An essential aspect of innate immunity is recognition of molecular patterns on the surface of pathogens or altered self through the lectin and classical pathways, two of the three well-established activation pathways of the complement system. This recognition causes activation of the MASP-2 or the C1s serine proteases followed by cleavage of the protein C4. Here we present the crystal structures of the 203-kDa human C4 and the 245-kDa C4-MASP-2 substrate-enzyme complex. When C4 binds to MASP-2, substantial conformational changes in C4 are induced, and its scissile bond region becomes ordered and inserted into the protease catalytic site in a manner canonical to serine proteases. In MASP-2, an exosite located within the CCP domains recognizes the C4 C345C domain 60 Å from the scissile bond. Mutations in C4 and MASP-2 residues at the C345C-CCP interface inhibit the intermolecular interaction and C4 cleavage. The possible assembly of the huge in vivo enzyme–substrate complex consisting of glycan-bound mannan-binding lectin, MASP-2, and C4 is discussed. Our own and prior functional data suggest that C1s in the classical pathway of complement activated by, e.g., antigen–antibody complexes, also recognizes the C4 C345C domain through a CCP exosite. Our results provide a unified structural framework for understanding the early and essential step of C4 cleavage in the elimination of pathogens and altered self through two major pathways of complement activation.

Results
Complete MBL-MASP-2 or ficolin-MASP-2 complexes bound to an activating carbohydrate structure and C4, to our knowledge, have never been reconstituted in a format suitable for crystallization. Instead, we used a MASP-2 fragment comprising the CCP1, CCP2, and the serine protease (SP) domains (9). By mutating the MASP-2 catalytic site serine 633 to alanine, we could crystallize the C4-MASP-2 complex, and in addition we crystallized C4 alone. We then determined the crystal structures of C4 and the enzyme–substrate complex at 3.6- and 3.75-Å resolution, respectively (Table S1, Fig. 1 B and C, and Fig. S1). Structure determination was promoted by the presence of two C4-MASP-2 complexes in the crystal, the known structure of the active MASP-2 fragment (10), and the ability to compare models of C4 either unbound or bound to MASP-2. As a result, the R_free values of the structures are 0.27 and 0.24 for unbound C4 and C4-MASP-2, respectively (Table S1). C4 exists in two isoforms C4A and C4B, differing within six residues, and the vast majority of individuals express both isoforms (11). No attempt was made to separate the isoforms, and we assume that both structures contain a mixture of C4A and C4B (Figs. S1 A and S2 A). The structures revealed that C4 is structurally similar to its paralogues C3 and C5 (12-14), with six N-terminal MG domains (MG1–6) forming an irregular superhelical arrangement, the β-ring (Fig. 1 B). The TE, MG8, and CUB domains form the tightly packed α-chain superdomain. The C4 TE bond formed between Cys-1010 and Glu-1013 is buried between the MG8 domain and the α-helical TE domain as in C3 (12, 14). The C4α domain is wedged between the β-ring and the α-chain superdomain, whereas the MG7 and the C345C domains are located at the α-chain superdomain periphery (Fig. 1 B).

The ability of pattern-recognition molecules to bind foreign markers such as pathogen-associated molecular patterns is central to the innate immune defense. One such defense mechanism is complement, which is capable of recognizing molecular patterns associated with microbes and apoptotic or necrotic cells. Recognition causes activation of proteolytic enzyme cascades, resulting in cleavage of the complement proteins C3, C4, and C5. Fragments of these proteins have important effector functions through binding to host cell receptors and pathogen surfaces (1). In the lectin pathway of complement, four pattern recognition molecules, mannan-binding lectin (MBL) or H-, L-, or M-ficolin, may bind to surface-linked carbohydrates or acetyl groups on the surface of pathogens or damaged self-tissue (Fig. 14) (2). The CL-11 protein is also a putative pattern recognition molecule acting in the lectin pathway (3). Pattern recognition leads to activation of the pattern recognition molecule-associated MASP-2 protease (4), and this enzyme then cleaves the 203-kDa protein C4 into the fragments C4a and C4b. The nascent C4b fragment becomes covalently linked through its reactive thioester (TE) to the surface bearing the pattern recognized (5). C4b recruits thezymogen C2, and subsequent C2 cleavage leads to formation of the surface-anchored and proteolytically active C3 convertase C4b-C2a, but C4b also contributes to immune clearance through interaction with the CR1 receptor (6). The classical pathway of complement, in which the MASP-2 paralogue C1s cleaves both C4 and C2, is initiated upon Clq recognition of, e.g., antigen–antibody complexes (7) and likewise results in C4b deposition and assembly of the C3 cleaving C4b-C2a complex on the surface recognized (Fig. 14). Such C3 cleavage leads to alternative pathway amplification and, subsequently, cleavage of the C5 protein as well (8). The downstream outcome of complement activation through the lectin and classical pathways is therefore C3 and C5 fragments, which have important effector functions through binding to host cell receptors and pathogen surfaces. The activation of complement receptors ultimately elicits inflammatory responses directing immune cells and molecules to the point of infection, tagging of pathogens for phagocytosis, lysis of pathogens, and stimulation of the adaptive immune response (1).
MASP-2 Interactions with the Scissile C4 Region. All three MASP-2 domains are in contact with C4, yielding an overall intermolecular interface of 1,800 Å² (Fig. 1C and Figs. S2C and D). The CCP domains contact the C4 C345C domain with an approximate interface area of 500 Å². The MASP-2 SP domain interacts with an extended loop comprising C4 residues Asp-748–Ile-760 (Fig. 2A–C). We shall refer to this loop as the R loop, which comprises the scissile bond region P2–P1–P1′–P2′ (Gln-755–Arg-756–Ala-757–Leu-758). In addition, the SP domain forms electrostatic interaction with the C4 sulfotyrosine region (see below) and a few contacts with the anchor region connecting the C4 MG8 and the C435C domains. The total interface area between the SP domain and C4 is ~1,300 Å². The SP domain loops A, B, E, D, 3, and 2, named according to ref. 10, interact with C4 (Fig. S2C). Whereas the R loop is bound identically to the SP domain in the two copies of the C4-MASP-2 complex in our crystal, there is a slight difference of the orientation of the SP domain relative to C4 residues outside the R loop (Fig. 2A). This flexibility is made possible by the extended R-loop conformation and makes one complex tighter and with better electron density for the SP domain; we describe this complex below. Because of the resolution of the C4-MASP-2 structure, hydrogen bonds and salt bridges mentioned below should be considered putative, although the presence of two copies of the C4-MASP-2 complex result in a better effective resolution than the nominal 3.75 Å.

The C4 R loop has the conformation of a twisted U, with a bend at C4 Gly-750–Gln-751 caused by van der Waals interactions between C4 Ala-752–Leu-754 and Met-658 in loop 2 of MASP-2 (Fig. 2B and Figs. S1B and S3A). The P1 arginine side chain is inserted into the MASP-2 S1 pocket and is held by hydrogen bonds to MASP-2 Ser-628 and -657 and an electrostatic interaction with Asp-627. The main chain oxygen of Arg-756 is placed in the oxyanion hole formed by the main-chain N–H of MASP-2 Gly-631 and Ser-633–Ala (Fig. 2C). In silico mutation of MASP-2 Ala-633 to serine in our structure suggests that the serine side chain from the catalytic triad can come within 2.8 Å of the P1 carbonyl C atom. P1′–P4′ residues Ala-757–Ile-760 adopt an extended conformation stabilized by putative main chain hydrogen bonds with MASP-2 Thr-467 in loop A and Gly-656. In a number of serpin–protease Michaelis complexes (15), likewise with their P1 side chain accommodated into the S1 site, the conformation of their P2–P2′ residues is rather similar to that of P2–P2′ residues in C4-MASP-2 (Fig. S3D). Furthermore, the main-chain conformations of C4 Glu-755–Glu-759 and residues Thr-29–Cys-33 in the MASP-2 inhibitor SGM1-2 bound to MASP-2 (16) strongly resemble each other (Fig. S3E), most likely due to the P1 residue carbonyl interaction with the oxyanion hole and main-chain hydrogen bonds between the P2′ residue and MASP-2 Thr-467 in both cases.

MASP-2 Exosites. Within residues 1,405–1,427 located at the C terminus of the C4 α-chain are three sulfotyrosines (Fig. S2D), and together with seven glutamates/aspartates and the terminal carboxyl group, these provide 11 negative charges to this region. Residues 1,415–1,420 fold into an extended region sandwiched between a large, well-conserved, positively charged surface patch on the MASP-2 SP domain and the likewise positively charged C4a domain (Fig. 2D). The SO₃ group of C4 Tyr-1,417 interacts electrostatically with MASP-2 Lys-503 (Fig. S1C), whereas C4 Asp-1,419 faces MASP-2 Lys-450, Arg-578, and Arg-583. In the opposite direction, C4 Asp-1,416 is directed toward Lys-744 in C4a (Fig. 2D). These long-range electrostatic interactions suggest that the C4 sulfotyrosine region acts as flexible electrostatic “Velcro” between MASP-2 and the C4a domain.

The CCP domains in both MASP-2 (9, 17, 18) and C1s (19, 20) have earlier been suggested to be important for C4 cleavage, but the regions of C4 possibly interacting with the CCP domains have never been identified. Our C4-MASP-2 structure reveals the MASP-2 regions forming what we will refer to as the CCP exosite. Conserved residues from both CCP domains provide an open negatively charged binding patch for four arginines within the C4 region 1,716–1,725 (Fig. 3A–E and Fig. S2D) located immediately before the large C-terminal helix of the C345C domain. MASP-2 CCP1 Glu-333—strictly conserved in C1s and MASP-2—forms long-range electrostatic interactions with C4 Arg-1724. The main chain of MASP-2 Asp-365 interacts with Thr-1,721, and the aspartate side chain is facing C4 residues Arg-1,716, -1,719, and -1,724 (Fig. 3A). Overall, the intermolecular interactions formed by MASP-2 with both the C4 C345C domain and the sulfotyrosine region appear to be dominated by electrostatic forces and hydrogen bonds.
In the C4 paralogue C3, the region equivalent to the MASP-2 interacting C4 region 1,716–1,725 is negatively charged, suggesting that electrostatic repulsion would prevent CCP exosite recognition of C3. Together with the C4-specific sulfotyrosine region, this feature could explain the strong preference for C4 over C3 as MASP-2 substrate (9).

Mutations in the CCP Exosite Affect C4 Recognition. To validate the physiological relevance of the MASP-2 CCP exosite, we mutated residues in the CCP1 and CCP2 domains with surface-exposed side chains facing C4 in the complex to arginines aiming at introducing steric hindrance and electrostatic repulsion. These MASP-2 variants were associated with MBL bound to the yeast carbohydrate mannan, causing MASP-2 activation and cleavage of C4 and resulting in C4b deposition on the mannan surface (Fig. 3F). Mutating CCP1 Glu-333 or CCP2 Asp-365 severely reduced MASP-2 cleavage of C4, and the double mutant was unable to mediate cleavage. Mutation of CCP1 Pro-340 or CCP2 Pro-368 resulted in ~60% cleavage activity, whereas the double mutant had a poor cleavage activity. The recently reported CCP1 Lys-342–Ala mutation (17) had no effect in accordance with the absence of Lys-342 from the C4 interface. All our mutants bound MBL and autoactivated as well as the wild type (Fig. S4 A and B), suggesting that impaired C4 cleavage activity was caused solely by weakened C4-MASP-2 interactions. We confirmed this hierarchy of activity for the MASP-2 variants in an even more physiological setting, where we measured C4 deposition onto mannan in MASP-2–deficient plasma reconstituted with the various forms of MASP-2 (Fig. S4C). In addition, surface plasmon resonance (SPR) measurements showed that the recombinant C4 C345C domain interacted with immobilized MASP-2 (Fig. S4D), with an apparent equilibrium dissociation constant $K_D$ of 21 nM and mutations in C4 residues interacting with MASP-2 in our structure affected this interaction. In summary, our functional experiments and biophysical measurements confirmed the importance of the C4-MASP-2 interactions at the CCP exosite observed in our crystal structure.

C1s Recognition of C4 and MASP-1 Discrimination. Based on prior data concerning the importance of the C1s CCP domains for C4 cleavage (19, 20), we hypothesized that the MASP-2 CCP exosite is conserved in C1s and interacts with the C4 C345C domain. In support of this hypothesis, we were able to inhibit C1s-mediated
C4 cleavage by the addition of recombinant C345C domain in 30-fold molar excess relative to C4. Single mutations at five different C345C residues located in the MASP-2 interface all decreased the ability of the mutated C345C domain to inhibit C4 cleavage by C1s (Fig. 4A–C). Similar to the C4-MASP-2 interaction, SPR measurements showed that the C1s interaction with the C4 C345C domain had an apparent dissociation constant $K_D$ of 37 nM (Fig. 4D), and mutations in C4 Arg-1,724 and Thr-1,721 affected this interaction. C1s (9) qualitatively shares the electrostatic properties at the CCP exosite, with MASP-2 supporting an interaction between C1s and C4 C345C similar to what we observed for C4-MASP-2 (Fig. S5A). Because a positively charged surface on the MASP-2 SP domain located next to the C4 sulfotyrosine region is also present on C1s (Fig. S5B), an interaction between the C1s SP domain and the C4 sulfotyrosine region may also occur. This surface area is identical to a very recently identified exosite in the C1s SP domain interacting with the C4 sulfotyrosine region, and mutations in this exosite decrease the efficiency of C4 cleavage (21). C4 sulfotyrosine interaction with C1s would also offer an explanation for the 10-fold higher concentration of C1s required to cleave nonsulforylated C4 compared with sulforylated C4 (22). Together, our experiments, prior functional data, and structural comparisons strongly support that the C1s SP domain interacts with the C4 SO3–Tyr region and that C1s contacts residues 1,710–1,725 in the C4 C345C domain through a CCP exosite. The inability of MASP-1 to cleave C4 may be due to the electrostatic properties of MASP-1 being different from those of MASP-2 and C1s, especially at the CCP site. Furthermore, a large MASP-1 insertion in the SP domain loop B appeared to prevent insertion of the C4 R loop into the catalytic site (Fig. SS C–E). MASP-3, the third protease of the lectin pathway, is an alternative splice product of the MASP-1 gene and contains a different SP domain but the same CCP modules. Their electrostatic properties probably explain why MASP-3, like MASP-1, is unable to cleave C4 and consequently cannot take over the function of MASP-2 in MASP-2–deficient or –depleted plasma.

**Conformational Changes in Enzyme and Substrate.** The conformation of the CCP1–CCP2 domain tandem is identical in C4-bound and zymogen MASP-2, whereas the SP domain has rotated by 24–29° (Fig. S4A) relative to the CCP2 domain of zymogen MASP-2 (23) or active unbound MASP-2 (10). The ability of MASP-2 to undergo this hinge movement at the CCP2–SP linker allows simultaneous substrate recognition by the CCP exosite and the SP domain. C4 binding induces only minor structural reorganization within the SP domain, suggesting that the exosite–C4 interactions do not further activate the SP domain. In accordance with this finding, the recombinant MASP-2 SP domain cleaves C2 efficiently, whereas at least the CCP2 domain is required for fast C4 cleavage (9). C4 also undergoes significant conformational changes upon binding to MASP-2 in two regions around the scissile bond.
unbound C4, the C4α α-helix, the R loop, and downstream Nt-α residues are disordered. In the MASP-2 complex, C4a formed a four-helix bundle, and residues downstream of the R loop were ordered and held between the C4α α-helix and the MG3 and MG8 domains (Fig. S6). MASP-2 recognition of the R loop most likely stabilized the conformation of downstream residues in the Nt-α region, and the CCP exosite interaction induces a rotation of the C4 C345C domain, which resulted in a 10° rotation of the MG1-5 domains (Fig. 5B and Movie S1). Hereby the MG3 domain could capture the C4α α-helix between itself and the rest of the C4α domain, which presumably also locks the conformation of the Nt-α region located downstream of the scissile bond. Hence, not only does the MASP-2 CCP exosite interaction help to correctly orient the protease toward C4, but through a relay of domain movements it probably also stabilizes the R-loop conformation, which feasibly facilitates cleavage.

Discussion

Our structures of C4 and especially the C4-MASP-2 complex represent a crucial step forward in reaching a structure-based understanding of the series of events starting with pattern recognition in the lectin or classical pathway. The crystal structures presented here also allow us to compare how the structurally similar C3, C4, and C5 are cleaved by two distinct types of proteolytic enzymes, the C1s and MASP-2 based enzymes cleaving C4 in the lectin and classical pathway vs. the factor B and C2-based C3 and C5 convertases in the alternative and terminal pathway. The substrate recognition mechanism revealed by the C4-MASP-2 complex is fundamentally different from that identified in the C3 and C5 convertases (Fig. 5 C and D and Fig. S7). When C4 is cleaved by MASP-2 or C1s bound to a pattern recognition molecule, exosites in the MASP-2 or C1s CCP domains are required (9, 17, 18), and our results imply that their function is in the recognition of the C4 C345C domain. In the C3 and C5 convertases, large exosites are present in the noncatalytic C3b or C4b subunits recognizing the MG4, MG5, and probably also the MG7 domains, which are all located far from the C345C domain (24). In addition, the catalytic subunit of the convertases approaches much more horizontally relative to the scissile loop region in the substrate compared with the C4-MASP-2 complex (Fig. 5 C and D). C4 is a remarkable substrate because it is cleavable by two types of proteolytic enzymes—MASP-2 or C1s as the catalytic subunit—becoming activated upon pattern recognition taking place within structurally very variable and unpredictable environments. In contrast, when C3 and C5 are substrates for the convertases, the primary substrate-binding noncatalytic C3b or C4b subunit is likely to be rather rigid (24). Substrate recognition by the convertases is therefore likely to be only slightly influenced by the environment in which the convertase is present. In C3 and C5, the scissile bond is present in or close to an exposed loop with five to nine disordered residues (Fig. S7B), and in C5 the P1 residue Arg-751 is not even exposed, suggesting that this residue is not involved in the initial recognition between the convertase and the substrate. However, in C4 the scissile bond is located in a region with 22 disordered residues. The much higher flexibility of the scissile bond region in C4 offers an elegant solution to the problem of simultaneously being a substrate for two different proteolytic enzymes deposited in extremely variable environments. This high degree of flexibility may allow C4 to form initial contacts with the SP domain of C1s or MASP-2, approaching the protease from quite different orientations, and in combination with the flexibility of the MASP-2 and C1s containing proteolytic complexes discussed below, the flexibility of the scissile bond region may significantly promote formation of productive enzyme–C4 complexes.

Within a full enzyme–substrate MBL-MASP-2-C4 complex bound to a carbohydrate layer through the MBL carbohydrate recognition domains (CRDs), MASP-2 is bound to the MBL collagen stem through its CUB domains and firmly holds the substrate through contacts with primarily the C4 C345C domain and the C4 scissile bond region. This finding implies that the MG1–MG4–MG5 domains at the opposite end of C4 (Fig. 1 B and C) are expected to be oriented approximately in the same direction as the MBL CRDs toward the carbohydrate layer. In such an orientation, the TE would be directed into the carbohydrate layer during a C3b-like conformational change of nascent C4b, as previously suggested for convertase-bound C3b (24). The MBL-MASP-2-C4 complex is anticipated to be structurally flexible for several reasons. Considerable flexibility is present in MBL at the predicted kink of the collagen stem, and likewise the orientation of the CRDs relative to the collagen stem is variable. MBL is also heterogeneous by stoichiometry, with the smallest form being dimer of polypeptide trimers formed through the collagen stem, but such dimeric MBL–MASP complexes do not bind carbohydrate patterns with a high enough avidity to allow for efficient activation of complement. MBL trimers and tetramers are the dominating form occurring naturally in humans, but higher oligomers are also present, and their intersubunit orientation is quite variable (25). In MASP-2, flexibility is presumably present within the CUB2 domain (26), at the CUB2–CCP1 linkage (27), and at the CCP2–SP linkage, as shown here. Our C4-MASP-2 structure surprisingly revealed that the orientation between MASP-2 and C4 is also variable, and autolysis of unbound C4 suggests high mobility of the scissile bond region and regions surrounding it. Obviously, all of these sources of flexibility increase the likelihood of forming productive MBL-MASP-2-C4 complexes within the highly variable glycan environment, and similar levels of flexibility can be expected for the ficolin-MASP-2-C4 and the C1-C4 complexes in the classical pathway, because ficolins and C1q share their basic architecture with MBL, as is also the case for MASP-2 and C1s.

Activation of complement triggers an aggressive proteolytic cascade, creating potent inflammatory effector molecules, with the risk of host tissue damage if not kept under tight control by an array of soluble and membrane-bound regulators (1, 28). Uncontrolled complement activation is seen in association with, e.g., sepsis, myocardial infarction, ischemic stroke, rheumatoid arthritis, glomerulonephritis, myasthenia gravis, lupus, age-related macular degeneration, and atypical hemolytic uremic syndrome (28, 29). Of particular relevance for the present study, MASP-2 inhibition has proven efficient in protection against ischemia/reperfusion injury (30).

Our C4-MASP-2 structure reveals in details the intermolecular interactions of a key event in the lectin pathway and should therefore facilitate further development of inhibitors of this pathway. Moreover, the exosite part of the C4 recognition by MASP-2 identified by our structure is likely to be common to both lectin and classical pathways. Thus, in conclusion, our results provide an essential structural framework for future studies of the lectin and classical pathway of complement activation and rationalize many prior functional studies of both pathways.

Materials and Methods

Human C4 was purified from human plasma by anion exchange chromatography, and recombinant MASP-2 CCP1–CCP2–SP S633A was prepared by refolding and activated with MASP-1 (9). Crystals of C4 and C4-MASP-2 were obtained by vapor diffusion and cryoprotection before data collection. The structures were determined by molecular replacement aided by single-wavelength anomalous diffraction phases for a TaBr12 derivative for the C4 structure. Rebuilding was done in O (31), and refinement was performed in PHENIX. REFINEx (32). Full-length MASP-2 variants were expressed in HEK293F cells (33) and activated by their association with recombinant MBL bound to mannose-coated plates. C4 was then added and digested MASP-2-mediated cleavage, and cleaved C4 was quantitated with a biotinylated monoclonal anti-C4 antibody through detection with europium-labeled streptavidin and fluorometry (34). The ability of the MASP-2 variants to mediate C4 deposition was
also investigated in the presence of MASP-2-deficient plasma (33). Recombinant wild-type and mutated C4 C345C was expressed in Escherichia coli, purified by Ni²⁺-chelate chromatography and gel filtration, and tested in cleavage experiments with C1s at a molar ratio of 30:1 for (C4 C345C):C4. For SPR measurements, MASP-2 or C1s were immobilized on CM5 sensor chips, and the interaction with recombinant C4 C345C was measured at a flow rate of 5 μL/min. Data were analyzed by global fitting to a 1:1 binding sensor chip. A detailed description of experimental methods may be found in SI Materials and Methods.

ACKNOWLEDGMENTS. We thank the staff members at the European Synchrotron Radiation Facility and SOLEIL beamlines for help with data collection; L. Kristensen and J. Balcerz for protein purification; S. Degn for help with expression; and A. M. Bundgaard for help with SPR. G.R.A. was supported by Danscatt, the Lundbeck Foundation, the Lundbeck Foundation Nanomedicine Centre for Individualized Management of Tissue Damage and Regeneration, and a Hallas-Møller stipend from the Novo-Nordisk Foundation. P.G. was supported by Hungarian Scientific Research Fund (OTKA) Grant NK77798 and National Development Agency Grant KMOP-1.1.2-07/1-2008-0003. J.D. was supported by the János Bolyai Foundation.

Supporting Information

Kidmose et al. 10.1073/pnas.1208031109

SI Materials and Methods

C4 and MASP-2 CCP1-CCP2-SP Purification. Purification was performed at 4–6 °C, and all columns were from GE Healthcare. Thawed citrated human plasma was made 10 mM in benzamidazine (BZA), 3 μg/mL in pancreatic trypsin inhibitor, 60 mM in BaCl2, and 25 mM in trisodium citrate. After centrifugation, the supernatant was loaded on a Q-Sepharose FF column equilibrated in 10 mM Tris-HCl, pH 7.5, 100 mM NaCl, 50 mM e-Amino-Caproic Acid, 5 mM EDTA, 0.5 mM phenylmethylsulfonylfluoride (PMSF), and 1 mM BZA. Bound protein was eluted with a 100–500 mM NaCl gradient. To C4 containing fractions were added PEG6000 to a final concentration of 12% (wt/wt). After centrifugation, the pellet was redissolved in 200 mM NaCl, 20 mM Tris-HCl, pH 7.5, 1 mM BZA, and 0.5 mM PMSF and loaded on a Q-Sepharose HR column equilibrated in the same buffer and eluted with a 200- to 800-mM NaCl gradient. C4 containing fractions were diluted with water before loading on a Source 15Q column equilibrated in 100 mM NaCl, 20 mM Tris-HCl, pH 7.5, and 1 mM BZA and eluted with a 200- to 800-mM NaCl gradient. Pooled fractions were dialyzed against 20 mM Hepes, pH 7.7, and 100 mM NaCl. Recombinant MASP-2 CCP1–CCP2–SP S633A was prepared as described (1). The refolded protein was loaded on a Source 30Q column, washed with 50 mM NaCl, and eluted with a 100- to 400-mM NaCl gradient in 10 mM Tris-HCl, pH 8.2, and 1 mM EDTA. Zymogen MASP-2 was cleaved with recombinant MASP-1 CCP1–CCP2–SP (2, 3) by using 5 μg/mL MASP-1 and 100 μg/mL MASP-2 in 180 mM NaCl, 10 mM Tris-HCl, pH 8.2, and 1 mM EDTA. After cleavage, it was loaded on a Source 15Q column and eluted with a 50- to 250-mM NaCl gradient in 10 mM Na-phosphate, pH 6.8. Cleaved MASP-2 was dialyzed against 300 mM NaCl, 5 mM Tris-HCl, pH 8.5, and 0.5 mM EDTA.

Structure Determination, Analysis, and Modeling. To crystallize the enzyme–substrate complex, C4 (15 mg/mL) was mixed with CCP1–CCP2–SP MASP-2 S633A (2.9 mg/mL) in a 1:1 molar ratio. Crystallization was performed by mixing equal volumes of pre-mixed C4-MASP-2 complex and reservoir [100 mM Tris-HCl, pH 8.0, 2% (vol/vol) 1,4-dioxiane, and 12–14% (wt/wt) PEG 3350]. Before flash freezing, crystals were gradually transferred to 25% (wt/vol) PEG 3350 and 15% (vol/vol) glycerol. To crystallize C4, it was concentrated to 6 mg/mL and mixed with an equal volume of reservoir [0.9 M sodium citrate, pH 6.4, 0.25 M potassium chloride, 1% (vol/vol) ethanol, 3% (vol/vol) acetonitrile, 6% (vol/vol) ethylene glycol, and 10 mM spermine] at 19 °C. Crystals were transferred to a solution containing 37.5% (vol/vol) PEG 400, 50 mM MesNaOH, pH 6.4, 0.25 M KCl, 1% (vol/vol) ethanol, 3% (vol/vol) acetonitrile, 6% (vol/vol) ethylene glycol, and 10 mM trimethylacetic acid before flash freezing. The latter increased resolution and decreased anisotropy, but no sites were identified. For the TaBr2+ derivative, the compound was resuspended in reservoir solution and added to the crystals for soaking. Before flash freezing these crystals were briefly transferred to 0.9 M sodium citrate, pH 6.4, and 30% (vol/vol) glycerol. Diffraction data collected at European Synchrotron Radiation Facility (ESRF) ID29 or SOLEIL PROXIMA1 (Table S1) were processed with XDS (4) and molecular replacement carried out with PHASER (5). Rebuilding was done in O (6) after refinement of intermediate models with PHENIX.REFINE (7). A substantial improvement was obtained after combination of single-wavelength anomalous diffraction phases from the TaBr2+ derivative (figure of merit, 0.33; phasing power, 2.5) calculated with CNS (8) with model phases from PHENIX.REFINE followed by solvent flipping in CNS. Inclusion of the isomorphous differences between the C4–TaBr2+ data and the native C4 data in phasing did not improve the resulting density maps, probably due to significant differences in cell parameters (Table S1). Because the sequence identity between human C3 and C4 is 28%, several cycles of manual rebuilding and refinement were carried out, but the electron density for the MG3 and MG4 domains remained poorly defined. At Rfree = 35% for the structure of bound C4, the C4-MASP-2 diffraction data became available and molecular replacement with Phaser using the intermediate C4 model and the X-ray structure of activated MASP-2 (PDB ID code 1Q3X) as search models placed two copies of both C4 and MASP-2. From this point structure determination was promoted by the presence of two C4-MASP-2 complexes in the crystal, the known structure of active MASP-2 (9), and the ability to compare models of C4 either unbound or bound to MASP-2. In an averaged 2mFo – DFo density for the C4-MASP-2 complex generated with AVE (10) using individual domains of C4 or MASP-2 for averaging masks and generation of noncrystallographic symmetry (NCS) operators, the main chain of the MG3 and MG4 domains were easily traced. The C4 R-loop residues 750–761 were included in the mask for the MASP-2 SP domain, because these are interacting tightly. All subsequent rebuilding was done in such averaged electron densities, which could not be further improved by multiple crystal averaging using 2mFo – DFo or solvent-flattened densities resulting from the C4 and C4-MASP-2 structures. During refinement of C4-MASP-2, model restraints for MASP-2 derived from PDB ID code 1Q3X and restraints for parts of the C4 TE domain from the structure of C4Adg (PDB ID code 1HZF) were used together with tight NCS restraints applied domain-wise and secondary structure restraints as implemented in PHENIX.REFINE. Comparison with the C4-MASP-2 complex was also used to guide the iterative rebuilding of unbound C4, which was refined with secondary structure restraints in PHENIX.REFINE. For both structures translation/libration/screw (TLS) parameters were refined for the individual structural domains. The structures were refined with grouped B factors, but in the final refinement a few cycles of refinement with individual B factors were carried out as justified by a drop in Rfree. Models were validated with PROCHECK (11) and MOLPROBITY (12). The final C4 structure had 78.1% of the residues (excluding proline and glycine) in favored regions of the Ramachandran plot, 21.6% in allowed regions, and 0.3% in disallowed regions. The equivalent statistics for the C4-MASP2 complex were 79.3%, 20.2%, and 0.5%. Molecular graphics figures were prepared with PYMOL (13). Sequence alignments were done with CLUSTALW (14).

C4 Deposition Assay. MASP-2 variants were generated from a MASP-2 expression plasmid (15) by site-directed mutagenesis (Quikchange II XL; Agilent Technologies). Plasmids were mixed with Lipofectamin-2000 and OptiPRO SFM (Invitrogen) and used for transfection of HEK293F cells (Invitrogen), according to the manufacturer’s instructions. Cells were cultivated for 72 h in Freestyle 293 Expression Medium (Invitrogen), and the supernatants were harvested. FluoroNUNC plates were coated with mannann in 15 mM Na2CO3, 35 mM NaHCO3, and 15 mM Na2SO4, pH 9.6, and blocked with human serum albumin (HSA) in Tris-buffered saline [TBS; 10 mM Tris-HCl, 140 mM NaCl, pH 7.4 with 0.01% (wt/vol) Na2SO4]. The wells were washed in TBS/Tw/Ca2+ [TBS with 0.05% (vol/vol) Tween-20 and 5 mM CaCl2] and added 100 ng of rMBL (16) in 100 μL of TBS/Tw/Ca2+. After 2 h at room temperature, the wells were washed, and to 12 wells were added 50 ng of WT MASP-2 or mutants in 100 μL of...
TBS/Tw/Ca\(^{2+}\) and incubated overnight at 4 °C. The wells were washed, activation buffer (4 mM barbital, 145 mM NaCl, 2 mM CaCl\(_2\), 1 mM MgCl\(_2\), pH 7.4) was added, and they were incubated for 3 h at 37 °C. After wash, the wells were eluted with SDS/PAGE buffer [30 mM Tris-HCl, 10% (vol/vol) glycerol, 8 M urea, 3% (wt/vol) SDS, and 0.1% (wt/vol) bromophenol blue, pH 8.9] diluted 1:1 in TBS, and the eluate was analyzed by SDS/PAGE and Western blotting using an anti-MASP-2 antibody (1:3B7) (17). MASP-2 wild-type and mutant supernatants were analyzed without prior activation as above. The ability of the different MASP-2 variants to deposit C4 was tested as described (18). Mannan–MBL microrattle plate wells were generated as above, and dilutions of culture supernatant MASP-2 WT or mutants were added. After incubation overnight at 4 °C, the wells were washed and C4 was added, and after incubation for 1 h at 37 °C, the wells were developed by addition of biotinylated monoclonal anti-C4 antibody, followed by wash and addition of europium-labeled streptavidin. The amount of europium bound in the wells was measured by time-resolved fluorometry (Victor 3; PerkinElmer). The ability of the MASP-2 variants to mediate C4 deposition was also investigated in the presence of plasma. MASP-2–deficient plasma (15) was diluted to 1/200 in 1 M NaCl, 20 mM Tris Base, 0.05% (vol/vol) Triton X-100, 10 mM CaCl\(_2\), and 0.1% (wt/vol) HSA, pH 7.4; rMBL was added to 125 ng/mL, and rMASP-2 supernatant containing the MASP-2 variants was added. The mixtures were added to mannan-coated microrattle plate wells and tested for ability to C4 deposition as described above.

Inhibition of C1s by the C4 C345C Domain. DNA encoding C4 residues 1,591–1,744 and a cleavable His\(_6\)-tag (Genscript) was inserted in the pET32a(+) vector (Novagen). Mutations were made with the Quikchange Lightning Mutagenesis kit (Stratagene). Proteins were expressed in the Escherichia coli SHuffle T7 strain (New England Biolabs). Cells were suspended in lysis buffer (50 mM Hepes-NaOH, pH 7.5, 300 mM NaCl, 30 mM imidazole-HCl, pH 7.6, and 1 mM PMSF), sonicated, and centrifuged, and the supernatant loaded on a HisTrap FF crude column. Bound proteins were eluted with 500 mM imidazole

---

Fig. S1. Stereo representation of 2mFo − DFo electron densities from the C4 MASP-2 structure. MASP-2 labels are underlined. (A) Structures of C4 MASP-2 after refinement based on an input model with either the C4A sequence (blue carbon atoms, black labels) or the C4B sequence (green carbon atoms, green labels). The models are displayed together with an averaged omit 2mFo − DFo electron density contoured at 1σ calculated by omitting C4 residues 1,115–1,130. (B) Twofold averaged 2mFo − DFo electron density around the C4 R loop (red until Arg-756; yellow after) contoured at 1.2σ. The MASP-2 SP domain (gray carbon atoms) is shown as a Cα trace. (C) Averaged omit 2mFo − DFo electron density around the sulfotyrosine region contoured at 1σ calculated by omitting C4 residues 1,417–1,420. In contrast to Y1417, the sulfo group on Y1420 does not form direct electrostatic interactions with any positive charge and is therefore likely to be free to rotate around the ester bond resulting in weak omit electron density. (D) The C4 C345C region (brown carbon atoms) recognized by the MASP-2 CCP exosite (magenta carbon atoms) in averaged 2mFo − DFo omit density contoured at 1σ. C4 residues 1,716–1,724 and MASP-2 residues 333–348, 364–365, and 416–417 were omitted for map calculation.
Fig. S2. Conservation of MASP-2 and C4 regions involved in intermolecular contacts. (A) Domain structure of C4 and definitions of the chains. Domains are colored as in Fig. 1B. (B) Domain structure of MASP-2 and the chain structure in zymogen and activated MASP-2. Domains are colored as in Fig. 1C. (C) Alignment of the CCP1–CCP2–SP region of selected mammalian MASP-2 and C1s sequences. ■, cleavage site for activation; ★, C4 contact;▲, catalytic triad residue. Dashed squares indicate SP loop regions as defined by Perona and Craik (1) and applied to MASP-2 (2). (D) Alignment of regions involved in MASP-2 contacts in selected mammalian C4 sequences, residue numbering is according to human C4, for which both isoforms are shown. ■, C1s/MASP-2 cleavage site; ★, MASP-2 contact; ●, SO₃-tyrosine. In A and B, species abbreviations are as follows: HS, Homo sapiens; SS, Sus scrofa (pig); BT, Bos taurus (bovine); MM, Mus musculus. C and D were prepared with ALINE (3).

Fig. S3. Details of the R-loop interaction with MASP-2 and comparison of the R-loop main chain conformation to substrates of other SPs and a MASP-2 inhibitor. (A) Stereoview of the R loop (red carbon atoms until Arg-756; yellow after) bound to the MASP-2 SP domain (gray carbon atoms; labels underlined). (B) Footprint (MASP-2 residues within 3.8 Å of C4 shaded green) of C4 on the surface of the SP domain. (C) Footprint (green residues) of the MASP-2 SP domain on C4. D and E are related by a 180° rotation around a vertical axis. (D) Comparison of the overall path for the PS–P4′ region of the substrates in two SP serpin complexes (light blue cartoon, PDB ID code 1K9O serpin 1B·trypsin (1); green cartoon, PDB ID code 1SR5 (2) antithrombin·anhydrothrombin·heparin complex) and in the C4·MASP-2 complex (MASP-2, gray cartoon; C4 R loop, red and yellow). The MASP-2 loop regions are labeled (underlined) according to ref. 3. (E) Legend continued on following page.
Peptide main-chain conformation of the substrate C4 (red-yellow carbon atoms) and the inhibitor SGMI_2 (4) (green carbons, PDB ID code 3TVJ) P3–P2\(^{\prime}\) region bound to MASP-2. The protease is shown as a gray surface illustrating the deep S1 pocket accommodating the P1 arginine. MASP-2 active site residues His-483 and Ser-633 (here mutated to alanine) are colored blue and green, respectively.

Fig. S4. The autoactivation and the activity of rMASP-2 variants. (A) Culture supernatants containing the various rMASP-2 variants (indicated above the lanes) were analyzed by SDS/PAGE under reducing conditions and Western blotting using an anti-MASP-2 antibody. (B) Recombinant MASP-2 containing supernatants were incubated in wells with rMBL bound to mannan, washed, and left at 37 °C. The wells were subsequently washed, and MBL and MASP-2 were eluted from the wells with SDS/PAGE sample buffer. The samples were then analyzed by SDS/PAGE at reducing conditions, followed by Western blotting developing with anti-MASP-2 antibody. The migration of the molecular mass markers is given on the right in kilodaltons. The position of the intact polypeptide chain (A+B in A) and the position of the A-chain seen after cleavage (B) are indicated on the left-hand side. When an MBL/MASP complex binds to an activating surface, pro-MASP-2 may autoactivate. When this happens, the polypeptide chain of MASP-2 is cleaved into two chains (the 52-kDa A and the 31-kDa B chain) held together by a disulfide bond. (C) The activity of the different MASP-2 variants in the presence of MASP-2 deficient plasma. rMBL and dilutions of rMASP-2 supernatants were added to MASP-2–deficient plasma, and the solutions were added to microtiter plate wells coated with mannan. After incubation and wash, C4 was added, and the wells were incubated and developed for C4 deposition as described. The result is depicted as for the experiment using purified component (Fig. 3F). (D) Assessment by SPR of the interaction between the C4 C345C domain and immobilized MASP-2. An overlay of sensorgrams representing 20 nM (close to the calculated $K_D$) of wild-type C4 C345C and two mutants is shown. Dissociation constants ($K_D$) are 20.7, 22.1, and 29 nM for WT C345C, C345C R1724A, and C345C T1721A,R1724E, respectively.
Fig. S5. Putative exosites in C1s for C4 binding and discrimination against C4 by MASP-1. (A and B) Surface electrostatic potential of a C1s model at the putative CCP exosite (A; same orientation as Fig. 3E) and the SP domain (B; same orientation as Fig. 2D). (C and D) As for A and B, but displaying MASP-1. (E) Comparison of the SP domains of MASP-2 (Left; gray surface), C1s (Center; pink surface), and MASP-1 (Right; wheat-colored surface). C4 is shown as a cartoon with the MG3 domain in blue, the C4a domain in red, the Nt-α’ region in yellow, and the MG8 domain in green. C1s and MASP-1 were docked onto MASP-2 through their SP domains. Although it seems plausible that C1s can bind C4 similar to MASP-2, MASP-1 apparently would be prevented from proper R-loop recognition due to major steric hindrance (dotted rectangle) exerted by the C4 MG3 domain and minor steric hindrance (dotted ellipse) exerted by the C4 MG8 domain.
Fig. S6. Ordering and reorientation of C4a upon MASP-2 binding to C4. (A) The traceable parts of the C4a domain (red) and the Nt-α’ region (yellow) in 2mF₀ − DF, electron density contoured at 1σ from the structure of unbound C4. The dashed square indicates electron density for a symmetry-related molecule and can therefore not be attributed to the disordered R loop. (B) Same electron density but with C4a, R loop (red and yellow), and the Nt-α’ region from the C4·MASP-2 structure superimposed onto C4a from unbound C4, thereby for comparison showing where the untraceable regions—the α1-helix, the R loop, and part of the Nt-α’ region—are located in the C4·MASP-2 complex. (C) Comparison of the C4a (red), R loop (red-yellow), and Nt-α’ region (yellow) in C4·MASP-2 with the corresponding regions (gray) in unbound C4. Here, the structures of unbound C4 and C4·MASP-2 were superimposed on their α-chain superdomains to emphasize the relocation of C4a relative to α-chain superdomain upon MASP-2 binding.
Fig. S7. Comparison of C3, C4, and C5. (A) Cartoon representation of the three paralogues emphasizing the high similarity in their overall 3D architecture. The three proteins [unbound C3 (1) and C4, CVF-bound C5 (2)] were superimposed on their MG1 and MG5 domains to obtain a common orientation. The anaphylatoxin domains are colored red, and the C345C domain is brown. (B) SP view of the scissile bond region in unbound human C4, C3 (1), and C5 (3). Cleavage of the scissile bond causes formation of the two products, the anaphylatoxins (C3a, C4a, or C5a) and the nascent large fragment (C3b, C4b, or C5b). The dotted line represents the mobile region (residue numbers indicated) connecting the α4-helix with the Nt-α′ region. In C5 the P1 residue Arg-751 is ordered. The proteins were superimposed on the α2-, α3-, and α4-helices of their anaphylatoxins.

### Table S1. Statistics for data collection and refinement

<table>
<thead>
<tr>
<th></th>
<th>C4</th>
<th>C4 Ta$<em>2$Br$</em>{12}$</th>
<th>C4-MASP-2</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Data collection</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Beamline</td>
<td>ESRF ID29</td>
<td>SOLEIL Proxima 1</td>
<td>ESRF ID29</td>
</tr>
<tr>
<td>Space group</td>
<td>P2$_1$2$_1$</td>
<td>P2$_1$2$_1$2$_1$</td>
<td>P2$_1$</td>
</tr>
<tr>
<td>a, b, c, Å</td>
<td>85.5, 103.3, 256.0</td>
<td>87.0, 106.8, 258.1</td>
<td>99.0, 215.0, 142.8</td>
</tr>
<tr>
<td>Resolution, Å</td>
<td>45.89–3.60 (3.73–3.60)</td>
<td>44.82–6.00 (6.24–6.00)</td>
<td>48.98–3.75 (3.86–3.75)</td>
</tr>
<tr>
<td>$\beta$,°</td>
<td>110.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Refinement</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Resolution, Å</td>
<td>45.89–3.6 (3.73–3.60)</td>
<td>48.98–3.75 (3.86–3.75)</td>
<td></td>
</tr>
<tr>
<td>No. Reflections</td>
<td>25,600/1348</td>
<td>56,608/1727</td>
<td></td>
</tr>
<tr>
<td>$R_{\text{work}}/R_{\text{free}}$</td>
<td>20.0/26.9 (27.0/35.3)</td>
<td>17.4/24.2 (27.4/33.4)</td>
<td></td>
</tr>
<tr>
<td>No. of atoms</td>
<td>12,794</td>
<td>32,218</td>
<td></td>
</tr>
<tr>
<td>rms deviations</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bond lengths, Å/angles, °</td>
<td>0.005/1.117</td>
<td>0.006/1.126</td>
<td></td>
</tr>
</tbody>
</table>

*R$_{\text{sym}}$ = (Σ$_h$|I(h)| – <|I(h)|>/Σ$_h$_|I(h)|) for the intensity of reflection h measured N times. Values in brackets are for outer resolution shell. R factor = (Σ$_h$|F$_o$| – k|F$_c$|)/Σ$_h$_|F$_o$| where $F_o$ and $F_c$ are the observed and calculated structure factor respectively, and $k$ is a scaling factor. $R_{\text{free}}$ factor is identical to the $R$ factor on a subset of test reflections not used in refinement. One crystal was used for each dataset. Highest resolution shell is shown in parentheses.
Movie S1. Conformational changes in C4 occurring upon MASP-2 binding. C4 is displayed with the domains colored as in Fig. 18. To prepare the model of unbound C4 containing the same residues as C4 from the MASP-2 complex, which is required for interpolation between the two states, disordered residues in the C4a α1-helix, the R loop, and the sulfotyrosine region were modeled with O (1). These regions were placed merely to illustrate a conformational change upon MASP-2 binding, and their location is not built on experimental evidence. For the C4a α1-helix (residues 681–696) disordered in unbound C4, we fitted the CsA-desArg structure (PDB ID code 3HQA) onto C4a helices α2, α3, and α4. In this CsA structure, the α1 helix is ordered and exposed, but it does not interact with the remaining three helices. This location of the α1-helix is likely in the proforms of C3, C4, and C5 before cleavage between the β- and α-chain (2). We thereafter modeled the disordered C4a α1-helix after the exposed CsA α1-helix. A 30-state interpolation between the resulting modeled unbound C4 and C4 from the MASP-2 complex was conducted with the RIGIMOL option in PYMOL. The conformational changes shown here for the C4a α1 helix, the R loop, and the sulfotyrosine region are not built on experimental evidence, but help to illustrate a transition from a mobile conformation in unbound C4 (R-loop and sulfotyrosine region modeled) to an ordered conformation in MASP-2 bound C4 (R loop and sulfotyrosine region observed in our C4-MASP-2 structure). During the movie, C4 repeatedly shifts between the partially modeled unbound C4 conformation with the C4a α1-helix exposed, and the MASP-2 bound C4 conformation with the C4a α1-helix inserted between the MG3 domain (blue) and the rest of the C4a domain (red). Notice how the rotation of the C4 C345C domain (top) is transmitted through the MG7 and MG6 to the MG1-5 domains (bottom).

Movie S1