**The Crystal Structure of the Globular Head of Complement Protein C1q Provides a Basis for Its Versatile Recognition Properties**

Christine Gaboriaud, Jordi Juanhuix, Arnaud Gruez, Monique Lacroix, Claudine Darnault, David Pignol, Denis Verger, Juan C. Fontecilla-Camps, and Gérard J. Arlaud

From the Laboratoire de Cristallographie et Cristallogénèse des Protéines and the Laboratoire d’Enzymologie Moléculaire, Institut de Biologie Structurale Jean-Pierre Ebel, Commissariat à l’Énergie Atomique (CEA-CNRS-Université Joseph Fourier, Grenoble Cedex 1, France, Laboratoire de Llum Sincrotró, Ed. Cn, Campus UAB 08193 Bellaterra, Barcelona, Spain, and Département d’Ecophysiologie Végétale et de Microbiologie, CEA-Cadarache, 13108 Saint Paul les Durance, France

C1q is a versatile recognition protein that binds to an amazing variety of immune and non-immune ligands and triggers activation of the classical pathway of complement. The crystal structure of the C1q globular domain responsible for its recognition properties has now been solved and refined to 1.9 Å of resolution. The structure reveals a compact, almost spherical heterotrimeric assembly held together mainly by non-polar interactions, with a Ca2+ ion bound at the top. The heterotrimeric assembly of the C1q globular domain appears to be a key factor of the versatile recognition properties of this protein. Plausible three-dimensional models of the C1q globular domain in complex with two of its physiological ligands, C-reactive protein and IgG, are proposed, highlighting two of the possible recognition modes of C1q. The C1q/human IgG1 model suggests a critical role for the hinge region of IgG and for the relative orientation of its Fab domain in C1q binding.

Innate immunity involves a combination of cell-surface receptors and soluble proteins with the ability to recognize microbial pathogens and thereby to generate signals that both orientate subsequent adaptive immune responses and trigger effector mechanisms (1, 2). Most of these molecules are oligomeric and recognize molecular patterns on microorganisms (3). An archetypal molecule of this type is C1q, the recognition subunit of C1, the complex that triggers activation of the classical pathway of complement, a major element of innate immunity. C1q is a 460-kDa protein with the overall shape of a bouquet of flowers, comprising six heterotrimeric collagen-like triple helices that associate in their N-terminal half to form a “stalk,” then diverge to form individual “stems”, each terminating in a C-terminal heterotrimeric globular domain (4). It is well documented that most of the C1 complex ligands are recognized by these peripheral globular domains, or heads, of C1q, thus triggering activation of C1r and C1s, the proteases associated with C1q (5). It is also established that C1q binds to immune complexes containing IgG or IgM, but not to those having IgA, IgD, or IgE (6). The major C1q binding site on IgG has been mapped to the CH2 domain of the Fc portion of the molecule (7–9). Although C1q shows marked differences in its reactivity toward IgG subclasses, the reason for this selectivity is not known.

C1q is traditionally known for its ability to bind antibodies. However, it recognizes an amazing variety of other ligands. These include certain bacteria, viruses, parasites, and mycoplasma (6, 10–12), underscoring its role as an antibody-independent defense protein. C1q also binds to C-reactive protein (CRP)1 when complexed with exposed phosphocholine residues on bacteria, providing a further means of host defense (13). C1q is also capable of recognizing aberrant structures from self. Thus, in addition to cellular debris and sub-cellular membranes (14), it is established that C1q binds to, and induces clearance of, apoptotic cells (15), thereby playing a major role in immune tolerance. Recent studies also indicate that abnormal proteins such as β-amyloid fibrils (16, 17) and the prion protein (18, 19) are recognized by C1q. There are no obvious structural features shared by these ligands, but the fact that many polymers are C1q ligands (6) suggests that C1q may function as a charge pattern recognition molecule.

The globular domain of C1q is a heterotrimeric association of protein modules known as gC1q domains found at the C-terminal end of various proteins, including types VIII and X collagens, the adipocyte complement-related protein (ACRP)-30, precrebellin and, multimerin (4). The structures of the globular domains of ACRP-30 (20) and collagen X (21) have been solved by x-ray crystallography, revealing the gC1q fold and indicating that both are homotrimers held together by both hydrophobic and polar interfaces. We report here the x-ray structure of the heterotrimeric globular head of C1q, the domain responsible for the versatile recognition properties of this protein. The structure reveals how three different gC1q modules achieve an assembly homologous to, but structurally more diverse than, the one observed in homotrimers and provides insights into the molecular mechanisms of the recognition function of C1q.

**EXPERIMENTAL PROCEDURES**

Preparation and Analysis of the C1q Globular Domain—C1q was purified from human serum as described previously (22) and digested with Achromobacter iophagus collagenase (Roche Applied Science) (enzyme/protein ratio = 0.2, w/w) for 24 h at 37 °C in 250 mM NaCl, 5 mM

---

1 The abbreviations used are: CRP, C-reactive protein; Mes, 4-morpholinethanesulfonic acid; r.m.s.d., root mean square distance; ACRP-30, adipocyte complement-related protein-30.

---

Received for publication, July 18, 2003, and in revised form, August 29, 2003 Published, JBC Papers in Press, September 5, 2003, DOI 10.1074/jbc.M307764200
that the crystal had a triclinic P1 space group (24).

Laser desorption ionization technique on a Voyager Elite XL instrument. Mass spectrometry analysis was performed using the matrix-assisted laser desorption technique in the presence of 0.2–0.4% agarose, 2% glycerol, and 100 mM non-detergent Sulfobetaine 195 as a solubilizing agent (25). Crystals were obtained at 20°C, with satisfactory overall statistics (R_{free} = 0.04; 92% completeness) but with only 60% completeness of the anomalous signal.

The initial C1q-CRP model was built manually by positioning residues Lys92, Tyr175, and Arg155, from the C1q head directed toward CRP residues Asp29, Tyr175, and Arg132, respectively. This C1q model was then subjected to molecular dynamics and refined using CNS (34). The collagen-like arms were oriented in a way that the 36–39-insertion segment of the A chain was restrained to orientations where the collagen arms had no steric clashes with either the targets or the underlying surface. To take into account these various constraints, the two structures were initially positioned in a modified docking program Hex (38, 39) was used to search for solutions in a more exhaustive and objective way. Hex calculates interaction energies that include a hydrophobic excluded volume model derived from the notion of overlapping surface skins with or without soft electrostatic potential complementarity. Bad contacts at the main-chain level are checked by the program. Because most of the solutions obtained from a free rotation of the C1q globular domain were not compatible with the above selection criteria, the search was restricted to the ligand orbit protocol.

Interaction with IgG b12 was constrained by the location of the C1q binding site in human IgG1 as defined by mutagenesis data (9, 41, 42).

Interaction between the ACRP-30 and collagen X structures were superimposed on this minimal model to help tracing and building the complete C1q globular head structure. The automated refinement procedure (30) was used to improve the quality of the maps and to reduce the model bias. Model building was easily carried out to achieve very clear electron density maps using program O (31). Refinement was done using CNS (29) using the very last steps, which were performed using REFMAC (32) after the introduction of water molecules and of alternative conformations for some amino acids in the model. The following residues have a disordered side chain and have been modeled as Ala: Gin410, Arg632, Gin635, Thr636, Gin638, Arg639, Arg640, Arg642, Arg643, Gin645, Lys675. Atomic coordinates have been deposited in the Protein Data Bank with accession code PDB 6G16.

Structure Modeling—The C1q collagen-like stem model was based on published statistical information derived from collagen-like structures (33). The relative positioning of the collagen triplets of the A, B, and C chains in the triple helix is the only one compatible with the N-terminal end of the present globular domain structure. The conformational parameters for amino-rich regions (33) were used for the modeling of segments A40–A48, A55–A60, and A78–A87 and of the corresponding segments B42–B50, B57–B62, B80–B89, C39–C47, C54–C59, and C77–C86.

The conformational parameters for imino-rich regions were used for the other segments. The globular domains were equally spaced in a circle with a radius of ~100 Å (34). The collagen-like arms were oriented in a way that the two platinum sites were first positioned in the center of the ACRP structure (20) was used as a model to solve the platinum binding site in human IgG1. These solutions are under severe restricting steric constraints. The solutions obtained from a free rotation of the C1q globular domain were not compatible with the above selection criteria, the search was restricted to the ligand orbit protocol.

The initial C1q-CRP model was built manually by positioning residues Lys92, Tyr175, and Arg155, from the C1q head directed toward CRP residues Asp29, Tyr175, and Arg132, respectively. This C1q model was then subjected to molecular dynamics and refined using CNS (34). The collagen-like arms were oriented in a way that the 36–39-insertion segment of the A chain was restrained to orientations where the collagen arms had no steric clashes with either the targets or the underlying surface. To take into account these various constraints, the two structures were initially positioned in a modified docking program Hex (38, 39) was used to search for solutions in a more exhaustive and objective way. Hex calculates interaction energies that include a hydrophobic excluded volume model derived from the notion of overlapping surface skins with or without soft electrostatic potential complementarity. Bad contacts at the main-chain level are checked by the program. Because most of the solutions obtained from a free rotation of the C1q globular domain were not compatible with the above selection criteria, the search was restricted to the ligand orbit protocol.

Interaction between the ACRP-30 and collagen X structures were superimposed on this minimal model to help tracing and building the complete C1q globular head structure. The automated refinement procedure (30) was used to improve the quality of the maps and to reduce the model bias. Model building was easily carried out to achieve very clear electron density maps using program O (31). Refinement was done using CNS (29) using the very last steps, which were performed using REFMAC (32) after the introduction of water molecules and of alternative conformations for some amino acids in the model. The following residues have a disordered side chain and have been modeled as Ala: Gin410, Arg632, Gin635, Thr636, Gin638, Arg639, Arg640, Arg642, Arg643, Gin645, Lys675. Atomic coordinates have been deposited in the Protein Data Bank with accession code PDB 6G16.

Structure Modeling—The C1q collagen-like stem model was based on published statistical information derived from collagen-like structures (33). The relative positioning of the collagen triplets of the A, B, and C chains in the triple helix is the only one compatible with the N-terminal end of the present globular domain structure. The conformational parameters for amino-rich regions (33) were used for the modeling of segments A40–A48, A55–A60, and A78–A87 and of the corresponding segments B42–B50, B57–B62, B80–B89, C39–C47, C54–C59, and C77–C86.

The conformational parameters for imino-rich regions were used for the other segments. The globular domains were equally spaced in a circle with a radius of ~100 Å (34). The collagen-like arms were oriented in a way that the two platinum sites were first positioned in the center of the ACRP structure (20) was used as a model to solve the platinum binding site in human IgG1. These solutions are under severe restricting steric constraints. The solutions obtained from a free rotation of the C1q globular domain were not compatible with the above selection criteria, the search was restricted to the ligand orbit protocol.

Interaction between the ACRP-30 and collagen X structures were superimposed on this minimal model to help tracing and building the complete C1q globular head structure. The automated refinement procedure (30) was used to improve the quality of the maps and to reduce the model bias. Model building was easily carried out to achieve very clear electron density maps using program O (31). Refinement was done using CNS (29) using the very last steps, which were performed using REFMAC (32) after the introduction of water molecules and of alternative conformations for some amino acids in the model. The following residues have a disordered side chain and have been modeled as Ala: Gin410, Arg632, Gin635, Thr636, Gin638, Arg639, Arg640, Arg642, Arg643, Gin645, Lys675. Atomic coordinates have been deposited in the Protein Data Bank with accession code PDB 6G16.

The conformational parameters for imino-rich regions were used for the other segments. The globular domains were equally spaced in a circle with a radius of ~100 Å (34). The collagen-like arms were oriented in a way that the two platinum sites were first positioned in the center of the ACRP structure (20) was used as a model to solve the platinum binding site in human IgG1. These solutions are under severe restricting steric constraints. The solutions obtained from a free rotation of the C1q globular domain were not compatible with the above selection criteria, the search was restricted to the ligand orbit protocol.

The initial C1q-CRP model was built manually by positioning residues Lys92, Tyr175, and Arg155, from the C1q head directed toward CRP residues Asp29, Tyr175, and Arg132, respectively. This C1q model was then subjected to molecular dynamics and refined using CNS (34). The collagen-like arms were oriented in a way that the 36–39-insertion segment of the A chain was restrained to orientations where the collagen arms had no steric clashes with either the targets or the underlying surface. To take into account these various constraints, the two structures were initially positioned in a modified docking program Hex (38, 39) was used to search for solutions in a more exhaustive and objective way. Hex calculates interaction energies that include a hydrophobic excluded volume model derived from the notion of overlapping surface skins with or without soft electrostatic potential complementarity. Bad contacts at the main-chain level are checked by the program. Because most of the solutions obtained from a free rotation of the C1q globular domain were not compatible with the above selection criteria, the search was restricted to the ligand orbit protocol.

The initial C1q-CRP model was built manually by positioning residues Lys92, Tyr175, and Arg155, from the C1q head directed toward CRP residues Asp29, Tyr175, and Arg132, respectively. This C1q model was then subjected to molecular dynamics and refined using CNS (34). The collagen-like arms were oriented in a way that the 36–39-insertion segment of the A chain was restrained to orientations where the collagen arms had no steric clashes with either the targets or the underlying surface. To take into account these various constraints, the two structures were initially positioned in a modified docking program Hex (38, 39) was used to search for solutions in a more exhaustive and objective way. Hex calculates interaction energies that include a hydrophobic excluded volume model derived from the notion of overlapping surface skins with or without soft electrostatic potential complementarity. Bad contacts at the main-chain level are checked by the program. Because most of the solutions obtained from a free rotation of the C1q globular domain were not compatible with the above selection criteria, the search was restricted to the ligand orbit protocol.
respectively, and are therefore unlikely to be available for protein-protein interaction. Other studies based on expression of the individual A, B, and C modules of C1q indicate that, although both modules A and B show significant binding to IgG, only the latter has marked binding selectivity for IgG relative to IgM (45–47). Taken together the above information led us to the working hypothesis that most of the C1q residues involved in IgG recognition are contributed by module B. The initial model was built manually by positioning the IgG residues Asp270 and Lys326 facing C1q residues ArgB114 and GluB162, respectively. A cluster of similar models (within 3-Å r.m.s.d. from each other) was obtained with the docking program, and no other alternative solution was found that meets the selection criteria mentioned above. In the proposed model(s), additional ionic interactions possibly form between C1q residues ArgB114 and AspB163 and IgG residues GluM195 and Glu287, and several hydrophobic residues of C1q (IleB103, ValB105) and IgG (LeuM154, ProM204) show a decreased access to the solvent in the assembly. Other studies based on expression of the individual C1q residues ArgB114 and ArgB161 and IgG residues GluM195 and Glu287, and several hydrophobic residues of C1q (Leu114, Pro204) show a decreased access to the solvent in the assembly. The IgG residues GluM33 and LyM206 restrict the access of C1q GluB162 and ArgB163, respectively, and improved values are obtained when these two residues are converted to Ala in the computation (Table II). Although these computed differences are exaggerated because the two models are arbitrarily kept rigid to simplify the modeling process, they are coherent with the observed effects of the corresponding mutations (see “Discussion”).

**RESULTS**

**Overall Structure**—The C-terminal globular domain of C1q was obtained after digestion of the collagenous part of the protein with collagenase, treated with periodate, and purified by ion-exchange chromatography as described under “Experimental Procedures.” N-terminal sequence analysis of the purified material after separation of the three chains by SDS-PAGE yielded the following sequences: Gly-Asn-Ile-Lys-Asp-Glu (A chain), Gly-Asp-(OH)GLys-Glu-Glu-Ser (B chain), and Gly-Glu-Pro-Gly-Glu-Glu (C chain). Analysis by mass spectrometry yielded three major peaks with mass values of 17,339 ± 20 Da (A chain), 16,812 ± 20 Da (B chain), and 15,600 ± 20 Da (C chain). Both analyses were consistent with each other and indicated that the purified material comprised residues Gly185, Ala223 of the A chain, Gly81-Ala226 of the B chain, and Gly286, Asp217 of the C chain.

The crystal structure of the C1q globular domain was solved by molecular replacement and refined to 1.9-Å resolution. The final Rwork and Rfree factors are 0.193 and 0.238, respectively, and the refined model has excellent stereochemistry (Table I). Residues Glu120 to Ser222, Thr392 to Asp222, and Lys389 to Asp217 show clear and continuous electron densities, with only a few disordered side chains (see “Experimental Procedures”). The N-terminal collagen-like triplets not digested by collagenase are absent from the electron density map and, consequently, were not included in the model. The structure reveals a tight heterotrimeric assembly with non-crystallographic pseudo-3-fold symmetry, the subunits arranged clockwise in the order A, B, C when viewed from the top (Fig. 1A). The assembly exhibits a globular, almost spherical structure with a diameter of about 50 Å (Fig. 1B). As observed in the case of the ACRP-30 and collagen X homotrimers (20, 21), the N and C termini of the three subunits emerge at the base of the trimer. A further feature reminiscent of the collagen X structure (21) is the presence of a Ca2+ ion bound to the apical side of the trimer (Fig. 1B).

The subunit structure shows a 10-stranded β sandwich with a jellyroll topology homologous to the one described initially for tumor necrosis factor (48, 49), consisting of two five-stranded β-sheets (A', A, H, C, F) and (B', B, G, D, E), each made of anti-parallel strands (Fig. 2A). Compared with each other the C1q subunits show r.m.s.d. values of 0.73–0.94 Å, based on their overall structures, and of only 0.56–0.71 Å, based on the β-strands. These comparisons indicate strong conservation of the latter and significant variability in the loops, particularly A-A' and G-H on the apical side (Fig. 2A). Compared with ACRP-30 and collagen X, the β-sheets of the C1q modules show r.m.s.d. values of 0.59–0.70 Å, indicating strong structural homology within the C1q family. The free cysteines homologous to those found in ACRP-30 and collagen X (Fig. 2B) are essentially buried in the structure, consistent with the fact that they are not alkylated, despite treatment of the protein with iodoacetamide (see “Experimental Procedures”). A specific feature of the C1q modules is that they contain two extra cysteines that form a disulfide bond (Cys150-Cys168 in module A) (Fig. 2A). These cysteines are conserved in the sequences of the C1q chains from other species, except for Cys168, which is missing in the mouse C1q A chain (50). In agreement with previous studies (51, 52), a carbohydrate chain is attached to Asn124 in the B’-B loop, as shown by the observed partial electron density corresponding to the proximal two N-acetylglucosamine residues.

**The Heterotrimeric Assembly**—The β-sheet formed by the B’, B, G, D, and E strands and loop A-A’ form the exterior of the trimer, whereas the second β-sheet (strands A’, A, H, C, F) in the second strand of the C1q trimer, whereas the second strand of the C1q module shows r.m.s.d. values of 0.59–0.70 Å, indicating strong structural homology within the C1q family. The free cysteines homologous to those found in ACRP-30 and collagen X (Fig. 2B) are essentially buried in the structure, consistent with the fact that they are not alkylated, despite treatment of the protein with iodoacetamide (see “Experimental Procedures”). A specific feature of the C1q modules is that they contain two extra cysteines that form a disulfide bond (Cys150-Cys168 in module A) (Fig. 2A). These cysteines are conserved in the sequences of the C1q chains from other species, except for Cys168, which is missing in the mouse C1q A chain (50). In agreement with previous studies (51, 52), a carbohydrate chain is attached to Asn124 in the B’-B loop, as shown by the observed partial electron density corresponding to the proximal two N-acetylglucosamine residues.

**Experimental Procedures**—The crystal structure of the C1q globular domain was solved by molecular replacement and refined to 1.9-Å resolution. The final Rwork and Rfree factors are 0.193 and 0.238, respectively, and the refined model has excellent stereochemistry (Table I). Residues Glu120 to Ser222, Thr392 to Asp222, and Lys389 to Asp217 show clear and continuous electron densities, with only a few disordered side chains (see “Experimental Procedures”). The N-terminal collagen-like triplets not digested by collagenase are absent from the electron density map and, consequently, were not included in the model. The structure reveals a tight heterotrimeric assembly with non-crystallographic pseudo-3-fold symmetry, the subunits arranged clockwise in the order A, B, C when viewed from the top (Fig. 1A). The assembly exhibits a globular, almost spherical structure with a diameter of about 50 Å (Fig. 1B). As observed in the case of the ACRP-30 and collagen X homotrimers (20, 21), the N and C termini of the three subunits emerge at the base of the trimer. A further feature reminiscent of the collagen X structure (21) is the presence of a Ca2+ ion bound to the apical side of the trimer (Fig. 1B).

The subunit structure shows a 10-stranded β sandwich with a jellyroll topology homologous to the one described initially for tumor necrosis factor (48, 49), consisting of two five-stranded β-sheets (A’, A, H, C, F) and (B’, B, G, D, E), each made of anti-parallel strands (Fig. 2A). Compared with each other the C1q subunits show r.m.s.d. values of 0.73–0.94 Å, based on their overall structures, and of only 0.56–0.71 Å, based on the β-strands. These comparisons indicate strong conservation of the latter and significant variability in the loops, particularly A-A’ and G-H on the apical side (Fig. 2A). Compared with ACRP-30 and collagen X, the β-sheets of the C1q modules show r.m.s.d. values of 0.59–0.70 Å, indicating strong structural homology within the C1q family. The free cysteines homologous to those found in ACRP-30 and collagen X (Fig. 2B) are essentially buried in the structure, consistent with the fact that they are not alkylated, despite treatment of the protein with iodoacetamide (see “Experimental Procedures”). A specific feature of the C1q modules is that they contain two extra cysteines that form a disulfide bond (Cys150-Cys168 in module A) (Fig. 2A). These cysteines are conserved in the sequences of the C1q chains from other species, except for Cys168, which is missing in the mouse C1q A chain (50). In agreement with previous studies (51, 52), a carbohydrate chain is attached to Asn124 in the B’-B loop, as shown by the observed partial electron density corresponding to the proximal two N-acetylglucosamine residues.

| Table II: Scores for suitable C1q-ligand complex models obtained using the docking program Hex |
|-----------------|---------|---------|---------|-----------------|--------|
| C1q ligand (PDB code) | E total* | E shapeb | E forcec | r.m.s.d. (Å)c | Subunitd |
| CRP (1B09) | -613 | -415.6 | -197.4 | 1.97 | A |
| -579 | -389.0 | -190.0 | 3.9 | C |
| IgG1 b12 (1HZH) | -557.7 | -397.6 | -160.1 | 11.9 | E |
| -165.7 | -189.1 | 20.4 | 0.9 | |
| b12-mutated model | -155.5 | -115.3 | -40.2 | 2.5 | |
| -245.8 | -219.9 | -25.9 | 1.2 | |

* Total energy (eV).
b Scored shape complementarity.
c Scored electrostatic interactions.
d Root mean square distance from the initial manually built model.

e CRP subunit interacting with C1q Ty3175.
C. Whereas the Ca^{2+} cluster observed in collagen X is buried (21), the single Ca^{2+} ion of C1q is well exposed to the solvent and defines the upper entrance of a discontinuous central channel. (iii) Main-chain polar interactions take place between residues of loops E-F before the channel closes at the level of Ser^{A180}, Thr^{B182}, and Ser^{C176}. (iv) In the central part of the assembly a pocket is formed containing several water molecules bridging residues mainly located on strands C and F through a network of hydrogen bonds. (v) Tyr^{A138}, Tyr^{B142}, and Tyr^{C139} are found at the boundary between the upper, hydrophilic and the lower, hydrophobic parts of the central interface, the latter mainly involving residues from the C-terminal half of strand H. In addition to this central interface, a number of lateral contacts take place between the three modules. Again,
these interactions are hydrophobic near the base and become more polar toward the top, with a number of hydrogen bonds and two intermodular salt bridges, GluB209-LysC160 and AspA211-LysB166.

The heterotrimeric C1q assembly has many characteristics in common with the collagen X homotrimer (21), notably the different nature of the upper and lower halves of the structure. However, comparison of the overall β-sheet structures in the C1q and collagen X trimers shows a r.m.s.d of 0.78 Å, indicating subtle differences in the relative positioning of the three modules in the two structures. Tentative modeling of homotrimeric C1q assemblies clearly results in a loss of shape complementarity and in destabilization of the lateral salt bridges. Amino acid sequence comparisons (Fig. 2B) reveal that residues conserved in the gC1q modules of C1q, ACRP-30, and collagen X either belong to the hydrophobic core or participate in hydrophobic interactions in the lower half of the trimers, but they indicate substantial variability in the residues found at the interfaces.

Surface Properties—The three C1q modules exhibit marked differences in their electrostatic surface potentials. Thus, the pseudo 3-fold symmetry seen at the scaffold level disappears when the charge distribution at the C1q surface is considered. Module A (Fig. 3A) shows a combination of arginine and acidic residues scattered on its external face. The top part of the module is also markedly charged, with a predominance of lysine residues. The surface of module B (Fig. 3B) shows a net predominance of positive charges with, in particular, a continuous patch of three basic residues, ArgB110, ArgB114, and ArgB129, flanked by AspB116 and GluB127. Residues in italics have a disordered side chain and have been modeled as Ala. The free cysteines and disulfide bonds of subunits A, B, and C are superimposed, and only the cysteines of subunit C are displayed for clarity. B, structure-based sequence alignment of the gC1q domains of C1q, ACRP-30, and collagen X. The residue numbering (A, B, C, from top to bottom) and the secondary structure elements shown are those of the C1q chains. Residues in italics have a disordered side chain and have been modeled as Ala. C-terminal residues AlaA223, MetB224, and AlaB225 have no matching electron density. Residues involved in the C1q interfaces are highlighted in orange. Ca2+ binding ligands in C1q and collagen X are in red. Asterisks (*) indicate residues conserved in C1q, collagen X, and ACRP-30. The free Cys residues are indicated by SH, and the carbohydrate attachment site in subunit A of C1q is indicated by CHO.

Several hydrophobic residues are exposed on the external side of each subunit, with most of them located in concave areas of the structure. Notable exceptions to this include ProA103, MetA104, ProA119 (Fig. 3A), AlaC105, ProC106, and ValC185 (Fig. 3C) as well as a hydrophobic cluster (IleB101, ValB105, ProB106) extending over the ArgB110, ArgB114, and ArgB129 triad in the upper part of module B (Fig. 3B). Another hydrophobic patch, ValB118, IleB119, IleB129, lies in between the charged cluster of module B and the B-C interface. Only module C displays solvent-accessible aromatic residues (TyrC155, TrpC190) on its equatorial side (Fig. 3C).}

**Surface Properties**—The three C1q modules exhibit marked differences in their electrostatic surface potentials. Thus, the pseudo 3-fold symmetry seen at the scaffold level disappears when the charge distribution at the C1q surface is considered. Module A (Fig. 3A) shows a combination of arginine and acidic residues scattered on its external face. The top part of the module is also markedly charged, with a predominance of lysine residues. The surface of module B (Fig. 3B) shows a net predominance of positive charges with, in particular, a continuous patch of three basic residues, ArgB110, ArgB114, and ArgB129, flanked by AspB116 and GluB127. Residues in italics have a disordered side chain and have been modeled as Ala. The free cysteines and disulfide bonds of subunits A, B, and C are superimposed, and only the cysteines of subunit C are displayed for clarity. B, structure-based sequence alignment of the gC1q domains of C1q, ACRP-30, and collagen X. The residue numbering (A, B, C, from top to bottom) and the secondary structure elements shown are those of the C1q chains. Residues in italics have a disordered side chain and have been modeled as Ala. C-terminal residues AlaA223, MetB224, and AlaB225 have no matching electron density. Residues involved in the C1q interfaces are highlighted in orange. Ca2+ binding ligands in C1q and collagen X are in red. Asterisks (*) indicate residues conserved in C1q, collagen X, and ACRP-30. The free Cys residues are indicated by SH, and the carbohydrate attachment site in subunit A of C1q is indicated by CHO.

**Surface Properties**—The three C1q modules exhibit marked differences in their electrostatic surface potentials. Thus, the pseudo 3-fold symmetry seen at the scaffold level disappears when the charge distribution at the C1q surface is considered. Module A (Fig. 3A) shows a combination of arginine and acidic residues scattered on its external face. The top part of the module is also markedly charged, with a predominance of lysine residues. The surface of module B (Fig. 3B) shows a net predominance of positive charges with, in particular, a continuous patch of three basic residues, ArgB110, ArgB114, and ArgB129, flanked by AspB116 and GluB127. Residues in italics have a disordered side chain and have been modeled as Ala. The free cysteines and disulfide bonds of subunits A, B, and C are superimposed, and only the cysteines of subunit C are displayed for clarity. B, structure-based sequence alignment of the gC1q domains of C1q, ACRP-30, and collagen X. The residue numbering (A, B, C, from top to bottom) and the secondary structure elements shown are those of the C1q chains. Residues in italics have a disordered side chain and have been modeled as Ala. C-terminal residues AlaA223, MetB224, and AlaB225 have no matching electron density. Residues involved in the C1q interfaces are highlighted in orange. Ca2+ binding ligands in C1q and collagen X are in red. Asterisks (*) indicate residues conserved in C1q, collagen X, and ACRP-30. The free Cys residues are indicated by SH, and the carbohydrate attachment site in subunit A of C1q is indicated by CHO.
ProC103, ValC200, IleC202) in module C. Several aromatic residues are also accessible to the solvent, notably TrpA147, TyrB175, PheB178, TyrC196, and TyrC197.

**DISCUSSION**

The three new gC1q module x-ray structures described here and those determined previously for ACRP-30 and collagen X (20, 21) confirm the generality of the topology identified initially from these latter structures. Remarkably, the trimeric globular domains assemble in similar ways in C1q, ACRP-30, and collagen X despite the fact that interactions are achieved by three identical modules in the latter two proteins but by three different modules in C1q. In this respect, tentative assembly in silico of C1q homotrimers reveals in all cases a number of severe steric hindrances, especially at the level of lateral contacts, for example, in the A-A-A pseudo homotrimer, where MetA183 and LeuA165 clash with IleA115 and IleA98, respectively. This provides a structural basis for the specific property of the C1q subunits to associate only as heterotrimers (4). Interestingly, the highest degree of conservation between C1q and its homotrimeric homologues among interface residues is seen at the level of the hydrophobic interfaces near the base of the trimers, suggesting a critical role for this region in the alignment of the three chains in the globular and collagen regions on either sides. In this respect, the C1q structure provides information that is essential to model the collagen-like triple helix on the N-terminal side of the globular domain (see "Experimental Procedures"). The overall similarity between C1q and collagen X along the interface also suggests a common assembly process driven by hydrophobic interactions near the base and locked by hydrophilic contacts toward the top. Interestingly, the total buried surface in C1q (5490 Å²) is similar to

---

**Fig. 3. Surface properties of the C1q globular domain.** A, side view of the heterotrimer seen from module A. B, side view seen from module B. C, side view seen from module C. The overall orientation of the trimer is similar to that in Fig. 1B. D, view of the heterotrimer seen from the top. The side chains of Arg, Lys, His, Asp, and Glu residues are shown in *deep blue*, *light blue*, *green*, *red*, and *magenta*, respectively. Hydrophobic residues (Ile, Leu, Val, Pro, Met) are shown in *yellow*, and aromatic residues (Phe, Trp, Tyr) are in *orange*. The lines in D indicate the approximate module boundaries. This figure was made using GRASP (59).

**Fig. 4. Model of the complex between CRP and the C1q globular domain.** A, side view. Subunits B and C of CRP have been omitted for clarity. B, perpendicular bottom view. Color coding for the C1q subunits is as in Fig. 1. The lysines at the top of the C1q head (A173, A200, A201, C170) and TyrB175 are in *light blue*. A–E designate the CRP protomers as described by Shrive et al. (55). The phosphocholine (PC) ligand is in *red*, and the nearby Ca²⁺ ion is in *green*. Color coding for CRP mutations (40) is as follows. Mutations impairing complement activation (Glu-88, Asp-112, Tyr-175) are *magenta*, and mutations enhancing complement activation (Lys-114) are *blue*.
that of ACRP-30 (5320 Å²) but considerably less than that of collagen X (7360 Å²), confirming that the exceptional stability of the collagen X NC1 trimer arises from a particularly compact structure compared with other C1q-like proteins (21).

A further interesting feature of the C1q structure is the presence of a Ca²⁺ binding site at the top of the assembly. This finding is consistent with the fact that crystallization was performed in the presence of Ca²⁺ ions and accounts for previous data (53), indicating that Ca²⁺ binding is an intrinsic property of the C1q molecule. Whether this site only contributes to the stability of the heterotrimeric assembly or has also a functional role remains to be determined. In this respect it is noteworthy that Ca²⁺ is coordinated by six ligands, among which is a single carboxyl group. Although these characteristics are shared by many of the known protein Ca²⁺ binding sites (54), the observation that the Ca²⁺ ion is fully accessible to the solvent (Fig. 1C) opens the possibility that some of the charged targets recognized by C1q may interact directly with the Ca²⁺ ion by displacing one or both of the water molecules (Fig. 1C). It is also noteworthy that the residues involved in Ca²⁺ binding in C1q are homologous to their counterparts in collagen X. Because the latter define a consensus sequence that is also found in ACRP-30, it appears likely, therefore, that the partially disordered structure observed at the top of ACRP-30 (20) arises from the absence of Ca²⁺ in the structure.

The heterotrimeric structure of the C1q globular head has direct implications in generating the versatile recognition properties of this protein. Each of the three subunits exhibits particular surface patterns in terms of charged and hydrophobic residues and may, therefore, be expected to display specific individual recognition properties. In addition, the compact trimeric structure of the C1q head clearly allows ligand recognition through residues contributed by two or even three subunits, thereby broadening the recognition spectrum of C1q. This may explain in part the observation that many of the C1q ligands exhibit significant interaction with several of the three subunits of its globular domain (47). The diversity of the recognition modes of C1q is illustrated below by the proposed models of interaction between the top of the C1q head and CRP (Fig. 4) and between the lateral side of subunit B and human IgG1 (Fig. 5).

CRP is a major acute phase plasma protein in man that binds to phosphocholine head groups of membrane phospholipids and is in turn recognized by C1q (13). Its crystal structure (35, 55) shows a pentraxin fold. The C1q binding site identified by mutagenesis experiments (40) lies close to the central pore of

---

**Fig. 5. The C1q-IgG1 interaction.** A, space-filling representation of the proposed interacting faces of human IgG1 b12 (36, 37) and C1q head, showing their shape complementarity. The IgG Fc domain and Fab arms are indicated. Color coding for the mutated residues of IgG1 (9, 41, 42) is as follows. Major charged residues crucial for the interaction (Asp-270, Lys-322) are red, charged residues impairing the interaction (Lys-326, Glu-333) are blue, other crucial residues in the contacting zone (Pro-329, Pro-331) are magenta, and those in the hinge region (Leu-234, Leu-235) are black. The arginine residues of C1q B proposed as possible interaction site are displayed in light blue. B, same representation of the C1q/IgG1 assembly. C, overall view of the interaction between C1q and IgG1. The C1q model was constructed as described under “Experimental Procedures.” The black arrow indicates the antigenic site. D, structure of Mcg (57) highlighting the inability of this IgG1 molecule to bind C1q. The residue color coding is the same as in A. The Fab arms are shown in green.
its pentameric structure, on the face opposite to that attached to the phospholipid-bearing surface. As illustrated on Fig. 4, the top of the C1q head structure, predominantly basic, can be accommodated by the negatively charged central pore of CRP, with a striking shape complementarity between the two proteins. Asp112 and Tyr175 have been identified as major C1q residues of human IgG abolishes its C1q binding activity (56). The fact that, in this model IgG1 Glu333 and Lys326 severely interfere with the access of C1q Glu B162 and ArgB129, respectively (Fig. 5C), raising the possibility of direct additional contacts between C1q and the antigen itself.

In summary, as exemplified above by the models proposed for CRP and IgG, the x-ray structure of the globular domain of C1q provides a structural basis for the versatility of its recognition properties and opens the way to decipher its interaction with many of its immune and non-immune ligands.

Acknowledgments—J. J. is grateful to J. Bordas for support. We thank T. Rabilloud and L. Vuillard for the gift of non-detergent Sulfo betaine 195 and P. Carpentier for help with the use of the x-ray diffraction equipment at D2AM, European synchrotron radiation facility.

REFERENCES

1. Fearon, D. T., and Locksley, R. M. (1996) Science 272, 50–54
2. Hoffmann, J. A., Kabatse, F. C., Janeway, C. A., and Szalai, A. J. (1999) Science 284, 1313–1318
3. Janeway, C. A. (1989) Cold Spring Harbor Symp. Quant. Biol. 54, 1–13
4. Ollier, W., and Reed, K. B. (2000) Immunol. Rev. 170, 159–170
5. Arlau, G. J., Gaboriau, C., Thielens, N. M., Budayova-Spano, M., Rossi, V., and Fontenelle-Camps, J. C. (2002) Mol. Immunol. 39, 383–394
6. Cooper, N. R. (1985) J. Immunol. 135, 151–214
7. Burton, D. R., Boyd, J. A., Esterbrook-Smith, S. B., Emanuel, E. J., Norotyn, J., Rademacher, T. W., van Schravendijk, M. R., and Dwek, R. A. (1980) Nature 288, 378–384
8. Duncan, A. R., and Winter, G. (1988) Nature 332, 728–740
9. Idiosue, E. E., Preza, L. G., Gazzano-Santoro, H., Topal, K., Wong, P. Y., Ulbich, M., Meng, Y. G., and Mullerkin, M. G. (2000) J. Immunol. 164, 4178–4184
10. Thielens, N. M., Tacnet-Delorme, P., and Arlaud, G. J. (2002) Immunology 105, 562–574
11. Santoro, F., Ouaissi, M. A., Pestel, J., and Capron, A. (1989) J. Immunol. 142, 288–2891
12. Bretz, W., Welke, B., Brunner, H., and Loos, M. (1977) Infect. Immun. 15, 317–322
13. Szalai, A. J., Agrawal, A., Greenhough, T. J., and Volanakis, J. E. (1999) Clin. Chem. Lab. Med. 37, 265–270
14. Stasers, S. B., Kolb, W. P., Piekackr, R. N., and Olson, M. S. (1981) J. Biol. Chem. 256, 10924–10929
15. Navaza, J. (2001) Structure 9, 9, 41, 42
16. Rogers, J., Cooper, N. R., Webster, S., Schulte, J. A., McGee, L. P., Styn, S. D., Cwinn, W. H., Brachova, L., Bracht, B., Ward, P., and Lieberburg, I. (1992) Proc. Natl. Acad. Sci. U. S. A. 89, 10140–10142
17. Schulmeister, P., Chevalier, G., and Arlaud, G. J. (2001) J. Immunol. 166, 6381–6391
18. Klein, M. A., Kaefer, P. S., Schwarz, P., Weyd, H., Xenarios, I., Zinkernagel, R. M., Carroll, M. C., Verheek, J. S., Botto, M., Walport, M. J., Molina, H., Kalinke, A., Ucha-Orbea, H., and Aguzzi, A. (2001) Nat. Med. 7, 488–492
19. Mahbott, N. A., Bruce, M. E., Botto, M., Walport, M. J., and Pepey, M. B. (2001) Nat. Med. 7, 485–487
20. Shapiro, L., and Scherer, P. E. (1997) J. Biol. Chem. 272, 335–338
21. Bogin, O., Vransakul, M., Rom, E., Singer, J., Yayon, A., and Hohenester, E. (2002) Structure (Lond.) 10, 165–173
22. Arlau, G. J., Sim, R. B., Dutta, A.-R., and Colomb, M. G. (1979) Mol. Immunol. 16, 445–450
23. Rossi, V., Gaboriau, C., Lacroix, M., Ulrich, J., Fontenelle-Camps, J. C., Gagnon, J., and Arlau, G. J. (1996) Biochemistry 34, 7311–7321
24. Lacroix, M., Rossi, V., Gaboriau, C., Chevalier, S., Saquinod, M., Thielens, N. M., Gagnon, J., and Arlau, G. J. (1997) Biochemistry 36, 6270–6282
25. Vuillard, L., Rabilloud, T., Leberman, B., Berthet-Colominas, C., and Cusack, S. (1994) FEBS Lett. 355, 294–296
26. Pfenner, J.-L., Simon, J.-P., Bérat, J.-F., Cailiot, B., Fanchon, E., Kaikati, O., Arnaud, S., Guidotti, M., Piroch, M., and Roth, M. (1998) J. Synchrotron Radiat. 5, 1346–1356
27. Moy, J. P. (1991) Nucl. Instrum. Methods Phys. Res. A 348, 641–644
28. Narazaka, T. (1991) Acta Crystallogr. Sect. D 47, 1367–1372
29. Bourgeois, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J. N., Pannu, N. S., Read, R. W., Rice, L. M., Simonson, T., and Warren, G. L. (1998) Acta Crystallogr. Sect. D 54, 905–921
30. Jones, A. T., Zhou, J.-Y., Cowan, S. W., and Kjeldgaard, M. (1991) Acta Crystallogr. Sect. A 47, 215–221
31. Poddubskiy, S. A., and Dodson, E. J. (1991) Acta Crystallogr. Sect. D 53, 240–255
32. Rainey, J. K., and Goh, M. C. A. (2002) Protein Sci. 11, 2748–2754
33. Kilchherr, E., Hofmann, H., Steigemann, W., and Engel, J. (1985) J. Mol. Biol. 186, 403–415
34. Thompson, D., Pepys, M. B., and Wood, S. P. (1989) Structure (Lond.) 7, 169–177
35. Saphire, E. O., Parren, P. W., Pontoppidotl, R., Zwick, M. B., Morris, G. M., Rudd, P. M., Dwek, R. A., Stafnell, R. L., Burton, D. R., and Wilson, I. A.
37. Saphire, E. O., Stanfield, R. L., Crispin, M. D., Parren, P. W., Rudd, P. M., Dwek, R. A., Burton, D. R., and Wilson, I. A. (2002) J. Mol. Biol. 319, 9–18
38. Ritchie, D. W., and Kemp, G. J. (2000) Proteins 39, 88–106
39. Ritchie, D. W. (2003) Proteins 52, 98–106
40. Agrawal, A., Shrive, A. K., Greenhough, T. J., and Volanakis, J. E. (2001) J. Immunol. 166, 3998–4004
41. Idusogie, E. E., Wong, P. Y., Presta, L. G., Gazzano-Santoro, H., Totpal, K., Ultsch, M., and Mulkerrin, M. G. (2001) J. Immunol. 166, 2571–2575
42. Hezareh, M., Hessel, A. J., Jensen, R. C., van de Winkel, J. G. J., and Parren, P. W. H. I. (2001) J. Virol. 75, 12161–12168
43. Comis, A., and Easterbrook-Smith, S. B. (1985) Biochim. Biophys. Acta 842, 45–51
44. Marqués, G., Anton, L. C., Barrion, E., Sanchez, A., Ruiz, A., Gavilanes, F., and Vivanco, F. (1993) J. Biol. Chem. 268, 10393–10402
45. Kojouharova, M., Panchev, I. D., Tchorbadjeva, M. I., Reid, K. B. M., and Hoppe, H.-J. (1998) J. Immunol. 161, 4325–4331
46. Kishore, U., Leigh, L. E., Eggleton, P., Strong, P., Perdikoulis, M. V., Willis, A. C., and Reid, K. B. M. (1998) Biochem. J. 333, 27–32
47. Kishore, U., Gupta, S. K., Perdikoulis, M. V., Kojouharova, M. S., Urban, B. C., and Reid, K. B. M. (2003) J. Immunol. 171, 812–820
48. Jones, E. Y., Stuart, D. I., and Walker, N. P. (1989) Nature 338, 225–228
49. Eck, M. J., and Sprang, S. R. (1989) J. Biol. Chem. 264, 17595–17605
50. Petry, F., Reid, K. B. M., and Loo, M. (1994) J. Immunol. 151, 2988–2993
51. Mizuochi, T., Yoneyama, K., Yamashita, K., and Kohata, A. (1978) J. Biol. Chem. 253, 7404–7409
52. Reid, K. B. M., Gagnon, J., and Frampton, J. (1982) Biochem. J. 203, 559–568
53. Villers, C. L., Arlaud, G. J., Painter, R. H., and Colomb, M. G. (1988) FEBS Lett. 217, 289–294
54. Pidcock, E., and Moore, G. M. (2001) J. Biol. Inorg. Chem. 6, 479–489
55. Shrive, A. K., Cheetham, G. M., Holden, D., Myles, D. A., Turnell, W. G., Volanakis, J. E., Pepys, M. B., Bloomer, A. C., and Greenhough, T. J. (1996) Nat. Struct. Biol. 3, 346–352
56. McCall, M. N., and Easterbrook-Smith, S. B. (1989) Biochem. J. 257, 845–851
57. Goddard, L. W., Herren, J. N., and Edmunson, A. B. (1993) Proc. Natl. Acad. Sci. U. S. A. 90, 4271–4275
58. Isenman, D. E., Dorrington, K. J., and Painter, R. H. (1975) J. Immunol. 114, 1726–1729
59. Nicholls, A., Sharp, K. A., and Honig, B. (1991) Proteins Struct. Funct. Genet. 11, 281–296
The Crystal Structure of the Globular Head of Complement Protein C1q Provides a Basis for Its Versatile Recognition Properties
Christine Gaboriaud, Jordi Juanhuix, Arnaud Gruez, Monique Lacroix, Claudine Darnault, David Pignol, Denis Verger, Juan C. Fontecilla-Camps and Gérard J. Arlaud

J. Biol. Chem. 2003, 278:46974-46982.
doi: 10.1074/jbc.M307764200 originally published online September 5, 2003

Access the most updated version of this article at doi: 10.1074/jbc.M307764200

Alerts:
• When this article is cited
• When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts

This article cites 53 references, 23 of which can be accessed free at http://www.jbc.org/content/278/47/46974.full.html#ref-list-1