Structure and activation of C1, the complex initiating the classical pathway of the complement cascade


Edited by Douglas T. Fearon, Cornell University, Cambridge, United Kingdom, and approved December 22, 2016 (received for review October 13, 2016)

The complement system is an important antimicrobial and inflammation-generating component of the innate immune system. The classical pathway of complement is activated upon binding of the 774-kDa C1 complex, consisting of the recognition molecule C1q and the tetrameric protease complex C1r2s2, to a variety of activators presenting specific molecular patterns such as IgG and IgM-containing immune complexes. A canonical model entails a C1r2s2 with its serine protease domains tightly packed together in the center of C1 and an intricate intramolecular reaction mechanism for activation of C1r and C1s, induced upon C1 binding to the activator. Here, we show that the serine protease domains of C1r and C1s are located at the periphery of the C1r2s2 tetramer both when alone or within the nonactivated C1 complex. Our structural studies indicate that the C1 complex adopts a conformation incompatible with intramolecular activation of C1, suggesting instead that intermolecular proteolytic activation between neighboring C1 complexes bound to a complement-activating surface occurs. Our results rationalize how a multitude of structurally unrelated molecular patterns can activate C1 and suggests a conserved mechanism for complement activation through the classical and the related lectin pathway.

Significance

The complement system is an essential arm within the innate immune defense. Complement contributes to elimination of objects presenting danger signals such as pathogens, dying host cells, and abnormal molecular structures and is capable of inducing an inflammatory response stimulating further immune responses. The C1 complex is a giant proteolytic enzyme, which plays a leading role, because it is the first component in a proteolytic cascade initiated when complement is activated. On the basis of structural characterization of the C1 complex with X-rays and electron microscopy, we suggest that the first proteolytic reaction in the cascade, activation of the C1 complex itself, involves neighboring C1 complexes located near each other rather than a reaction within individual C1 complexes.


The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Data deposition: The data reported in this paper have been deposited in the Small Angle Scattering Biological Data Bank (SASDB), https://www.sasdb.org (accession nos. SASDB27, SASDB28, and SASDB38).

1To whom correspondence should be addressed. Email: gra@mbg.au.dk.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1616998114/-/DCSupplemental.
models for C1 activation state that activator-binding leads to a conformational rearrangement in C1q collagen stems that causes structural reorganization of the associated C1r2s2, bringing the two C1r SP domains close enough to activate each other (9, 10). Upon C1r activation and subsequent activation of C1s within the same tetramer, the C1s SP domains are suggested to become accessible to the substrates, C4 and C4b2.

A similar intramolecular activation mechanism was until recently believed to govern activation within the LP. However, biochemical evidence indicated that activation of the LP occurs through intermolecular proteolytic cleavage between SP domains from mannan-binding lectin (MBL) dimers situated on different PRMs. The role of an activating glycan pattern is therefore to create a local high concentration of the PRM-MASP complexes and to orientate the MASPs correctly relative to each other (11). Support for this model was obtained through small-angle X-ray scattering (SAXS) and EM studies of MBL in complex with a nonactivated MASP-1 dimer (12). We set out to investigate whether the structure of the C1 complex favors intercomplex proteolytic activation or indeed is compatible with intracomplex activation of C1 as suggested by the prevalent model.

Results

The C1r2s2 Tetramer Is Highly Extended. To gain structural insight into the solution state of C1r2s2, we collected synchrotron SAXS data from extensively purified recombinant C1r2s2 (Fig. 2A) with the S/A mutation in both proteases making the tetramer an inactive enzyme. Guinier analysis showed a linear plot for C1r2s2 in the concentration range of 1.0–6.8 mg/mL, but with slightly steady increasing values of $R_g$. We extrapolated the data to 0 mg/mL and created a merged dataset from the extrapolated data and the data recorded at 6.8 mg/mL (Fig. 2E and F and Fig. S1A). For the subsequent steps, we used merged data with $q < 2.5$ nm$^{-1}$ [$q = (4 \pi \times \sin(\theta)) / \lambda$, where $2\theta$ is the scattering angle]. The Kratky plot of the data showed that C1r2s2 is well folded and has limited flexibility (Fig. S1B). The pair-distance distribution function ($P(r)$) revealed that $D_{\text{max}}$ for the particle is just below 45 nm (Fig. 2F), which was immediate evidence against a compact conformation of the tetramer. To obtain a SAXS-based solution structure of the C1r2s2 tetramer, we used rigid-body modeling starting from tetramer models based on known substructures of C1r and C1s (Fig. S1C). The core of the tetramer models was formed by the CUB1-EGF domain from two C1s molecules related by a twofold rotation axis each surrounded by the CUB1-EGF domains from C1r (Fig. 3A and B). The best results in terms of fit to the experimental data and internal consistency between the resulting models were obtained by dividing both C1r and C1s into four rigid bodies consisting of CUB1-EGF, CUB2, CCP1-CCP2, and the SP domain. Models from 50 rigid-body refinements were sorted into two groups according to $\chi^2$ values of their fits to the data. The major group of 43 models fitted the SAXS data with $\chi^2 = 1.78 \pm 0.05$, whereas the remaining seven models had $\chi^2$ of 2.60 $\pm$ 0.15. The output models in the major group were rather similar with the CCP1-CCP2-SP domains curving to either side of the plane compared with the quite planar starting model (Fig. 3B and C). Removal of the glycans from the input model resulted in 10 models with $\chi^2 = 3.56 \pm 0.45$, showing that the glycans contributed significantly with respect to matching the models to the data although the 12 Asn-linked glycans constitute less than 10% of the total molecular weight of the tetramer. Importantly, the models obtained in the absence of glycans were similar to those obtained with glycans included. The model with the lowest $\chi^2 = 1.68$, which is representative, is displayed in Fig. 3D and it is fit to the data in Fig. 2E. The SP domains from both C1r and C1s are protruding from the central CUB1-EGF tetrameric interface. The catalytic sites of the two C1r SP domains are separated by 39 nm, whereas those of C1s are 28 nm apart from each other. The activation site in one C1r is separated by 37 nm from the active site in the second C1r molecule.

We do not know how the C1r CUB1-EGF domain tandem interacts with their C1s domains equivalent in the core of the tetramer,
and C1r,C1s may have a different structure in solution than as a component of the C1 complex. We therefore also investigated two alternative tetramer models in which C1r had been translated 27 Å and 54 Å (Fig. S1 D and E) in the plane defined by the C1s CUB1-EGF dimer compared with the first starting model where the C1r-C1s CUB1-EGF arrangement was based on a C1s CUB1-EGF tetramer formed by crystal packing (3). Rigid-body refinements gave average χ² values of 1.68 ± 0.04 and 1.34 ± 0.16 for 27 and 54 Å C1r translation, respectively, for 10 CORAL runs in each case. Hence, in terms of fitting the experimental data, both C1r-translated models were better than models derived from the nontranslated starting model. However, the resulting ensembles of models were structurally more diverse, perhaps reflecting that the C1r hinge points were located further out at the periphery of the complex. Lower shows computed projection images of the SAXS model from D using the same scaling as for the class averages. For each class average, a similar projection image has been assigned manually. (Scale bars: D, 5 nm; E, 50 nm.)

To elucidate the structural arrangement of the domains within C1, we performed single-particle electron microscopy on C1r,C1s. We recorded negative stain images of the C1r,C1s complex at low voltage conditions (60 kV) to achieve optimum contrast. Furthermore, we selected specimen regions with one single carbon layer appearance to avoid flattening effects from additional carbon films covering the particles that would influence the apparent size. We computed class averages of the particle dataset with on average 34 images per class by using unbiased clustering techniques independent of any SAXS-based model (Fig. 3E, Upper). Clustering enhances the signal-to-noise ratio and serves as a statistical tool that brings out recurrent particle views. All class averages show particles with an elongated shape and a mean maximum dimension of approximately 43 nm in agreement with a Dmax of 45 nm observed by SAXS. The particles show a central broadening and small peripheral domains protruding from the complex.

To compare the EM data to the SAXS model, we generated a density map from a representative atomic SAXS-derived model by using the same sampling as for the EM images, and computed projection images from this map to illustrate how the SAXS structure would appear in EM under different angles of view. For comparison between EM and SAXS, we manually assigned a computed projection to each class average (Fig. 3E, Lower). The comparison shows that the SAXS-derived model exhibits the same shape and overall dimensions as the particles imaged by EM. Overall, the SAXS and EM models appear in line with each other up to the imposed twofold symmetry in the SAXS model, i.e., some EM class averages suggest a deviation from a perfect twofold symmetry in particular in the outer protuberances.

C1 is a Hollow Particle with the SP Domains Pointing Away from the Core. We next examined the solution structure of nonactivated C1 with SAXS. C1 was assembled by mixing Clq purified from plasma with purified recombinant C1r2(S/A)2(S/A) tetramer (Fig. 2 A and B) in which both C1r and C1s were inactive because the serine in the catalytic triade was mutated to alanine. We used sucrose gradient centrifugation to separate the reconstituted C1r2S2 (Fig. 2C). The SAXS data were recorded in line immediately following elution of the C1 from a size exclusion chromatography (SEC) column. This approach allowed us to separate a small amount of oligomeric C1 from monomeric C1. We selected scattering curves from the center of the elution peak for merging to optimize the signal-noise ratio and to minimize the possible contribution from unbound Clq and C1r(C1r(S/A))2(S/A) eluting after the C1 complex. Furthermore, a stable value for the radius of gyration Rg (Fig. S2A) argued against significant C1 dissociation in the late eluting part of the peak, and interestingly, the Rg value of C1 was smaller than those observed for C1r and Clq (Figs. S2A and S3 and Table S1). The Guinier plot of the C1 data were linear (Fig. S2B), and for subsequent data analysis, we used data with q < 2.5 nm⁻¹. The Kratky plot of the data suggested that C1 is a multidomain protein with several well-folded domains with some flexibility (Fig. S2C). Ab initio modeling produced a model featuring a hollow shape similar to a saddle, which is incompatible with a densely packed core in which the C1r and C1s protease domains are tightly packed in the center of C1 (Fig. S2D).

To elucidate the structural arrangement of the domains within C1, we performed rigid-body modeling analyzing a central twofold axis running in parallel with the N-terminal segment of the Clq collagen stem. We started from the same tetramer model as for C1r,C1s except that the SP domain formed a single rigid body together with the CCP domains for both C1r and C1s, and glycans were omitted to reduce computational complexity. The collagen stems of Clq were restrained to the CUB domains of C1r,C1s (Fig. 4A). Common to all of the resulting models derived by rigid-body refinement was a hollow core as also observed in the ab-initio reconstruction, with the protease domains of C1r and C1s protruding away from the core. Fifty models were divided into two groups according to their fit to the experimental data. The first group comprised 44 models with χ² = 2.38 ± 0.050, whereas the remaining 6 models had χ² = 2.6 ± 0.082. In the major group, there was a clear twist of the tetramer compared with the planar starting model of the C1r,C1s tetramer (Fig. 4 B and C). Generally, the protease domains stretch out to form a longer and slimmer particle perpendicular to the twofold symmetry axis compared with the starting model. In a few models, the protease domains of C1s and C1r are pointing upwards toward the N-terminal collagen stem of Clq (Fig. S2E). In 4 of the 44 models, the protease domains occupy the same general area as the globular heads of Clq. In these models, there is a shift upwards of two or more of the globular heads to compensate for the SP domains of C1r,C1s such

---

**Fig. 3.** SAXS rigid-body modeling and negative stain EM suggest a highly elongated structure of C1r>C1s. (A) Schematic representation of the domains within the C1r>C1s. (B) The C1r>C1s tetramer starting model for rigid-body refinement. (C) The 43 rigid-body-derived structures are all curved but have a roughly equal probability of curving in opposite directions relative to the central plane. (D) The rigid-body structure with the best fit to the data. (E) Negative stain EM of the C1r>C1s tetramer. Upper shows eight selected class averages of C1r>C1s. The complex exhibits an elongated shape with a central broadening and small globular domains sticking out at the periphery of the complex. Lower shows computed projection images of the SAXS model from D using the same scaling as for the class averages. For each class average, a similar projection image has been assigned manually. (Scale bars: D, 5 nm; E, 50 nm.)
that the mass distribution between gC1q and SP domains is roughly conserved.

The minor group of output models is much more heterogeneous, but generally, the same features are observed. We observe the same orientations of the SP domains, where they can either be in the central plane, pointing upwards, but never packed inside the central void delimited by the collagen stems, the globular heads of C1q, and the CUB1-EGF domains of C1r2s2. To confirm that the input model did not bias our results, we conducted another 50 rigid-body refinements where the C1r CCP domains and the SP domain were folded into the core of C1 similar to the prevalent C1r zymogen model (10), with only minor adjustments due to geometrical constraints (Fig. S4A). In all of the resulting rigid-body output models, the C1r CCP1-CCP2-SP domains had swung out from the core to point away from the core as seen with the first input model (Fig. S4B).

We also recorded negative stain images of C1 molecules under identical conditions as for C1r2s2. Again, we selected specimen regions with a typical single carbon layer appearance to avoid flattening effects. We clustered the resulting dataset into an average 28 images per class by using unbiased clustering techniques independent of any SAXS model. The resulting class averages show maximum dimensions of 32–36 nm (Fig. 4D), which is slightly lower than the Dmax derived from SAXS. In the predominant views, 8–10 globular domains are arranged around a central network of densities. The central portions of the negatively stained class averages appear rather dark, which may be explained by an accumulation of heavy metal ion stain in this region. This particular staining behavior is thus in agreement with a rather hollow structure of the particle as suggested by both SAXS ab-initio and rigid-body modeling.

When comparing the EM class averages to a representative SAXS model of the C1 complex, the SAXS model predicts up to 10 globular domains that protrude from the stalk of C1 in top/bottom views of C1. Furthermore, the SAXS model predicts side views, where the C1q hub is seen as one of the protruding domains (compare Fig. 4C, Left, with e.g., upper protrusion labeled “h” in Fig. 4D). An EM class average with 10 individual globular protrusions is shown in row 3, view 3 of Fig. 4D. However, the majority of class averages show eight to nine discernable peripheral domains. For interpretation of views with less than 10 protruding domains, it is relevant to note that in certain projection views, some domains may be superimposed and, thus, masked in the class averages.

We interpreted class averages, where all protruding domains appeared as small roundish densities as top/bottom views (Fig. 4D), which implies that the central densities may represent overlays of the C1q hub with central portions of C1r2s2. The frequent occurrence of top/bottom views may be explained by a preferred binding of the C1q globular domains to the carbon film. The distances between the globular domains appear to be somewhat variable, indicating a certain degree of conformational variability. Such variability is not reflected in the SAXS models where twofold symmetry was imposed to minimize the number of rigid bodies to reduce overfitting and computational complexity.

Although the level of detail does not allow distinguishing C1q head domains and C1r2s2 SP domains, the class averages we obtained for C1 in negative stain strongly argued against a compact conformation of C1r2s2 within the complex. Such a compact conformation would give rise to projection views showing up to six protruding domains dependent on the angle of view and degree of overlay (2). Moreover, opposing globular domains were typically separated by a projected distance of at least 29 nm, which likewise argues in favor of a rather long distance between C1r2s2 SP domains in the C1 complex.

To further investigate the indicated C1 model, we also performed native cryo-EM (Fig. 4E, columns 1 and 3). A particular challenge was the low contrast of the C1 complex under cryo conditions, which can be explained by the extreme sparsity of protein density in this complex, as a mass of only 774 kDa is distributed in a sphere of 38 nm diameter. For comparison, the C1 complex is more than 50% larger than a prokaryotic ribosome by size while containing only ∼1/3 of its mass. We used highly diluted samples of C1 to minimize aggregation, resulting in approximately 1–10 particles per exposure. The cryo class averages (Fig. 4E) comprise 65 images per class. As cryo images are taken at a higher defocus compared with negatively stained particles, there is a somewhat higher amount of blurring, which, in turn, means that individual globular domains may appear superimposed. Nevertheless, we could distinguish six to nine protruding domains in most class averages (Fig. 4E, columns 1 and 3). Moreover, the angular distribution of views appeared to be more uniform than in the negative stain data, leading to a mean diameter of ∼38 nm in the cryo class averages. As for the C1r2s2 complex, we converted a representative C1 SAXS model into an identically scaled density map and computed projections under different angles.
of view (Fig. 4E, columns 2 and 4). Again, a computed projection was manually assigned to each class average. In this comparison, it can be seen that there is a good overall agreement in the outlines of the class averages and assigned SAXS projections. However, the exact position of the individual globular domains may vary between the EM class averages and the single SAXS model used here for projection. These discrepancies may originate from structural heterogeneity of C1, and the individual flexibility of both the C1q and C1r2s2 components may lead to a considerable deviation from the twofold symmetry assumed for the SAXS rigid-body modeling. Such a conformational flexibility is in good agreement with the variations found between the different SAXS models that we computed as described above.

Discussion

The structural data presented here contradict the generally accepted structure of the C1 complex where the tetrameric C1r2s2 is folded inside the C1q stalks. Autoactivation occurring between two C1r molecules embedded in a single C1 molecule appears impossible because of the separation of their SP domains we observe for nonactivated C1 both in solution and in electron microscopy. We also observe that the unbound C1r2s2 is highly extended, further arguing against a compact conformation of C1r2s2 once bound inside C1. The simplest solution to the activation conundrum appears to be that zymogen C1r in one C1 complex is activated by C1r in a neighboring C1 complex. Our study does not allow us to exclude that C1r activates a neighboring zymogen C1s in the same C1 molecule. At both ends of the protease tetramer, the SP domains of one C1r and one C1s will be present near each other. However, the peripheral location of the active site at the far end of the C1r SP domain appears better compatible with activation of C1s in a neighboring C1 molecule compared with activation of the C1s belonging to the same C1 molecule. Sophisticated kinetic studies using several variants of C1r and C1s and a model system in which exchange of C1r and C1s between tetramers is under stringent control are likely to be needed to settle this question.

With the first step of C1 activation being an intercomplex reaction, the major role for the activator is that it brings C1 molecules together and orients the SP domains in neighboring C1 molecules roughly in plane, allowing inter-C1 activation. This mechanism would be akin to the one we showed for the lectin pathway (11). This suggestion is also in agreement with the finding that a gC1q-specific mAb and its F(ab′)2 fragment are able to activate C1 in the fluid phase through aggregation of C1q (13). Further supporting that tethering of C1 in the correct orientation at a certain density is sufficient for CP activation, bispecific antibody fragments, diabodies of scFv fragments specific for gC1q and lysozyme, were able to target C1 to lysozyme-coated erythrocytes and induce their lysis (14). Both results appear to be compatible with the intercomplex activation model rather than with intracomplex activation.

The steadily growing list of natural activators of the CP is also a strong point in favor of an intercomplex activation mechanism. C1q binding is mainly governed by electrostatic interaction between gC1q and the more than 100 established ligands (15). Not all are able to activate complement, but structurally unrelated ligands like IgG/IgM, C-reactive protein, gC1qR, and the Alzheimer’s-associated oligomers of Ab, suggest to mention a few, activate the classical pathway (16). The prevalent intercomplex activation model is rather intricate mechanistically and is difficult to reconcile with such a large number of diverse activators.

The prevalent model for the C1 complex is based on older negative stain EM studies of C1 and C1r2s2, C1r–C1r interactions observed in crystal packing (2, 9, 10) and mass-spectrometry analysis (5). This model suggests that the CCP1-CCP2-SP domains of C1r and C1s pack tightly in the center of C1. Activation of C1 is also suggested to be an intramolecular reaction elicited by conformational changes propagating through the C1q collagen stem upon activator binding. Besides our structural studies presented here, there are additional arguments against the prevalent model: First, to our knowledge, there is no direct experimental data showing that activator binding induces a conformational change in the C1q collagen stems. Second, crystal packing interactions between two C1r CCP1-CCP2-SP fragments in which the SP domain of one molecule contacts the CCP1 domain in the second have been suggested to mirror C1r-C1r interactions in nonactivated C1 (2, 9, 10). However, the in vivo importance of the crystal contact has never been supported by mutations in C1r. Third, the putative compact conformation of C1 underlying the prevalent model has never been unambiguously observed by others (2). Fourth, activation of C1q-bound C1r requires C1s CUB1-EGF-CUB2 domains, but not the C1s CCP1-CCP2-SP domains (17, 18). This result is odd if the C1s domains and their C1r equivalents were tightly packed in the center of the C1 complex but it is readily compatible with C1r and C1s SP domains extending away from the center. Fifth, the putative compact C1 conformation would also require the C1r CUB2 domain to adopt a severely bent or partially unfolded conformation to simultaneously engage in the interaction with a C1q collagen stem and allow the CCP1-CCP2-SP domains to be positioned in the core of C1. Sixth, mass spectrometry quantitation of lysolecithin accessibility in C1r2s2 as a free tetramer compared with being incorporated in C1 suggested that several lysines in the C1s SP domain are protected upon incorporation into C1, which was taken as evidence in support of the compact C1 model underlying intracomplex molecular activation (5). However, this difference could also be due to insertion of the tetramer into C1, which most likely will change the dynamic properties of C1s and its SP domain and apparently also leads to a more compact organization of the tetramer as indicated by the lower D_max value for C1 compared with the free tetramer.

Our results do not exclude a conformational change in C1 upon activator binding. Such a change could actually change both the accessibility and conformation of the C1r2s2 SP domains and promote intercomplex activation. The crucial difference between the prevalent model and the model we present is the conformation of the fluid-phase nonactivated C1. Our structural studies were conducted with two complementary techniques with one of them, SAXS, directly addressing the fluid phase conformation. These studies indicate that C1r2s2 both unbound and present in nonactivated C1 is extended rather than compact with respect to the SP domain. The recent cryo-EM tomography studies of C1 bound to a hexameric IgG platform is likely to represent the activated state of C1. Here, the C1r2s2 CCP1-CCP2-SP domains are also proposed to protrude from the center of the C1 molecule (19).

An obvious strategy for distinguishing between intermolecular and intramolecular C1 activation is to investigate the kinetics of the reaction. The wild-type C1r2s2 tetramer can slowly activate spontaneously, but the activation is accelerated by formation of the C1 complex and fast activation occurs upon binding to IgG containing immune complexes (17, 20). The kinetics for C1 activation is extremely complicated because it involves binding of C1 to an activator, autoactivation of C1r, and subsequent activation of C1s by C1r possibly accompanied by conformational changes in the intricate C1 complex. A simplistic model is to assume that the conversion of nonactivated C1 to activated C1 has a single overall rate-limiting step. One study in favor of intramolecular C1 activation supports first-order kinetics for spontaneous activation of C1 in solution in the absence of an activator with a t_1/2 of 4 and 7 min at 37 and 30 °C, respectively (21). Here, C1 was reconstituted from subunits derived from serum. However, another study displayed no measurable fluid-phase activation of reconstituted C1 after 20 min unless significant amounts of IgG oligomers or monomers were present (22). A third study using antigen-antibody aggregates as activator also showed little activation of reconstituted C1 in the absence of immune complexes (17). A fourth using C1 purified directly from fresh plasma showed only slow activation in the absence of immune complex (23). Later it was suggested that small
amounts of contaminating active proteases (including activated C1 and C1r) could contribute to an apparent first-order activation of C1 (24, 25). Overall, existing kinetic data on C1 activation cannot distinguish between the two activation models.

In conclusion, our structural studies indicate that the first step in C1 activation involves cleavage of zymogen C1r in one C1 complex by C1r from a neighboring C1 complex while further data are required to decide whether C1s activation by C1r also occurs between C1 complexes. Our results lead us to propose a universal model for activation occurring through the related lectin and classical pathway of complement. Elucidation of structure and function are important for future design of therapeutic intervention strategies aiming at stimulating or inhibiting complement activation. Complement-dependent cytotoxicity inducing antibodies used in cancer therapy rely on their Fc segment being organized in oligomers to which C1 can bind and augmentation of C1 targeting to cancer cells is a proven strategy (26). In combination with the above-mentioned existing experimental data showing that tethering of C1 to a target is sufficient to induce CP activation, our results provide support for IgG-independent targeting of C1 to cancer cells or pathogens as a viable strategy.

**Methods**

C1q was purified from plasma whereas recombinant C1r2s2 both carrying the S/A mutations in both proteases was expressed in HEK293-F cells. The reconstituted C1 complex was purified by ultracentrifugation in a sucrose gradient. SAXS data for C1r2s2 and C1q were collected in batch mode at ESRF BM29, whereas the Superose 6 SEC inline SAXS data for the C1 complex were collected at Petra P12. Starting models for rigid-body refinement of C1r2s2 and C1 were constructed from available substructures available at the Research Collaboratory for Structural Bioinformatics protein data bank. For preparation of negative-stain EM samples for C1r2s2, a peak fraction from a SEC Superdex 200 column was used, whereas for C1, the complex was first purified by sucrose gradient centrifugation and then by SEC on Superose 6 column. Grids were imaged in a Tecnai T12 microscope. For cryo-EM analysis, C1 was purified by sucrose gradient and SEC, diluted and adsorbed on a continuous carbon film. After freezing in liquid ethane, samples were imaged in a Titan Krios EM (FEI) equipped with a Gatan US4000 camera operated at 200 kV. Full experimental details are presented in SI Methods.

**ACKNOWLEDGMENTS.** We thank the beamline staff at Petra P12 and ESRF BM29 for support during data collection and, in particular, Cy Jeffries for help with in-line data reduction and Maxim Petoukhov for compiling CORALXL. G.R.A. was supported by Biorstruct-X, Dancsatt, The Danish Council for Independent Research for Natural Sciences, and the Lundbeck Foundation centre BRAINSTRUC. S.T. was supported by the Danish Council for Independent Research for Medical Sciences and the Novo-Nordic Foundation. B.S. acknowledges funding from the Danish Council for Independent Research for Natural Sciences. M.M.G. received funding from the Sapere Aude Program of the Danish Council for Independent Research and a Lundbeck Foundation Fellowship.

15. Thielens NM, Illy C, Bally IM, Arlaujl GI (1994) Activation of human complement serine-protease C1r is down-regulated by a Ca(2+) dependent intramolecular control that is released in the C1 complex through a signal transmitted by C1q. Biochem J 301(Pt 2):509–516.
Supporting Information

Mortensen et al. 10.1073/pnas.1616998114

SI Methods

Expression and Purification of Proteins. The genes according to the sequence for C1r (NM_001733.4) and C1s (NM_201442.2) were synthesized by GenScript Inc. USA. The synthesized genes were cloned into the vector pcDNA3.1/myc-His(-) A (Invitrogen) using the restriction sites EcoRI/XbaI. Inactive C1r and C1s were generated by mutating the active serine of the catalytic triad in C1r(S637A) and C1s(S632A). The human embryonic kidney cell line HEK293-F (Invitrogen) was used for transient expression of recombinant inactive versions of C1r and C1s. HEK293-F cells were grown in protein-free medium (FreeStyle 293 Expression Medium; Gibco) with agitation at 37 °C and 8% CO2. Transfection was done by using PEI (25 kDa; Polysciences) as transfection reagent in a PEELDNA ratio 3:1. After transfection, the cells were grown for 4–5 d, and the supernatants were harvested after centrifugation.

C1r and C1s Purification for SAXS Studies. The tetramer C1r(S637A)C1s(S632A) was purified according to Bally et al. (7), whereas C1q was purified from serum according to Tenner et al. (27). To reconstitute C1, C1q was mixed with a molar excess C1r(S637A) and loaded on a 10–30% (wt/vol) sucrose gradient in 50 mM EPPS (4-(2-hydroxyethyl)-1-piperazinepropanesulfonic acid) (pH 8.5), 145 mM NaCl, 3 mM CaCl2, and centrifuged in a TH-660 rotor (Sorvall, ThermoFisher Scientific) at 185,795 × g for 17 h at 4 °C. After ultracentrifugation, fractions were collected from the bottom of the centrifuge tube by using a peristaltic pump and analyzed by SDS-PAGE. Before in-line SAXS analysis of C1, it was dialyzed against 50 mM EPPS (pH 8.5), 145 mM NaCl, 3 mM CaCl2, and concentrated immediately before SAXS measurements. C1r(S637A)C1s(S632A) and C1q were purified like C1 on a 10–30% (wt/vol) sucrose gradient in 50 mM Tris-HCl (pH 7.4), 145 mM NaCl, 3 mM CaCl2. Before SAXS data collection, the samples were dialyzed against 50 mM Tris-HCl (pH 7.4), 145 mM NaCl, 3 mM CaCl2, and concentrated immediately before SAXS measurements.

SAXS Data Collection and Processing. The data for C1r(S637A)C1s(S632A) and C1q were collected in batch mode (ranging from 1.0 to 6.8 mg/mL) at the ESRF BM29 beamline by using a PILATUS 1M pixel detector and λ = 0.992 Å at a temperature-controlled capillary at 4 °C. The sample-to-detector distance was 2,867 m, covering a range of momentum transfer 0.004 < q < 0.5 Å−1[q = (4π×sinθ)/λ, where 2θ is the scattering angle]. For the C1 complex, in-line SAXS size exclusion chromatography was done at the PETRA III/EMBL P12 beamline by using a Superose 6 10/30 GL column, equilibrated in 50 mM EPPS (pH 8.5), 145 mM NaCl, 3 mM CaCl2. The flow rate was 0.3 mL/min. The data were recorded at 20 °C by using a PILATUS 2M pixel detector (DECTRIS) and λ = 1.240 Å. The sample-to-detector distance was 3.0 m covering 0.002 < q < 0.48 Å−1. Each frame covers a 0.995-s exposure performed every second during the SEC run. Normalization and radial averaging was performed at the beamline by using the automated pipeline (28, 29). Data reduction to produce the scattering profile were performed by following standard procedures (30). The background and C1 data were obtained from frames 3800–3900 and frames 1851–1950, respectively.

SAXS Ab Initio Modeling and Rigid-Body Refinement. In the C1r(S637A)C1s(S632A) tetramer, two C1s molecules are located centrally with one C1r molecule on the outside of each C1s molecule (2, 3). The input model for CORAL refinement of the C1r(S637A)C1s(S632A) tetramer was constructed in three steps (Fig. S1C). First, the CCP1-CCP2-SP fragments were constructed by the combination of structures of zymogen C1r (pdb ID codes 1GPZ and 1MD7), whereas for zymogen C1s, ID code 4J1Y was used. Second, a CUB1-EGF tetramer was constructed based on a crystal packing tetramer in the structure of the C1s CUB1-EGF-CUB2 fragment (ID code 4LMF) suggested to represent a possible model for the C1r(S637A)C1s(S632A) tetramer. The C1r CUB1 and CUB2 domains were modeled by using phenix.sculptor (31) starting from the structure of the C1s CUB1-EGF-CUB2 fragment (ID code 4LOR). These models were then combined with the structure of the C1r EGF domain (ID code 1APQ). In the ID code 4LMF, the CUB2 domain is bent out of the plane of the CUB1-EGF tetramer. We modeled an approximate in-plane orientation of both the C1r and C1s CUB2 domains by comparison with the structure of MAP-1 (ID code 4AOB), which is better compatible with simultaneous binding of collagen stems to both CUB1 and CUB2 domains. In the third step, the C1r(S637A)C1s(S632A) tetramer model was constructed by combining of the CUB1-EGF-CUB2 tetramer with the CCP1-CCP2-SP fragments of C1r and C1s using the structure of the C1s CUB2-CCP1-CCP2 fragment (ID code 4LOT) as a guide. Finally, smaller missing regions in this tetramer model were manually modeled within the program “O” (32), and two Asn-linked glycans on C1s and four on C1r were modeled as described (33). However, because the spatial relation between the C1r CUB1-EGF domains and the corresponding domains from C1s is unknown, we also constructed two alternative start models with 27 and 54 Å in-plane translations of C1r relative to the fixed C1s (Fig. S1 D and E).

The backbone and β-carbons of the C1q collagen stems were modeled by using the interactive Triple-Helical collagen Bul lign Script (THe BuScr) (34). The C1q collagen stem was modeled in two separate pieces with one fragment consisting of the amino acids before kink, and the other consisting of the amino acids after the kink. The chains were modeled with the glycine of chain A hydrogen bonding to the next amino acid in the tripeptide unit in chain C, the glycine of chain C hydrogen bonding to chain B, and the glycine of chain B hydrogen bonding to chain A as described in ref. 4. An all-atom model was obtained from the backbone and β-carbon model by using a side chain rotamer library (35) followed by the addition of the hydroxylprolines with THe BuScr. In the final model of the collagen stem of C1q, the collagen stem after the kink was added as two separate pieces to allow for better solvation in CORALXL. The kink was not modeled in details, instead the two parts of C1q was held together by two distance restraints to allow flexibility in the angle between the two stems. Finally, the complete C1 molecule was assembled by placing lysine residues in the C1q collagen stems known to be important for interaction close to the calcium sites in CUB1 and CUB2 domains in C1r and the C1s CUB1 domain (7) to establish a starting geometry for interaction between the collagen stem and CUB domains as observed in ID codes 4LOR (C1s CUB1-collagen) and 3POB (MASP-1 CUB2-collagen). Only the C1r(S637A)C1s(S632A) tetramer model with a C1r translation of 0 Å was used as subsequent placement of the collagen stems onto the CUB domains in the two models with C1r translations was found to require a tight clustering of the collagen stems at opposite ends of the CUB1-EGF-CUB2 tetramer.

The rigid-body refinements were done by using data range in the range s < 0.25 Å−1 with CORALXL, a custom-made version of CORAL (36) allowing 30 rigid bodies and 1,000 distance restraints compiled by Maxim Petoukhov at the EMBL Hamburg. A twofold symmetry operation was applied by using P2 symmetry to the input models to minimize the number of rigid bodies. Distance restraints were used to maintain the interface between symmetry-related C1s and C1r molecules.
CUB1-EGF domains across the twofold rotation axis. For C1\(_{2}\), three distance restraints were applied, whereas 28 restraints were used for C1. Several values of these restraints (connectivity and distance) were tested before the final refinements and evaluated based on internal consistency between the resulting models and biological relevance. A representative output model and the data used for rigid-body refinement are deposited at the SASDB for both C1\(_{2}\) and the C1 complex (Table S1). Plots were prepared with GraphPad Prism 5.03. The Porod volume was determined by in-house software written by J.S.P. A Guinier expression was fitted to the low-q part of the data to determine \(I(0)\), and the Porod law \(I(q) \propto q^{-4}\) with an added constant was fitted to the high-q part to determine the constant to be subtracted from the data so that they follow the Porod Law. The Porod invariant was calculated by numerically integrating the experimental data multiplied by \(q^2\) with appropriate extrapolations to \(q = 0\) and \(q = \infty\).

**EM.** For negative-stain EM of the C1\(_{2}\) complex, a peak fraction from size exclusion chromatography on a 24-mL Superdex 200 column (GE Healthcare) equilibrated in 10 mM Hepes (pH 7.4), 150 mM NaCl, and 3 mM CaCl\(_2\) was diluted appropriately. Carbon film prepared on mica was floated on freshly prepared sample for 2 min, and grids were prepared by using the sandwich method (37). To prepare C1 samples for negative-stain EM and cryo-EM, the C1 complex was subjected to sucrose gradient centrifugation as described above for preparation of the SAXS sample. Fractions containing the C1 complex (Fig. 2C) were subjected to SEC on a Superose 6 10/300 GL column, equilibrated in 50 mM EPPS (pH 8.5), 145 mM NaCl, 3 mM CaCl\(_2\), and the peak fraction was prepared according to the sandwich method for the negative-stain EM. Grids were imaged in a Tecnai T12 microscope (FEI/Thermo Fisher Scientific) at room temperature at a high tension of 60 kV on a Gatan Multiscan 794 CCD camera (Gatan) at a nominal magnification of 42,000x resulting in a pixel size of 4.2 Å per pixel. For C1\(_2\), a defocus of −1.5 to −2.1 μm was used. A total of 7,551 particles were manually selected from the images, corrected for defocus (38) and clustered as described (12) into on average 34 images per class. For the C1 complex, a defocus of −1.2 to −1.4 μm was used, and 14,331 particles were selected and clustered into on average 28 images per class.

For cryo-EM analysis of C1 material from SEC was also used. A continuous carbon film was mounted on Quantifoil (Jena, Germany) 3.5/1 grids. Appropriately diluted sample was adsorbed on a glow-discharged grid for 1 min, followed by manual blotting and freeze-plunging into liquid ethane by using a Leica EM CPC system (Leica Microsystems). The sample was imaged in a Titan Krios EM at 200 kV (FEI), a defocus of −5 to −8 μm and 59,000x nominal magnification on a Gatan US4000 camera resulting in a final pixel size of 1.2 Å per pixel. Because the C1qrs complex generates a rather low contrast under cryogenic conditions, a low-pass Wiener filter was applied to the images, and 12,890 images were selected manually. Images were clustered into on average 65 images per class.
Fig. S1. SAXS analysis of C1r\(_{52}\). (A) The Guinier plot of the merged C1r\(_{52}\) data. The residual plot is shown at Right. (B) Kratky plot of the data suggests C1r\(_{52}\) to be a rigid and highly folded molecule. (C) Schematic representation in the same style as Fig. 3A outlining the process of generating the C1r\(_{52}\) starting model. PDB entries used for the various steps as described in SI Methods are indicated. (D) Alternative input model in which the C1r molecule is translated 27 Å horizontally compared with the input model displayed in Fig. 3B. The output models from 10 rigid-body refinements are superimposed at the bottom. (E) As in C, but with a translation of 54 Å.
Fig. S2. SAXS analysis of the C1 complex. (A) The forward scattering and the $R_g$ of the inline SAXS data for C1 is plotted against the elution volume. The gray area marks the volume corresponding to the frames used for further processing. (B) The linear Guinier plot calculated from the SAXS inline data displays no signs of C1 aggregation or intermolecular repulsion. The residual plot is shown at Right. (C) Kratky plot of the data suggests that C1 is mostly rigid but with some internal flexibility. (D) An averaged model of C1 based on 100 ab initio models calculated with DAMMIN. (E) An example of a rigid-body output model in which the CCP1-CCP2-SP fragments are pointing upwards compared with the central plane of the complex.
Fig. S3. SAXS analysis of C1q. (A) Scattering data of C1q presented as $I(q)$ on a logarithmic scale plotted against $q$. (B) The Guinier plot calculated from the SAXS data. The residual plot is shown at Right. (C) Pair-distance distribution function for C1q indicating a particle with $D_{\text{max}}$ approaching 45 nm. (D) Kratky plot of C1q suggests that C1q is mostly rigid but with some flexibility within the molecule.
Fig. S4. Rigid-body refinement of C1 starting with the C1r CCP1-CCP2-SP moiety placed inside the central void. (A) Input model for refinement with the C1r CCP1-CCP2-SP moieties placed inside the central void. (B) Representative output model with all four SP domains located at the periphery of C1r located in the central plane of C1, showing that the C1r conformation in the input model did not bias the output models significantly during rigid-body refinement.

Table S1. Key statistics and software used for SAXS data collection, processing, analysis, and rigid-body refinement

<table>
<thead>
<tr>
<th>SASBDP entry</th>
<th>SASDBZ7</th>
<th>SASDB28</th>
<th>SASDB838</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data-collection parameters</td>
<td>C1r (batch)</td>
<td>C1q (batch)</td>
<td>C1 (inline)</td>
</tr>
<tr>
<td>Instrument</td>
<td>ESERF BM29</td>
<td>ESERF BM29</td>
<td>Petra P12</td>
</tr>
<tr>
<td>Wavelength, Å</td>
<td>0.992</td>
<td>0.992</td>
<td>1.24</td>
</tr>
<tr>
<td>q range, Å⁻¹</td>
<td>4.6×10⁻³ to 0.50 (0.25)*</td>
<td>3.6×10⁻³ to 0.50</td>
<td>2.2×10⁻³ to 0.48 (0.25)*</td>
</tr>
<tr>
<td>Exposure time per frame, s</td>
<td>2</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>No. of frames averaged</td>
<td>10</td>
<td>10</td>
<td>100</td>
</tr>
<tr>
<td>Concentration range, mg·ml⁻¹</td>
<td>1.01–6.79</td>
<td>3.66</td>
<td>N/A</td>
</tr>
<tr>
<td>Temperature, K</td>
<td>277.15</td>
<td>277.15</td>
<td>293.15</td>
</tr>
<tr>
<td>Structural parameters</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I(0) [from P(r)]</td>
<td>436</td>
<td>534</td>
<td>N/A</td>
</tr>
<tr>
<td>Rg, Å [from P(r)]</td>
<td>130</td>
<td>126.6</td>
<td>114.6</td>
</tr>
<tr>
<td>I(0) (from Guinier)</td>
<td>436 for 1.8 mg·ml⁻¹</td>
<td>535</td>
<td>N/A</td>
</tr>
<tr>
<td>Rg, Å (from Guinier)</td>
<td>126 ± 3.0</td>
<td>126 ± 2</td>
<td>115 ± 1</td>
</tr>
<tr>
<td>Dmax, Å</td>
<td>450</td>
<td>450</td>
<td>365</td>
</tr>
<tr>
<td>Porod volume estimate, Å³</td>
<td>645×10³</td>
<td>965×10³</td>
<td>2,050×10³</td>
</tr>
<tr>
<td>Dry volume calculated from sequence and glycans¹, Å³</td>
<td>399×10³</td>
<td>563×10³</td>
<td>992×10³</td>
</tr>
<tr>
<td>Molecular mass M, [from I(0)]², kDa</td>
<td>332</td>
<td>407</td>
<td>N/A</td>
</tr>
<tr>
<td>Calculated monomeric M, from sequence, O-linked glycans, N-linked glycans, hydroxylations, kDa</td>
<td>330</td>
<td>444</td>
<td>774</td>
</tr>
</tbody>
</table>

Software used

| | PyFAI/EDNA | PyFAI/EDNA | SASFLOW |
| | N/A | N/A | DAMMIN |
| Ab initio analysis | N/A | N/A | DAMAVER Suite |
| Validation and averaging | N/A | N/A | CORALXL |
| Rigid-body modeling | N/A | N/A | CORALXL |
| Computation of model intensities | | | |

*The maximum q range used for rigid-body modeling is indicated in the parentheses.

¹The volume was calculated as M, times 1.21 Å³/Da for the protein component. A volume of 2,535 Å³ was used for an Asn-linked glycan and 361 Å³ for the O-linked Glc-Gal disaccharide on hydroxylated lysine.

²Using BSA with I(0) = 94.52 and Mw = 72 kDa as a standard.