and Lectin Pathways of Complement

Sofia Mortensen, Rune T. Kidmose, Steen V. Petersen, Agnes Szilágyi, Zoltan Prohászka and Gregers R. Andersen

*J Immunol* published online 24 April 2015
http://www.jimmunol.org/content/early/2015/04/24/jimmunol.1500087

**Supplementary Material**
http://www.jimmunol.org/content/suppl/2015/04/24/jimmunol.1500087.DCSupplemental

**Subscription**
Information about subscribing to *The Journal of Immunology* is online at: http://jimmunol.org/subscription

**Permissions**
Submit copyright permission requests at: http://www.aai.org/About/Publications/JI/copyright.html

**Email Alerts**
Receive free email-alerts when new articles cite this article. Sign up at: http://jimmunol.org/alerts
Structural Basis for the Function of Complement Component C4 within the Classical and Lectin Pathways of Complement

Sofia Mortensen,* Rune T. Kidmose,* Steen V. Petersen,† Ágnes Szilágyi,‡ Zoltan Prohászka,‡ and Gregers R. Andersen*  

Complement component C4 is a central protein in the classical and lectin pathways within the complement system. During activation of complement, its major fragment C4b becomes covalently attached to the surface of pathogens and altered self-tissue, where it acts as an opsonin marking the surface for removal. Moreover, C4b provides a platform for assembly of the proteolytically active convertases that mediate downstream complement activation by cleavage of C3 and C5. In this article, we present the crystal and solution structures of the 195-kDa C4b. Our results provide the molecular details of the rearrangement accompanying C4 cleavage and suggest intramolecular flexibility of C4b. The conformations of C4b and its paralogue C3b are shown to be remarkably conserved, suggesting that the convertases from the classical and alternative pathways are likely to share their overall architecture and mode of substrate recognition. We propose an overall molecular model for the classical pathway C5 convertase in complex with C5, suggesting that C3b increases the affinity for the substrate by inducing conformational changes in C4b rather than a direct interaction with C5. C4b-specific features revealed by our structural studies are probably involved in the assembly of the classical pathway C3/C5 convertases and C4b binding to regulators. The Journal of Immunology, 2015, 194: 000–000.

The lectin pathway (LP) and the classical pathway (CP) (Fig. 1A) are two of the three main activation pathways within the complement system that are involved in resolution of infections, clearance of immune complexes, removal of apoptotic/necrotic cells, and stimulation of adaptive immunity (1). Both pathways are initiated by the binding of pattern recognition molecules (PRMs) to molecular patterns presented by pathogens or altered self-surfaces. In the CP, the PRM C1q recognizes IgG or IgM present in immune complexes or bound to nonself surfaces, pentraxins, and other ligands (2). Activation of the LP is based on five PRMs—mannan-binding lectin, H/M/L-ficolin, and CL-L1/K1—that bind specific carbohydrate structures or acetylated molecules presented by pathogens (3). The PRMs in both pathways form complexes with paralogous serine proteases; C1q binds the heterotrimer C1r/C1s, and the LP PRMs bind mannan-binding lectin–associated serine protease (MASP)-1 and MASP-2. Upon pattern recognition, the associated proteases become activated, and C1s or MASP-2 cleave C4 into C4b (195 kDa) and C4a (9 kDa). Although the small fragment has no established biological function, C4b acts as an opsonin and attaches to surface nucleophiles on the activator through their reaction with the thioester (TE) exposed upon C4 cleavage (4). Thezymogen C2 may then bind C4b to form the CP proconvertase C4b2a (standardized notation for the C4b–C2 complex and other complexes is used throughout), in which C2 is cleaved by C1s, MASP-2, or MASP-1, and the C2b fragment is released. The C4b2a complex (the CP C3 convertase) remains on the activator and mediates downstream activation by cleaving C3 into C3b and C3a (5). Strong amplification of C3 cleavage proceeds through the alternative pathway (AP) via formation of the AP C3 convertase C3bBb. At a high density of C3b on the activator surface, C3b may combine with the C3 convertases, resulting in the CP and AP C5 convertases, C4b2a3b and C3bBb3b, respectively (6). In both the C3 and the C5 CP convertase, C4b acts as the noncatalytic substrate-binding subunit that presents the catalytic subunit C2a to the substrate. In the CP C5 convertase, C4b also contacts C3b, and they can even become covalently linked because C4b may offer a nucleophile for the cleavage of the TE in nascent C3b (7).

C4 is paralogous to complement proteins C3 and C5 and shares up to 30% sequence identity with the two proteins. However, maturation of C4 is more complex compared with C3 and C5 and includes generation of the three chains (α, β, and γ) from the precursor and the introduction of posttranslational modifications, including four N- and one O-linked glycosylations and sulfation of three tyrosine...
residues (8, 9) (Fig. 1B). C4 exists in two isoforms: the products of the two genes C4A and C4B. Although C4A and C4B differ in only 4-aa residues (C4A: P1120PCPVLD1125; C4B: L1120LSPVIH1125), they display markedly different hemolytic activities and affinities toward diverse substrates. Thus, C4A is more prone to attach to amine groups of an activator, whereas C4B reacts preferentially with hydroxyl groups (10, 11). Additionally, DNA and protein sequencing detected 23 polymorphic amino acid residues in both C4A and C4B (12).

Upon activation of C4, nascent C4b is expected to undergo conformational changes similar to those observed for C3/C3b, and the role of C4b in the CP convertases is assumed to be analogous to that of C3b in the AP convertases with respect to binding of the C3 and C5 substrates (13). There have been several structural studies of the AP C3 proconvertase and convertase (14–16), but it remains to be shown whether C4b2 and C4b2a are organized in the same manner or have additional unique features. There are also important differences between C3b and C4b with regard to their regulation on host cells and recognition by complement receptors (CRs). For example, an important regulator of C4b is C4b binding protein (C4BP), which acts as a cofactor for factor I (FI) cleavage of C4b into C4c and C4d via generation of an intermediate product iC4b (C4BP), which acts as a cofactor for factor I (FI) cleavage of C4b into C4c and C4d via generation of an intermediate product iC4b (17–19); the main cofactor for C3b degradation, factor H (FH), is not considered a regulator of C4b. Another striking example is recognition of C3b/C4b and their degradation products by the CRs; only CR1 has an established function as a receptor for C4b (20), whereas there are five known receptors for C3b and its fragments (1). Hence, although C3b and C4b provide the same platform for whereas there are five known receptors for C3b and its fragments (10, 11). Additionally, DNA and protein sequencing detected 23 polymorphic amino acid residues in both C4A and C4B (12).

To provide the structural basis for the dissection of the multiple functions of C4b, we determined the crystal structure of C4b at 4.2 Å resolution and characterized its solution structure by small angle x-ray scattering (SAXS).

Materials and Methods

Generation and crystallization of C4b

Human C4 was purified from outdated plasma pools without any attempt to separate isoforms C4A and C4B, essentially as described (21), although an additional step of anion-exchange chromatography on a Mono Q column (GE Healthcare) was added. To generate C4b, 50 mM Tris-HCl (pH 8.5), 30 mM glycine, 10 mM iodoacetamide, and 0.1% (w/v) to C4) C1s (Complement Technology) were added to the purified C4. The reaction was incubated at 37°C overnight, and 1% (w/w to C4) C1 Esterase Inhibitor (Complement Technology) was added to the reaction mixture. The digest was subsequently loaded onto a 9-mL Source 15Q column (GE Healthcare) equilibrated in 20 mM HEPES-NaOH (pH 7.5), 200 mM NaCl and eluted with an 80-ml 200–600-mM NaCl gradient. Fractions containing C4b were pooled and dialyzed against 20 mM HEPES-NaOH (pH 7.5), 100 mM NaCl overnight. For deglycosylation, C4b was made 50 mM in Tris-HCl (pH 7) and 10% each (w/v to C4b) in GST-tagged EndoF1 and PNGase F. The reaction mixture was incubated at 30°C for 36 h, after which the endoglycosidases were removed by affinity chromatography on a GStrap FF column (GE Healthcare). The flow-through containing deglycosylated C4b was dialyzed against 20 mM HEPES-NaOH (pH 7.5), 100 mM NaCl overnight, and used for crystallization. Crystals of C4b were grown at 19°C in sitting drops made by mixing in a 1:1.05 ratio, C4b at 4–5 mg/ml, reservoir solution containing 7.5% (v/v) PEG8000, 0.4 M MgCl2, and 0.1 M Tris-HCl (pH 7), and a crystal seed stock solution. Crystals were cryoprotected by soaking them sequentially for 5–10 s in reservoir solution supplemented with 10, 20, and 30% (v/v) PEG400 prior to flash cooling in liquid nitrogen.

Data collection and structure determination

Data were collected at the European Synchrotron Radiation Facility (ESRF) (Grenoble, France) and processed with XDS (22). The structure was determined with molecular replacement using Phaser (23) and two search models based on macroglutinin (MG) domains (24, 25) (RCSB Protein Data Bank [PDB] entry 4FXK) superimposed on the equivalent domains of C3b (RCSB PDB entry 2XWJ) and the TE domain derived from the structure of C4dg (RCSB PDB entry 1HZF). Models were iteratively refined in Phenix.refine (24) and manually rebuilt in the O software (25). Atomic positions were refined with simulated annealing and minimization, whereas temperature factors were refined with grouped B-factors and translation, libration, and screw-rotation groups. The two molecules in the asymmetric unit were restrained to each other with tight positional and loose B-factor noncrystallographic symmetry restraints. Electron-density maps used for model building were improved by real-space averaging. Figures were prepared with PyMOL v1.7.2 (26). Coordinates and structure factors for the C4b crystal structure were deposited in the RCSB PDB (http://www.rcsb.org) under accession code 4XAM.

SAXS data collection and modeling

C4b not treated with endoglycosidases was used for SAXS experiments after buffer exchange into 20 mM HEPES-NaOH (pH 7.5), 100 mM NaCl, 2 mM MgCl2. Data were collected at the EMBL P12 beamline at PETRA III using a PILATUS 2M pixel detector. Scattering of the C4b samples at concentrations of 10, 5, and 2.5 mg/ml was recorded doing 20 exposures of 0.045 s with periods of 0.05 s for each sample in a temperature-controlled cell at 10°C with a path length of 1.5 mm. The sample-to-detector distance was 3.1 m, covering a range of momentum transfer 0.0028 < s < 0.45 Å−1 (s = 4πsinθ/λ). Normalization, radial averaging, buffer subtraction, and concentration correction of the data were done at the beamline by the automated pipeline (26). Radial distribution functions (RDF) were calculated for each dataset using the ATSAS package (27) and compared. The two datasets calculated at the highest protein concentrations (10 and 5 mg/ml) were merged using ALMERGE (26). Data analysis and rigid body (RB) refinement were done with ATOMS, using the data range 0 < s < 2.5 nm−1. Calculation of theoretical scattering profiles of atomic structures and their fits to the experimental data was carried out with CRYSOL and CORAL, respectively, with GraphPad Prism 5.03. The input model for CORAL RB refinement was generated from the C4b crystal structure by addition of full biantennary Asn-linked glycan derived from RCSB PDB entry 3RY6.

Isolation of O-linked glycosylated peptides

To isolate the C4A isoform, we obtained plasma from a healthy donor (54 y old, female) with zero C4B gene copy number, as previously identified (29). C4A was purified essentially as C4 from normal plasma but on smaller columns. The conversion of C4A into C4Ac and C4Ad was monitored through incubation of C4A overnight at 37°C in 50 mM Tris-HCl (pH 8.7), 30 mM glycine, 10 mM iodoacetamide with C1s (1:500 w/w to C4), F1 (1:100 w/w to C4), and in-house purified C4BP (1:60 w/w to C4). The digest was separated by cation-exchange chromatography on a Source 15S column (GE Healthcare) equilibrated in 50 mM Tris-HCl (pH 7.6), 70 mM NaCl (buffer A). C4Ac and C4Ad were eluted with a linear gradient from 70 to 400 mM NaCl in buffer A. Fractions containing C4Ad were pooled and dialyzed into 50 mM NH4HCO3, 30 mM NaCl. To enable the analysis of O-linked glycosylation, the purified C4Ad was incubated overnight at 37°C with 1:30 (w/w to C4Ad) porcine trypsin (Promega) and subjected to subdigestion using asparaginyl endopeptidase (Asn-C; R&D Systems) to obtain smaller peptides compatible with mass spectrometry (MS) analysis. Prior to digestion, the Asn-C was activated in 1 M acetic acid (pH 4.4), 100 mM NaCl for 2 h at 37°C. Tryptic peptides were lyophilized and resuspended in 100 mM citric acid (pH 5.5), 100 mM NaCl, 5 mM DTT; activated Asn-C was added at an approximate ratio of 1:40 (w/w to C4Ad), and the solution was allowed to incubate for 16 h at 37°C. The digest was acidified by the addition of trifluoroacetic acid (TFA), and peptides were recovered using a StageTip (Thermo Scientific) containing C15 reverse-phase material. The peptides were eluted using 90% (v/v) acetonitrile, 0.1% (v/v) TFA, lyophilized, resuspended in 100 mM Tris-HCl (pH 7.4), 500 mM NaCl, and applied to a 5-μl Jacalin/Agarose (InvivoGen) micro-column equilibrated in the same buffer. The column was washed with 100 mM Tris-HCl (pH 7.4), 500 mM NaCl, and peptides subsequently were eluted by the same buffer containing 20 mM o-methylglucopyranoside. The eluate was acidified, and peptides were recovered on a C18 StageTip and eluted with a matrix solution of 2% (w/v) α-cyano-4-hydroxycinnamic acid, 70% (v/v) acetonitrile, 0.1% (v/v) TFA directly on the target plate for MALDI-MS analysis. The MS analysis was performed using an autotaxii III Smartbeam instrument operated in reflector or linear mode and calibrated in the mass range of 1000–3200 Da using Peptide Calibration Standard (both from Bruker Daltonics).

Characterization of cysteine 1121

Purified C4Ad was incubated overnight at 37°C with porcine trypsin (1:30 w/w to C4Ad). The digest was subsequently acidified by the addition of TFA, and peptides were separated by ultra performance liquid chromatography reverse-phase chromatography using a BEH300 C18 column (2.1 mm × 15 cm;
1.7 μm) operated by an ACQUITY UPLC System (Waters). The column was developed at a flow rate of 300 μl/min using a 1% B/min linear gradient of solvent B (90% v/v acetonitrile, 0.08% v/v TFA) in solvent A (0.1% v/v TFA). Fractions were analyzed by MS using α-cyano-4-hydroxycinnamic acid. The fraction containing the C4A isotope peptide (L1100-R1126) was analyzed in both MS and tandem MS mode to validate the detected cysteinylation of C4b (30). The lyophilized fraction was resuspended in 50 mM ammonium bicarbonate and 20 mM DTT and incubated at 37 °C for 20 min. After that, the sample was made 40 mM in iodoacetamide and incubated at room temperature for 20 min to allow for derivatization of free thiol groups. The sample was acidified, and peptide was recovered with a StageTip containing C18 reverse-phase material and eluted directly onto the target plate using a-cyano-4-hydroxycinnamic acid, as described above.

**Results**

The architecture of C4b

We prepared crystallizable C4b of human complement component C4 by cleaving purified C4 with C1s and trimming the N-linked glycans with the endoglycosidases Endo F1 and PNGase F. The resulting C4b formed fragile crystals with a typical size of 50 × 30 × 5 μm; after optimization of cryobuffers and mounting protocols, we obtained a complete data set extending to 4.2 Å resolution (Fig. 1C, Table I). The final model contained 1571 of 1637 C4b residues, whereas untraceable regions encompassed residues 1231–1255 (prepronumbering of C4) in the TE domain, the C-terminal end of the C4b α′-chain (residues 1414–1428), and the beginning of the γ-chain (residues 1454–1463). The final C4b model has Rwork/Rfree values of 0.217/0.273 (Table I), which is quite favorable considering the low resolution and the high Rmeas of the diffraction data. C4b is the major alterations in C4b are in the positions and orientations of the MG8, CUB, and TE domains. In C4, the TE is buried at the interface between the TE and MG8 domains (Fig. 2A); however, in C4b, it is highly exposed to nucleophiles from either solvent or an activator that may cleave the TE and form a covalent bond with Gln1013 (Fig. 2B). The location of the TE glutamine suggests that the C4b molecule attaches to activator surfaces with its bottom (the MG1, MG4, MG5, and TE domains) located closest to the activator, whereas the C345c domain points away from the activator (Fig. 2B).

In C4, the MG8 and MG3 domains are held in position by the anaphylatoxin domain; upon its release as C4a, the MG8 domain rotates by 53° and translates by 29 Å (Fig. 2, Supplemental Video 1). The MG3 domain also rotates by 18° and establishes hydrophobic interactions with the relocated MG8 domain in C4b. The structural consequences of C4 cleavage

Our structure reveals the dramatic changes occurring upon release of C4a and the subsequent conformational transition of nascent C4b (Fig. 2, Supplemental Video 1). Overall, they are very similar to the conformational changes occurring upon activation of C3 (see below). Compared with C4, the major alterations in C4b are in the top edge of the β-ring and form contacts with the MG2, MG3, and MG6 domains. The complement C1r/C1s, Uegf, Bmp1 (CUB) and TE domains are located on the long edge of the β-ring and interact with MG2 and MG1, respectively (Fig. 1D). The TE, MG1, MG4, and MG5 domains constitute the bottom of the molecule. The C-terminal C345c domain is positioned at the opposite end of the molecule, on top of the MG7–8 domains, and connected to the MG7 domain through a disulfide bridge. The two noncrystallographic symmetry–related C4b molecules in our crystals vary in the orientation of the C345c domain by a 5-Å translation and a 12° rotation, with no other noteworthy differences between the two molecules.

**FIGURE 1.** The function of C4 in complement and the structure of C4b. (A) Proteolytic processing and function of C4 and C4b in the CP (involving C1s) and the LP (involving MASP-1 and MASP-2). Active proteases are in bold font. C4b2a and C4b2a3b are the C3 and C5 convertases, respectively. (B) Schematic representation of the chain and domain structure of C4 and C4b. Four N-linked (N226, N862, N1328, N1391) and one O-linked (T1244) glycosylation sites are labeled. (C) The final model of C4b together with an omit 2mFo-DFc electron density map contoured at 1σ calculated following simulated annealing of a model with residues P1039-F1059 from the TE domain omitted. (D) The structure of C4b in two orientations. The position of the TE glutamine residue is indicated by the red sphere marked “Q.” The domains are colored as (B).
calculated structure factor, respectively, and $k$ is a scaling factor. $R_{\text{free}}$-factor is measured.

FIGURE 2. The structural changes accompanying activation of C4 and its transition to C4b. Overall views of C4 (RCSB PDB entry 4FXK) (A) and C4b (B) with close-up views of the two regions, the MG8 domain (close-up view, solid box) and the $\alpha'$-NT (close-up view, dashed box) in the two states. The two domains rotate by 35°, relocates to form the "ceiling" of the MG2 domain, and engages in new interactions with the MG8 domain. The MG7–6 domains and shown separately. Green spheres mark the beginning and end of the MG8 domain. The C4 and C4b structures are superimposed on the

Table I. Statistics for data collection and model refinement

<table>
<thead>
<tr>
<th>Data Collection and Processing</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Beam line/λ (Å)</td>
<td>ESRF ID23-1/0.9840</td>
</tr>
<tr>
<td>Space group</td>
<td>P2₁</td>
</tr>
<tr>
<td>Unit cell parameters</td>
<td>a, b, c (Å), β (°)</td>
</tr>
<tr>
<td>Resolution (Å)</td>
<td>49.54–4.20 (4.39–4.20)</td>
</tr>
<tr>
<td>$R_{\text{work}}$ (%)</td>
<td>28.5 (83.4)</td>
</tr>
<tr>
<td>$R_{\text{free}}$ (%)</td>
<td>7.15 (2.56)</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>98.3 (99.3)</td>
</tr>
<tr>
<td>Redundancy</td>
<td>7.04 (7.15)</td>
</tr>
</tbody>
</table>

Refinement

| Resolution (Å)                | 49.54–4.20 (4.39–4.20) |
| No. reflections               | 34,840                |
| $R_{\text{work}}/R_{\text{free}}$ (%) | 21.7/27.3 (30.3/39.4) |
| No. protein atoms             | 24,258                |
| RMS deviations from ideality  | 0.010/0.143          |
| Bond lengths (Å)/angles (°)   | 121.50, 161.08, 131.60, 107.262 |
| Average B values (Å²)         | 127.7                 |
| Ramachandran plot (%) Favored/allowed/outliers | 91.65/0.07/0.29 |

The values in parentheses are for outer resolution shell. $R_{\text{work}} = \left(\sum_{ijk} \sqrt{1/(n-1)}(\sum_{i} |F_{\text{obs}}(ijk) - \langle F(k) \rangle|) / \sum_{ij} \sum_{k} |F_{\text{calc}}(ijk)|\right)$ for the intensity of reflection $hkl$ measured $n$ times.

$R$-factor $= \left(\sum_{ijk} \frac{|F_{\text{obs}}(ijk) - k|F_{\text{calc}}(ijk)|}{\sum_{i} |F_{\text{obs}}(ijk)|}\right)$, where $F_{\text{obs}}$ and $F_{\text{calc}}$ are the observed and calculated structure factor, respectively, and $k$ is a scaling factor. $R_{\text{work}}$-factor is identical to the $R$-factor on a subset of test reflections not used in refinement.

$\chi^2$ Values given by MolProbity. RMS, root mean square. 

The solution structure of C4b

Crystallization of proteins can induce conformational artifacts by crystal packing. Furthermore, because C4b could only be crystallized upon trimming of the Asn-linked glycans, we decided to investigate the solution conformation of C4b with glycans by performing SAXS experiments (Supplemental Fig. 1). During RB refinement with CORAL, four RBs were defined within C4b corresponding to all MG domains, as well as the CUB, the TE, and the C345c domains (Fig. 3A). With all RBs locked in the positions as in the C4b crystal structure and modeling of residues not modeled in the crystal structure, a mean $\chi^2 = 2.99$ of the fit of the model to the experimental SAXS data was observed. With four glycans added as separate RBs, the mean $\chi^2$ decreased to 2.12, justifying their addition. By allowing the CUB and C345c domains to move, mean $\chi^2$ further decreased to 1.82, and the additional release of the TE domain gave a $\chi^2$ value of 1.40. In the C4b crystal structure, a contact between the MG1 and TE domains was observed; however, in the CORAL models obtained with the above scenario, the TE domain separated from the MG1 domain. To explore this difference between the crystal and solution structures, we evaluated three restraints based on the MG1–TE domains contact with lengths of 6, 10, and 14 Å, which gave rise to mean $\chi^2$ values of 1.73, 1.51, and 1.43, respectively. This suggested that the contact observed in the

FIGURE 2. The structural changes accompanying activation of C4 and its transition to C4b. Overall views of C4 (RCSB PDB entry 4FXK) (A) and C4b (B) with close-up views of the two regions, the MG8 domain (close-up view, solid box) and the $\alpha'$-NT (close-up view, dashed box) in the two states. The two close-up views are related by a 180° rotation, because the domains shown in these views are located on opposite faces of C4 and C4b. The MG8 domain undergoes a large rotation and translation upon release of C4a. The $\alpha'$-NT is inserted between the MG2, MG3, and the linker region in C4, whereas it has moved out through the hole formed by these domains in C4b and now extends toward the MG7 domain. The C4 and C4b structures are superimposed on the MG1–6 domains and shown separately. Green spheres mark the beginning and end of the MG8 domain. The $\alpha'$-NT region is displayed in yellow, and the $\alpha'$CT/sulfotyrosine region is displayed in magenta. The TE glutamine ("Q") is shown as a red sphere.
crystalline state is stabilized by crystal packing, whereas in solution the two domains are separated slightly (Fig. 3C). As the final SAXS-derived model of C4b, we present the CORAL solutions derived from the scenario with the 14 Å MG1–TE contact restraint described above with the best $\chi^2 = 1.41$ (Fig. 3A, 3B).

The SAXS models indicate the structural organization of residues 1231–1255 within the TE domain that were not traceable in the C4b crystal structure, suggesting its flexibility, as already proposed by the structure of the C4b fragment C4dg (31). These residues protrude from the TE domain in a roughly U-shaped exposed loop in all SAXS models (Fig. 3A, 3D). In C4, this region is folded into an ordered loop stabilized by interaction with the C4 CUB domain. Another functionally important area in C4 is the $\alpha\beta$CT/sulfotyrosine region encompassing residues 1405–1427. In C4, residues 1404–1415 form an $\alpha$-helix that is followed by the three sulfated tyrosines (Y1417, Y1420, Y1422) of importance for C4 interaction with C1s and MASP-2 (21, 32). In contrast, residues 1404–1413 adopt a random coil conformation in C4b, whereas the residues 1414–1427 could not be traced. Our SAXS modeling suggests that, in solution, the $\alpha$CT region runs across the MG7 domain roughly parallel to the $\alpha\beta$NT (Fig. 3C). In addition to the position of the TE domain, another major difference between the crystal structure and the SAXS model is the reorientation of the C345c domain by almost 90° (Fig. 3C). We tested the significance of this by repeating the restrained CORAL scenario described above where the C345c domain was fixed together with the MG domains as observed in the crystal structure, but this increased the average $\chi^2$ from 1.41 to 1.52, justifying release of the C345c domain during refinement. In conclusion, our SAXS studies confirm the overall conformation of C4b observed in the crystalline state but also identified two significant deviations from the crystal conformation in solution. Furthermore, the SAXS modeling suggests the approximate location of functionally important and exposed regions that could not be modeled in the crystal structure because of flexibility.

Two posttranslational modifications in the C4 TE domain

Human C4A recovered from cerebrospinal fluid was identified to contain an O-linked core 1 glycan (Gal-GalNAc-Thr; T-Ag) at Thr1244 (8), meaning that it is exposed in the loop 1231–1255 discussed above (Fig. 3D). To confirm these findings, we subjected the purified C4d fragment of C4A (C4Ad) to proteolytic cleavage using both trypsin and Asn-C. To enrich for O-linked glycopeptides, the generated peptides were fractionated by lectin-affinity chromatography using Jacalin/Agarose. Analysis of the isolated peptides identified peaks representing the A1240-R1248 peptide ($m/z_{calc}$ 895.5) modified by Gal-GalNAc (365.3 Da; $m/z_{obs}$ 1260.63) or by its sialylated variant (656.6 Da; $m/z_{obs}$ 1551.52) (Supplemental Fig. 2A). To further validate that these ions indeed represented the A1240-R1248 peptide, we subjected $m/z_{obs}$ 1260.63 to MS/MS analysis (Supplemental Fig. 2B). This analysis generated fragment ions representing the peptide containing GalNAc only ($m/z$ 1098.51) and the peptide with no carbohydrate moiety ($m/z$ 895.39). Moreover, we could assign ions representing the y2 and y4 fragments of the unglycosylated peptide (Supplemental Fig. 2B). Our data cannot
conclusively distinguish between residues S1242 or T1244 as carrier of the \( O \)-linked glycan, but they are consistent with the finding by Halim et al. (8) identifying Thr1244 as the modified residue. Interestingly, this residue is almost strictly conserved, which points to a functional importance of the \( O \)-linked glycan.

Additionally, human C4A encompasses an isotypic cysteine residue at position 1121. Another isotypic residue of C4B, H1125, was suggested to form an acyl-imidazole intermediate with the TE glutamin upon C4B activation (33). This implies that a free thiol group at position 1121 in C4A could play a role in TE cleavage or formation of a de novo disulfide bridge with C1010 released from the TE in C4b. To evaluate the status of residue 1121, we separated tryptic peptides generated from C4Ad by reverse-phase chromatography and analyzed the resulting fractions by MALDI-MS. One fraction presented an ion of \( m/z \) 3194.89, corresponding to the peptide encompassing C1121 (\( m/z_{\text{calc}} \) 3076.35) modified by cysteinylation (119 Da; \( m/z_{\text{calc}} \) 3194.89) (Supplemental Fig. 2C, upper panel). To confirm this, we subjected the peptide to reduction and alkylation, producing the \( S \)-carbamidomethylated peptide represented by \( m/z \) 3132.91 (Supplemental Fig. 2C, lower panel). The position of the modifications was validated further by tandem MS analysis, showing that C1121 is cysteinylated in human C4A. This finding is consistent with previous studies showing that C4b generated by digestion of C4, which was isolated from pooled plasma and expected to contain almost equivalent amounts of C4A and C4B, has a maximum of one free thiol group/C4b molecule (34). Hence, human C4A does not contain a free thiol, and C1121 and C1010 are separated by 15 Å in C4b, further arguing that disulfide formation in nascent C4b is unlikely to occur (Supplemental Fig. 2D).

**Structural mapping of the C4 isoforms and allotypes**

C4 is the most heterogeneous protein of the complement system. In addition to the two isoforms of the C4 gene, C4A and C4B, 23 allotypes of the genes were reported. With the C4 and C4b structures available, we can map these variants onto the two major functional states of C4 (Fig. 4). This shows that 14 of these variable residues are solvent accessible each in C4b and in C4 (additional variable residues are located in the C4a moiety). Eleven of the variants are located in the TE domain, and seven of them are exposed in C4b (Fig. 4). These variants are the basis for the serological distinction between the Rodgers or Chido blood group Ags (35). Our analysis suggests that eight of the variable isotypic and allotypic residues are close to the activator in C4b (Fig. 4). Hence, similar to the differential reactivity toward nucleophiles of the nascent TE in C4A and C4B, the variation within surface-exposed residues close to the activator might broaden the repertoire of interactions that can occur between C4b and activators.

**Discussion**

During the last decade, a wealth of structures of complement proteins has profoundly improved our understanding of structure–function relationships within the complement system. C3 and C4 are important proteins of the complement cascade, because their proteolytic fragments C3b and C4b act as opsonizing molecules, as well as the central subunits in the C3/C5 convertases providing a platform for bringing together the catalytic subunit and the substrate for proteolysis. Although several crystal and solution structures involving C3b are known, there has been a complete lack of structural information concerning C4b. The present study provides this information and identifies the structural principles shared by C3b and C4b, as well as several important C4b-specific features.

Structural similarity of C3b and C4b was expected, and comparison of our C4b structure with various structures containing C3b indeed reveals a remarkable resemblance of the two paralogues. The MG1–8 domains in C3b and C4b are arranged in a very similar manner (Fig. 5A), which is likely to be dictated by the function of the MG4, MG5, and MG7 domains in both C3b and C4b as interactors with the substrate within the convertases (13) and their common regulators (see below). Furthermore, the structural similarity extends to the location of the CUB and TE domains observed in our C4b crystal structure and all known C3b-containing crystal structures. In particular, the MG1–TE interaction is preserved between crystal structures of C3b and C4b. Likewise, the presentation of the \( \alpha’ \)NT region on the surface of the MG7 domain (Fig. 3C) is shared between C3b and C4b.

The close conformational match between C3b and C4b further strengthens the idea that the CP proconvertase C4b2 and convertase C4b2a are similar in structure to their AP equivalents C3bB and C3bBb and the fluid-phase convertase formed by cobra venom factor (CVF) and Bb, CVFBb (13). Within the AP proconvertase, the C3b \( \alpha’ \)NT region is sandwiched between the C3b MG7 domain and factor B (14), and the virtually identical presentation of the \( \alpha’ \)NT region in C4b on top of the MG7 domain suggests that this will apply to the CP proconvertase as well (Supplemental Fig. 3A). This is in accordance with reduced C2 binding to C4b carrying mutations in two acidic clusters within the \( \alpha’ \)NT region (30). Additionally, in our model of the CP proconvertase, the highly acidic and flexible \( \alpha’ \)CT/sulfitotyrosine region has the potential to interact with nearby positively charged areas of C2 (Supplemental Fig. 3A, 3B).

Our C4b structure also provides a means for investigating the mysterious architecture of the CP C5 convertase C4b2a3b, in which the presence of C3b shifts the specificity from C3 to C5. The SAXS-based modeling of residues 1231–1255 in the TE domain of C4b shows that the loop is highly exposed to the solvent (Fig. 3D). This loop is known to be involved in the interaction of C4b with C3b within the CP C5 convertase (7). Therefore, the exposure of these residues and their accessibility for C3b might be necessary for establishing contacts between C4b and C3b. Additionally, the strictly conserved Ser1236 within the exposed loop can act as a nucleophile.
attacking the TE in nascent C3b, leading to the formation of covalent C4b–C3b heterodimers (7). The SAXS structure of C4b hints at how C3b can be located in the CP C5 convertase because Ser1236 is the primary contact point between C4b and C3b in this convertase. By placing the cleaved TE of a C3b molecule roughly at the apex of the exposed C4b loop, we can search for possible arrangements of the C3b molecule relative to C4b (Supplemental Fig. 3D). We assume that C3b does not change dramatically in conformation when it is incorporated into the CP C5 convertase. This leads us to suggest a model for the CP C5 convertase in which the longest principal axes of C4b and C3b are arranged roughly in parallel (Fig. 4B, Supplemental Fig. 3D). In the model, we chose to position the C3b binding sites for FH complement control protein (CCP)1–4 and CCP19–20 (Supplemental Fig. 3C) toward C4b, suggesting that binding of these modules to C3b would interfere with a close approach of C3b and C4b in the CP C5 convertase, and additional steric hindrance may exerted by the intervening CCP modules 5–18 of FH believed to be looping out between CCP4 and CCP19 (36, 37). The resulting model for the CP C5 convertase offers an explanation for prior biochemical data showing that FH, in contrast to factor B, inhibits lysis of erythrocytes (38) and that C3b covalently bound to C4b is protected from FH-dependent FI degradation (39). However, our model for the CP C5 convertase must be considered with caution, keeping in mind that conformation of the loop containing the C4b–C3b contact point (Ser1236) is modeled by the use of low-resolution SAXS data. If this loop adopts a radically different conformation in activator-bound C4b, the suggested parallel orientation of C3b and C4b in the CP C5 convertase may not be possible.

An important feature of our CP C5 convertase model is that the C3b molecule cannot obtain direct contact with C5 because of the proximity between the two TE domains of C3b and C4b, suggesting that C3b alters the substrate specificity from C3 to C5 solely by altering the conformation of the C4b2a complex. An overall putative model of the substrate-bound C5 convertase is displayed in Fig. 5B; the MG6 domain of C3b is close to the MG3 and MG8 domains of C4b. Because the latter two domains are adjacent to the substrate-binding C4b domains MG4 and MG7, respectively, it is conceivable that an allosteric effect leading to higher C5 affinity could be transmitted through C3b contacts with these C4b domains. Based on the ability of both anti-C4 and anti-C3 Abs to inhibit substrate binding by the CP C5 convertase, it was suggested previously that both C4b and C3b form direct contact with C5 (6). However, our model is also consistent with these data: the anti-C3 Abs might simply interfere with the C3b–C4b interaction proposed in this article.

By combining our C4b and CVF-C5 structures, we suggest that the C4b-binding area for the substrates C3 and C5 is within the domains MG4, MG5, and MG7. This suggestion is supported by studies showing that changes in the two variant residues in the MG5 domain of C4b, R477, and P478 (Fig. 4), to tryptophan and leucine, respectively, cause defects in C5 binding by the CP C5 convertase (40, 41), and alanine scanning of residues in their vicinity also results in mutant C4b molecules impaired in CP C5 convertase activity (42). Moreover, the α’CT is also located near C5 in our substrate-convertase model (Fig. 5B), which suggests that, as in the C4:MASP-2 complex, the α’CT region may bridge the substrate and the catalytic subunit. In summary, the structures of C3b and C4b are surprisingly similar. This allows us to combine structural studies of the AP convertases with biochemical and genetic data on the CP proteins and convertases. Thereby, we are able to propose overall working models for the CP convertases.

The C4b α’CT/sulfotyrosine region is an extremely acidic region harboring 11 negative charges within residues 1405–1427. We and other investigators provided data showing that it facilitates C1s/MASP-2 recognition of C4 (21, 43) by acting as electrostatic “Velcro” between the SP domain of the protease and the C4 anaphylatoxin domain. However, the observed changes within the α’CT region in C4b compared with C4 prompted us to speculate on other functions for this C4-specific region. Our SAXS modeling indicated that the α’CT region is associated with the MG7 domain and is located close to the C345c domain of C4b and the acidic α’NT region (Fig. 3C). Again, the low resolution of the SAXS data means that suggestions based on the resulting RB models must be treated with caution. Even if the conformation of the α’CT region is different from that proposed by SAXS, this highly negatively charged and flexible region has a length and flexibility that probably enable it to interact with C2, C2a, convertase substrates, and even regulators.

Our crystal and SAXS structures differ with respect to the location of the C345c, TE, and CUB domains, suggesting that C4b harbors some degree of flexibility, which may be a prerequisite for the function of the molecule. The flexibility of the C345c domain is a well-known phenomenon, because reorientation of this domain is also observed in structures involving C3b (14, 15, 44). In C4/C4b,
this flexibility is probably required for its multiple functions because it is both a carrier of C2/C2a protease in the (pro)convertases and a binding site for MASP-2 and C1s proteases during C4 cleavage (21). The ability of the TE domain to dissociate from the MG1 domain, as observed in the SAXS structure, may support C4b function as an opsonin by improving its chance for covalent attachment to the myriad of activator surfaces that it encounters. Very recent solution structures of C3b and C3(H2O) in 137 mM NaCl indicated that the TE domain of C3b can separate by up to 60 Å from the MG1 domain, whereas a compact conformation resembling the C3b crystal conformation was observed in 50 mM NaCl (45). It was suggested that the low salt concentrations used for crystallization of C3b induced the compact conformation with the TE–MG1 contact. We observed the opposite trend; our SAXS modeling of C4b in solution indicates that the TE domain can move slightly away from the MG1 domain in the presence of 100 mM NaCl, whereas C4b exhibits the MG1–TE interaction inside crystals grown in the presence of 400 mM MgCl2. Cofactor-assisted FI cleavage presumably requires a rather fixed conformation of C3b and C4b to dock C3b/C4b, the regulator, and FI simultaneously into a productive complex (46). Therefore, the crystal structures of C3b and C4b are likely to resemble a substrate conformation for cofactor-assisted FI degradation. This does not exclude that the TE and CUB domains may be more or less separated from the MG domains in other functional states of C3b and C4b.

Binding of FH CCP1–4 domains relies on interaction with the C3b α′NT region, the MG1, MO2, MG6, and MG7 domains, and the CUB and TE domains (46). Thereby, FH provides a platform for positioning FI correctly relative to the C3b CUB domain for subsequent cleavage, and this binding mode probably applies to membrane cofactor protein (MCP) and decay acceleration factor (DAF) as well (46). The cofactors themselves also have a conserved structure. Usually, three consecutive CCP domains are required for cofactor activity, and these three are linked in an almost linear structure (47), whereas a bend occurs between the third and the following CCP domains, which may not be required for cofactor activity. The high conservation between C3b and C4b, the conserved positions of FI cleavage sites in the CUB domains of C3b and C4b, and the conserved structure of the regulators led us to suggest a simplistic model for the interaction of C4b with its regulators (Supplemental Fig. 3E). This suggests that cofactors (C4BP, MCP, CR1, MCP, the vaccinia virus complement control protein VCP, and the rodent specific Crry) for FI degradation bind to C4b in a manner similar to that observed in the C3b–FH complex. In agreement with this proposal, C4c encompassing all of the MG domains and the α′NT region binds only three times weaker than Does C4b to C4BP (48). In the proposed binding mode, the CCP1 and CCP2 domains of C4b regulators would align with the C4b CCP1–CCP2 domains (Supplemental Fig. 3D). All C4b cofactors indeed have several lysines and arginines that would face the α′CT and α′NT regions in a C3b–FH-like binding mode. The importance of a cluster of basic residues exactly in the C4BP CCP1–CCP2 domains for cofactor activity against C4b, as well as the strong electrostatic nature of the intermolecular interaction C4b–C4BP, were shown earlier by mutagenesis and binding studies (49). Another regulator of C4b, DAF, likewise contains positively charged residues in its CCP2 domain and the linker between CCP2 and CCP3 that might interact with the two aforementioned acidic C4b regions.

In conclusion, determination of both the crystal and solution structures of C4b confirms the anticipated unifying structural principles for C3b and C4b, supporting their shared functions as opsonins and as platforms for binding the catalytic subunit and substrates in convertases. Additionally, our results highlight the flexibility of the TE domain loop 1231–1255 and the sulfotyrosine-containing α′CT region specific to C4 and suggest important functions of these elements within the CP convertases. Our study sets the scene for dissecting structural requirements of C4b binding of C2/C2a, the role of C3b in shifting the specificity of the CP C3 convertase to C5, and the recognition of C4b by DAF and the FI cofactors CR1, C4BP, Crry, and MCP.

Acknowledgments

We thank F.X. Gomis-Ruth for endoglycosidase expression vectors, the beamline staffs at PETRA III and ESRF for support during data collection, and Jan Skov Pedersen for help with SAXS data processing and RB refinement.

Disclosures

The authors have no financial conflicts of interest.

References


Figure S1. SAXS data analysis. A) Guinier plot (black dots) in the region \( s^*R_g < 1.3 \) with the linear approximation line (red). Residuals between experimental points and the Guinier fit are shown on the plot to the right. B) Data analysis with Gnom. The left panel displays the regularized scattering curve computed by Gnom (\( \log(I_{\text{reg}}) \)) in comparison to the data (\( \log(I_{\text{exp}}) \)). In the right panel the resulting pair distance distribution function with \( D_{\text{max}} = 175 \, \text{Å} \). C) Kratky plot of the scattering data.
Figure S2. Mass spectrometric analyses of the C4 post-translational modifications. A) The ion of m/z 1551.52 represents A1240-R1248 (A-R) modified by a core 1 O-linked glycan (T-antigen) composed of sialic acid (star), galactose (circle) and N-acetylgalactosamine (square). The ion of m/z 1260.63 represents the same structure without sialic acid. B) To verify the identity of the glycosylated peptide the ion of m/z 1260.63 was subjected to MS/MS analysis and identified fragment ions are indicated (see text for details). C) The status of the isotypic C1121 was evaluated by separating tryptic peptides by reverse-phase chromatography. One fraction encompassed m/z 3194.89 representing L1100-R1126 modified by cysteinylation (top). Reduction (DTT) and subsequent alkylation using iodoacetamide (IAA) of this fraction produced the S-carbamidomethylated peptide represented by m/z 3132.91. D) The spatial position of C1010 and C1121 in the TE domain of C4b with the distance between the residues indicated. The glutamine residue of the thioester is indicated as a red sphere.
Figure S3. Models of the CP proconvertase, the C5 convertase-substrate complex and C4b-regulator complexes. A) Model of the CP proconvertase based on the structures of C3bB (RCSB entry 2XWJ), C2a (RCSB entry 2ODP), and C2b (RCSB entry 3ERB) illustrating the possible interaction of the C4b α'CT with C2. B) A putative electrostatic interaction between C4b and C2/C2a in the convertases. The electrostatic potential mapped on the surface of C2b (RCSB entry 3ERB, top left) and C2a (RCSB entry 2ODP, bottom left) reveals a buildup of positive charge. Boundaries between the domains in C2a are indicated by black lines. The electrostatic potential mapped onto C4b in the proconvertase model from panel A is shown to the right. The MG7 domain and the associated α'NT display a strongly negatively charged surface becoming even more negatively charged if the α'CT region associates with the MG7 domain as suggested by our SAXS modelling. The approximate location of C4b MG7 and the C345c domains are indicated by black boxes, while that of CCP1 within C2 is indicated by the broken line box. Electrostatic surface potentials were calculated using the APBS Pymol plugin and shown from negative $k_BT/ec = -3$ (red) to positive $k_BT/ec = 3$ (blue). C) The footprint of FH domains CCP1-4 (cyan) and CCP19-20 (orange) on the C3b surface. D) A schematic presentation of the modelling of the C3b-C4b dimer in the CP C5 convertase. The C3b molecule was placed on top of the C4b 1231-1255 loop so that the thioester glutamine of C3b ("Q", red sphere) forms a contact with the loop and then possible orientations of C3b were investigated by rotation of C3b. The final position of C3b (bottom) was chosen to make the C3b FH binding sites in panel C face towards C4b. E) A role of the C4b α'CT region in interaction with regulators is proposed by the comparison of the C4b structure with the C3b-FH CCP1-4 complex. FH was placed as an indicator of where the C4b specific regulators bind by superimposing C4b and C3b on
Crystal and solution structures of complement C4b

their MG1-5 domains. The similar positions of FI cleavage sites in the CUB domains of C3b and C4b are shown in the two panels to the right.
Movie S1. The conformational changes in nascent C4b. C4 from the C4:MASP-2 complex (RCSB entry 4FXG) was superimposed onto the crystal structure of C4b and energy minimized interpolated structures were calculated using “Morphinator” (www.morphinator.au.dk) and rendered in PyMol. In the first part of the animation, the red C4a is released, and the α’NT region (yellow) is shown in its C4:MASP-2 conformation. In the remaining part of the movie the α’NT region is shown in its C4b location. The relocation of the TE, CUB and MG8 domains are shown as a linear interpolation, but notice how this appears to conflict with the presence of the MG1 domain. This suggests that during the conformational change within nascent C4b the TE domain cannot reach the contact with the MG1 domain observed in the crystal structure without protruding more into the solvent as indicated by our SAXS studies. The C345c domain is colored sand, the CUB domain is shown in magenta while the TE domain is colored orange, remaining domains are colored cyan. The thioester is represented by the red sphere showing the location of glutamate 1013. The activator to which the exposed thioester in nascent C4b can cross-link is indicated as a grey surface at the bottom.