X-ray Structure of the Ca$^{2+}$-binding Interaction Domain of C1s

INSIGHTS INTO THE ASSEMBLY OF THE C1 COMPLEX OF COMPLEMENT

Received for publication, May 16, 2003
Published, JBC Papers in Press, June 4, 2003, DOI 10.1074/jbc.M305175200

Lynn A. Gregory‡, Nicole M. Thielens§, Gérard J. Arlaud‡, Juan Carlos Fontecilla-Camps‡, and Christine Gaboriaud‡ ¶

From the ‡Laboratoire de Cristallographie et Cristallogénèse des Protéines and the ¶Laboratoire d’Enzymologie Moleculaire, Institut de Biologie Structurale Jean-Pierre Ebel, 41 rue Jules Horowitz, 38027 Grenoble Cedex 1, France

C1, the complex that triggers the classical pathway of complement, is assembled from two modular proteases C1r and C1s and a recognition protein C1q. The N-terminal CUB1-EGF segments of C1r and C1s are key elements of the C1 architecture, because they mediate both Ca$^{2+}$-dependent C1r-C1s association and interaction with C1q. The crystal structure of the interaction domain of C1s has been solved and refined to 1.5 Å resolution. The structure reveals a head-to-tail homodimer involving interactions between the CUB1 module of one monomer and the epidermal growth factor (EGF) module of its counterpart. A Ca$^{2+}$ ion is bound to each EGF module and stabilizes both the intra- and inter-monomer interfaces. Unexpectedly, a second Ca$^{2+}$ ion is bound to the distal end of each CUB1 module, through six ligands contributed by Glu$^{45}$, Asp$^{53}$, Asp$^{98}$, and two water molecules. These acidic residues and Tyr$^{17}$ are conserved in approximately two-thirds of the CUB repertoire and define a novel, Ca$^{2+}$-binding CUB module subset. The C1s structure was used to build a model of the C1r-C1s CUB1-EGF heterodimer, which in C1 connects C1r to C1s and mediates interaction with C1q. A structural model of the C1q/C1r/C1s interface is proposed, where the rod-like collagen triple helix of C1q is accommodated into a groove along the transversal axis of the C1r-C1s heterodimer.

The classical pathway of complement, a major element of innate immunity against pathogens, is triggered by C1, a 790-kDa complex formed from association of a recognition protein C1q with two modular serine proteases, C1r and C1s, that respectively mediate internal activation and proteolytic activity of the complex (1–3). C1q is a protein with the overall shape of a bouquet of flowers, comprising six heterotrimeric collagen-like triple helices that associate to form a N-terminal “stalk” and then diverge to form individual “stems,” each terminating in a C-terminal globular domain (Ref. 4; see Fig. 5). C1r and C1s have homologous modular structures with a N-terminal C1r/C1s, uEGF, bone morphogenetic protein (CUB) module (5), an epidermal growth factor (EGF)-like module of the Ca$^{2+}$-binding type (6), a second CUB module, two complement control protein modules (7), and a chymotrypsin-like serine protease domain. This modular architecture is shared by the mannann-binding lectin-associated serine proteases (MASPs), a group of enzymes involved in the triggering of the lectin pathway of complement (8). Whereas the enzymatic properties of C1r and C1s are mediated by their C-terminal regions, the N-terminal CUB1-EGF domains have interaction properties that are essential to the assembly of the C1 complex. Thus, it is well established that C1s-C1r-C1r-C1s, the tetrameric catalytic subunit of C1, assembles through Ca$^{2+}$-dependent heterodimeric C1r-C1s interactions involving the CUB1-EGF segment of each protease (9–12). Furthermore, the current data (10, 11, 13) are consistent with the hypothesis that the CUB1-EGF moieties of C1r and C1s each contribute ligands for the interaction between the C1s-C1r-C1r-C1s tetramer and C1q sites located in the individual collagen-like stems of the protein (14, 15). Based on these and other features, several low resolution models of the C1 complex have been proposed (2, 16, 17). In an effort to decipher the structure-function relationships of the C1 complex at the atomic level, we have used a dissection strategy that has yielded precise insights into the activation mechanism of C1r (18, 19) and the proteolytic function of C1s (20). We now report the x-ray structure of the CUB1-EGF moiety of C1s, a domain that in the C1 complex associates with the corresponding part of C1r and has the additional ability to form homodimers in the absence of C1r (9–11). The structure reveals a novel, Ca$^{2+}$-binding CUB module subset and yields insights into the C1q/C1r/C1s interface in the C1 complex.

EXPERIMENTAL PROCEDURES

Production and Purification of the Recombinant C1s CUB1-EGF Domain—A DNA fragment encoding the C1s signal peptide and the N-terminal CUB1-EGF segment (residues 1–159 of the mature protein) was amplified by PCR using Vent$_R$ polymerase and the pBS-C1s plasmid (21) as a template, according to established procedures. The sequences of the sense (5‘-GGGATCCATGGATGTGGTAGCGTTCTCG-3’) and antisense (5‘-GGGGATTCTATAAATCTCGGATCTCTCC-3’) primers introduced a BamHI restriction site (underlined) at the 5’ end of the polymerase chain reaction product and a stop codon (bold type) followed by a KpnI site (underlined) at the 3’ end. The amplified DNA was purified using the GeneClean kit (Bio 101), digested with BamHI and KpnI, and cloned into the corresponding sites of the pFastBac1 baculovirus transfer vector (Invitrogen). The resulting construct was characterized by restriction mapping and checked by double-stranded DNA sequencing (Genome Express, Grenoble, France). The recombinant baculovirus was generated using the Bac-to-Bac$^\text{TM}$ system (Invitrogen Corp.), amplified, and titrated as described previously (22). High Five cells (1.75 × 10$^7$ cells/175-cm$^2$ tissue culture flask) were infected with the recombinant virus at a multiplicity of infection of 2 in SF900 II SFM medium (Invitrogen) for 96 h at 28 °C. The supernatant was collected by centrifugation, and diisopropyl phosphorofluoridate was added to a final concentration of 1 mM. The culture supernatant containing the C1s CUB1-EGF segment was dialyzed against 75 mM
Structure of Human C1s Interaction Domain

NaCl, 10 mM imidazole, pH 6.1, and loaded at 1.5 ml/min onto a Q-Sepharose-Fast Flow column (Amersham Biosciences) (2.8 × 12 cm) equilibrated in the same buffer. Elution was carried out by applying a 1-liter linear gradient from 75 to 500 mM NaCl in the same buffer. The fractions containing the recombinant fragment were identified by SDS-PAGE analysis, precipitated by addition of NH4SO4 to 60% (w/v), and left overnight at 4 °C. The pellets were resuspended in 145 mM NaCl, 1 mM EDTA, 50 mM triethanolamine hydrochloride, pH 7.4, and applied onto a TSK G3000 SWG column (7.5 × 600 mm) (Toso Haas) equilibrated in the same buffer. The purified fragment was concentrated to 1.0 mg/ml by ultrafiltration on a PM-10 membrane (Amicon).

Chemical and Functional Characterization of the Recombinant Protein—SDS-PAGE analysis was performed as described previously (9). Mass spectrometry analysis was performed using the matrix-assisted laser desorption ionization technique on a Voyager Elite XL instrument (PerSeptive Biosystems, Cambridge, MA) under conditions described previously (23). High pressure gel permeation chromatography was performed on a TSK G3000 SWG column (7.5 × 600 mm) (TosoHaas) equilibrated in 145 mM NaCl, 50 mM triethanolamine hydrochloride, pH 7.4, containing either 1 mM EDTA or CaCl2, and run at 1 ml/min.

Analysis by surface plasmon resonance spectroscopy of the interaction between the C1s CUB1-EGF domain and intact C1r was performed at 25 °C using an upgraded BIAcore instrument (BIAcore AB, Uppsala, Sweden). The running buffer for protein immobilization was 145 mM NaCl, 5 mM EDTA, 10 mM HEPEs, pH 7.4. The C1s CUB1-EGF domain was diluted to 35 μg/ml in 10 mM formate, pH 3.0, and coupled to the carboxymethylated dextran surface of a CM5 sensor chip (BIAcore AB) using the amine coupling chemistry (BIAcore amine coupling kit). Binding of purified plasma-derived human C1r (24) was measured over 250 resonance units of the immobilized C1s CUB1-EGF segment, at a flow rate of 10 μl/min in 145 mM NaCl, 1 mM CaCl2, 50 mM triethanolamine hydrochloride, pH 7.4. Equivalent volumes of the C1r samples were injected over a surface with immobilized ovalbumin to serve as blank sensorgrams for subtraction of bulk refractive index background. Regeneration of the surface was achieved by injection of 10 μl of 5 mM EDTA. The data were analyzed by global fitting to a 1:1 Langmuir binding model of both the association and dissociation phases for several concentrations simultaneously, using the BIAevaluating 3.1 software (BIAcore). The apparent equilibrium dissociation constant (Kd) was calculated from the ratio of the dissociation and association rate constants (koff/kon).

Crystallization and Data Collection—The C1s CUB1-EGF fragment was concentrated to 6.0–7.8 mg/ml in 145 mM NaCl, 1 mM CaCl2, 50 mM triethanolamine HCl, pH 7.4. The crystals obtained were grown from hanging drop vapor diffusion method by mixing equal volumes of the protein solution and of a reservoir solution composed of 30% (v/v) PEG 400, 0.2 M CaCl2, and 0.1 M HEPES, pH 7.5. Equivalent volumes of the C1r samples were injected over a surface with immobilized ovalbumin to serve as blank sensorgrams for subtraction of bulk refractive index background. Regeneration of the surface was achieved by injection of 10 μl of 5 mM EDTA. The data were analyzed by global fitting to a 1:1 Langmuir binding model of both the association and dissociation phases for several concentrations simultaneously, using the BIAevaluating 3.1 software (BIAcore). The apparent equilibrium dissociation constant (Kd) was calculated from the ratio of the dissociation and association rate constants (koff/kon).

Crystallographic data were collected at 1.5 Å resolution between the C1s CUB1-EGF domain and intact C1r was performed at 25 °C using an upgraded BIAcore instrument (BIAcore AB, Uppsala, Sweden). The running buffer for protein immobilization was 145 mM NaCl, 5 mM EDTA, 10 mM HEPEs, pH 7.4. The C1s CUB1-EGF domain was diluted to 35 μg/ml in 10 mM formate, pH 3.0, and coupled to the carboxymethylated dextran surface of a CM5 sensor chip (BIAcore AB) using the amine coupling chemistry (BIAcore amine coupling kit). Binding of purified plasma-derived human C1r (24) was measured over 250 resonance units of the immobilized C1s CUB1-EGF segment, at a flow rate of 10 μl/min in 145 mM NaCl, 1 mM CaCl2, 50 mM triethanolamine hydrochloride, pH 7.4. Equivalent volumes of the C1r samples were injected over a surface with immobilized ovalbumin to serve as blank sensorgrams for subtraction of bulk refractive index background. Regeneration of the surface was achieved by injection of 10 μl of 5 mM EDTA. The data were analyzed by global fitting to a 1:1 Langmuir binding model of both the association and dissociation phases for several concentrations simultaneously, using the BIAevaluating 3.1 software (BIAcore). The apparent equilibrium dissociation constant (Kd) was calculated from the ratio of the dissociation and association rate constants (koff/kon).

Crystallization and Data Collection—The C1s CUB1-EGF fragment was concentrated to 6.0–7.8 mg/ml in 145 mM NaCl, 1 mM CaCl2, 50 mM triethanolamine HCl, pH 7.4. The crystals obtained were grown from hanging drop vapor diffusion method by mixing equal volumes of the protein solution and of a reservoir solution composed of 30% (v/v) PEG 400, 0.2 M CaCl2, and 0.1 M HEPES, pH 7.5. The crystals obtained were used for micro-seeding in a drop with reservoir solution containing 16% (v/v) PEG 4000, 0.2 M MgCl2, 6.7% glycerol, and 0.1 M HEPES, pH 7.5. Fresh drops were obtained crystals suitable for high resolution X-ray data collection. A native data set with space group P1 was measured at the ESRF beamline ID34-H2 to a resolution of 1.5 Å. The images were processed using the MOSFLM program package (25), and the data were scaled using the Collaborative Computational Project 4 suite (26). Details are given in Table I.

Structure Determination and Refinement—A preliminary structure was solved using the single-wavelength anomalou dispersion method. Heavy atom derivatives were prepared by first transferring the crystal to a solution with a lower concentration of MgCl2 (0.15 M) and then by soaking in a mother liquor containing 0.5 mM TbCl3 and 0.15 M MgCl2. The heavy atom derivative data set was collected at the ESRF beamline BM29 with 1.4 Å resolution and indexed using XDS (27). Two heavy derivative sites were located using the Patterson heavy atom search method implemented in CNS (28). The correct enantiomorph was found by visual inspection of the electron density map. The protein region was distinguished from the solvent. Solvent flattening was carried out with DM (29).

Model building was carried out with the graphics program TURBO (30). The quality of the initial maps permitted the construction of up to 60% of the 318 residues in the asymmetric unit. This initial model was refined using CNS and then used as the molecular replacement search model for the high resolution native data set. The ensuing map was consistent with the amino acid sequence of the N-terminal Asn159 segment of C1s (calculated mass, 18,125 Da).

RESULTS

Characterization of the C1s Interaction Domain—Expression of the C1s CUB1-EGF domain in a baculovirus/insect cells system led to the production of large amounts of material (~20 mg/liter of cell culture). Purification was achieved by ion-exchange chromatography followed by (NH4)2SO4 precipitation and gel permeation. SDS-PAGE analysis showed that the purified protein was homogeneous and migrated as a single band with an apparent molecular mass of 25.5 and 20.5 kDa under reducing and nonreducing conditions, respectively (Fig. 1). Mass spectrometry analysis yielded a value of 18,129 ± 9 Da consistent with the amino acid sequence of the N-terminal Glu1–Asn159 segment of C1s (calculated mass, 18,125 Da).

Analysis by surface plasmon resonance spectroscopy showed that the CUB1-EGF heterodimer was carried out in three steps. First, the structure of the C1s CUB1 module was used as a scaffold for the C1r CUB1 model. After careful examination of the sequence alignment of the CUB1 modules of C1s and C1r, the residues in the structure of C1s CUB1 were replaced by the corresponding residues from the C1r sequence using the graphics program TURBO (30). Next, the C1r EGFR structure determined by NMR spectroscopy (32) was superimposed onto the atomic coordinates of the C1s EGFR B module using the interactive graphics program O (33). The program gave an root mean square of 1.05 Å based on the 33 Ca positions used to carry out the superimposition. Loop 10 of each EGFR module was excluded from the this calculation because of the difference in length and conformation. Finally, the heterodimer was assembled by taking the Protein Data Bank file of the C1s homodimer and superimposing the C1r CUB1 and EGFR models onto monomer B of the C1s homodimer.

The model of the C1q collagen arm is based on published statistical information derived from collagen-like structures (34). The arrangement of the A, B, and C chains in the heterotrimeric triple helix is derived from the crystal structure of the C1q globular domain, and the alignment of the collagen triplets shown in Fig. 5 is the only one compatible with this structure. Different configurations of the C1q/C1r/C1s interface were tested using computer graphics, looking for the most appropriate positioning in terms of shape and charge complementarity between the C1q and C1r/C1s models. Consistently with previously published biochemical analysis it was also included in the selection of the most plausible model (see “Discussion”). The model depicted in Fig. 5 allows ionic interaction between unmodified lysine residues A59, B61, and B68 of C1q and acidic residues of C1r (Asp61 Glu115 or Glu118, and Asp127, respectively). In this configuration, hydrophobic residues of C1q are positioned in a favorable environment at the C1q-C1r interface: methionines B68 and C67 point toward the central six-residue hydrophobic cluster, whereas residues A74, A77, C70, C71, and C74 are in the vicinity of the distal hydrophobic pocket.

Fig. 1. SDS-PAGE analysis of the recombinant C1s CUB1-EGF domain. Lane 1, CUB1-EGF domain, unreduced; lane 2, CUB1-EGF domain, reduced. The positions of unreduced and reduced standard proteins are indicated on the left and right sides of the figure, respectively.
that the immobilized C1s CUB1-EGF domain was able to bind C1r in a Ca\(^{2+}\)-dependent fashion, with a \(K_D\) value of 20.7 nm, similar to the values of 10.9 and 20.2 nm determined previously for intact C1s and the larger N-terminal C1so fragment, respectively (12). Gel filtration analysis of the C1s CUB1-EGF domain indicated that the protein eluted significantly earlier in the presence of Ca\(^{2+}\) ions than in the presence of EDTA, consistent with the known ability of the C1s interaction domain to form Ca\(^{2+}\)-dependent homodimers (9–11).

**Overall Structure**—The structure of the CUB1-EGF interaction domain of human C1s was solved by the single-wavelength anomalous dispersion method and refined at 1.5 Å into a very well defined electron density map (Fig. 2D). The final \(R_{	ext{work}}\) and \(R_{	ext{free}}\) factors are 0.229 and 0.246, respectively, and the refined model has excellent stereochemistry (Table I). In agreement with previous findings (9–11), the CUB1-EGF segment of C1s associates as a Ca\(^{2+}\)-dependent homodimer (Fig. 2). Within each monomer, the CUB1 and EGF modules are assembled in a linear fashion with a Ca\(^{2+}\) ion bound at the intermodular interface (site I). A second Ca\(^{2+}\) ion is bound to the distal part of each CUB1 module (site II). The two monomers interact in a head-to-tail fashion involving major contacts between the CUB1 module of one molecule and the EGF module of its counterpart, the resulting assembly displaying a noncrystallographic pseudo 2-fold symmetry. The C-terminal residues are located at either end of the dimer, indicating where the CUB2 modules follow (Fig. 2, A and B). The overall structure is rather elongated, with a length of approximately 85 Å and a width of 20–40 Å. A side view of the structure (Fig. 2C) reveals that whereas one side is relatively flat, the opposite side is markedly concave and forms a groove.

**A Novel, Ca\(^{2+}\)-binding CUB Module Structure**—Compared with the CUB domain topology established from the x-ray structure of two spermadhesins (35), the C1s CUB1 module reveals a number of particular features (Fig. 3). Like the N-terminal CUB module in C1r and the MASP family, the C1s CUB1 module shows a deletion at its N-terminal end (Fig. 3C). As a result, this module lacks not only the first of the two disulfide bridges characteristic of most CUB domains but also the first two \(\beta\)-strands present in the previously determined CUB structures (Fig. 3A). Thus, whereas CUB domains of the spermadhesin family are organized in two five-stranded \(\beta\)-sheets, each containing two parallel and four anti-parallel strands (35), the C1s CUB1 topology consists of two four-stranded \(\beta\)-sheets, each made of anti-parallel strands (strands 3, 10, 5, and 8 and strands 4, 9, 6, and 7). A further specific feature of the C1s CUB1 structure is the 3/10 helical conformation of the loop (H1) connecting strands \(\beta5\) to \(\beta6\), which is deleted in the spermadhesins (Fig. 3, A and C). Loops 3 and 9, on the same side of the module, and the large insertion loop 7 (a specific feature of the C1r/C1s/MASP family) also exhibit significant differences in length and/or conformation compared with their counterparts in the spermadhesin CUB structures (Figs. 3, A and C). In contrast to these modifications in solvent-exposed regions, the hydrophobic core observed in the spermadhesin family is highly conserved in the C1s CUB1 domain, the 18 hydrophobic or aromatic residues defining the CUB domain signature (5) being conserved in C1s (Fig. 3C). Compared with the spermadhesins, the C1s CUB1 domain shows root mean square deviation values of 1.40 Å (sSFP), 1.50 Å (PSP-I), and 1.54 Å (PSP-II), based on 80–86 homologous residues.

An unexpected feature of the structure is the occurrence of a Ca\(^{2+}\)-binding site (site II) at the distal end of each CUB1 module. The Ca\(^{2+}\) ion is coordinated by six oxygen ligands, namely one side chain oxygen of Glu\(^{45}\), both carboxylate oxygens of Asp\(^{53}\) and Asp\(^{98}\), the main chain carbonyl oxygen of Asp\(^{98}\), and two water molecules (Fig. 3B). The bond distances are in average 2.4 Å, the characteristic value for known Ca\(^{2+}\)-binding sites in proteins (36). In addition, the Ca\(^{2+}\) ion, its ligands, and the neighboring residues Tyr\(^{17}\), Asn\(^{101}\), and Phe\(^{105}\) also participate in an intricate network of hydrogen bonds that connect together strands \(\beta5\), \(\beta6\), \(\beta9\), and \(\beta10\) and loops L3 and L9 (Fig. 3B). Thus, the Ca\(^{2+}\) ion is the central element of a network of interactions that extensively stabilize the distal end of the C1s CUB1 module. The Ca\(^{2+}\) ion in site II is exposed to the solvent and exchangeable for Tb\(^{3+}\), as seen in the heavy atom derivative. Partial replacement by Mg\(^{2+}\) was also observed in crystals grown at high MgCl\(_2\) concentrations. Subtle differences in the coordinating ligands were observed between monomers A and B of the homodimer, including in some cases the involvement of a further water molecule contributing a seventh ligand.

Of the residues involved in the coordination of Ca\(^{2+}\) and the associated network of hydrogen bonds, Asn\(^{101}\) and Phe\(^{105}\) appear to be strictly specific to the CUB1 modules of the C1r/C1s/MASP family (Fig. 3B). In contrast, Tyr\(^{17}\) and the Ca\(^{2+}\) ligands Glu\(^{45}\), Asp\(^{53}\), and Asp\(^{98}\) are conserved in approximately two-thirds of the CUB module repertoire, strongly suggesting that these residues define a novel CUB module subset with the ability to bind Ca\(^{2+}\). This subset is possibly more representative in terms of structure than the spermadhesins.

**Ca\(^{2+}\)-binding Site I and the Intra-monomer CUB1-EGF Interface**—The C1s EGF module exhibits a fold similar to that described for other modules of this type (6), with one major and one minor anti-parallel double-stranded \(\beta\)-sheets (Fig. 2A).
Structure of Human C1s Interaction Domain

### DISCUSSION

We have previously solved the x-ray structure of the C-terminal catalytic domain of human C1s, comprising the second complement control protein module and the serine protease domain (20). The present study describes the structure of the N-terminal CUB1-EGF domain that mediates the interaction properties of C1s, hence establishing ~73% of the structure of this protease. The structure was solved ab initio and refined to a resolution of 1.5 Å. It reveals a number of interesting features that are directly relevant to some of the physicochemical and functional properties of C1s. In addition, as discussed below in light of the recently published structure of the rat MASP-2 CUB1-EGF-CUB2 region (38), the structure has wider implications, e.g. with respect to the assembly of the corresponding CUB1-EGF domain in the proteins of the C1r/C1s/MASP family.

The four Ca$^{2+}$ binding sites observed in the C1s dimer and their particular positions provide an explanation for the Ca$^{2+}$ dependence of this interaction. Site I is a key element of the hydrophobic clusters, Pro$^{19}$ from loop 7 of the CUB1 module makes a hydrophobic contact with Phe131 of the EGF module. (v) Although water molecules are excluded from the above hydrophobic interfaces, they are abundant at the periphery of the pockets, where they participate in two networks of hydrogen bonds providing indirect connections between residues Glu$^{31}$, Tyr$^{33}$, His$^{36}$, Tyr$^{38}$, Thr$^{40}$, Ser$^{78}$, and Pro$^{79}$ of a CUB1 module and Asn$^{133}$, Asn$^{134}$, Gly$^{138}$, and Phe$^{140}$ are either conserved or substituted by similar residues within the C1r/C1s/MASP family.

#### Table I

| Data collection and refinement statistics | Native<sup>a</sup> | SAD<sup>b</sup> |
|-----------------------------------------|------------------|-----------------|
| Space group                            | P1               | P1              |
| Unit cell (Å)                          | $a = 35.14 \ b = 47.50 \ c = 56.68$ | $a = 35.38 \ b = 47.31 \ c = 57.40$ |
| $\lambda$ (Å)                          | 0.933            | 1.649 Å         |
| Resolution (Å)                         | 30–1.50          | 35–3.00         |
| $R_{crys}$                              | 0.064 (0.143)<sup>f</sup> | 0.096 (0.191)<sup>f</sup> |
| % completeness                         | 94.4 (93.3)      | 94.5 (88.6)     |
| Redundancy                             | 5.5 (2.2)        | 2.98 (2.21)     |
| $I/\sigma(I)$ average                  | 6.9 (4.0)        | 13.02 (6.35)    |
| No. of reflections                     | 181271 (16763)   | 39958 (3794)    |
| No. of unique reflections              | 52518 (7588)     | 13394 (1713)    |
| Figure of merit                        | 0.4362           | 0.4539          |
| Phasing power                          | 2.4539           |                 |

---

<sup>a</sup> Collected at ID14-eh2, ESRF.

<sup>b</sup> Collected at BM30 FIP, ESRF.

<sup>c</sup> Statistics for high resolution bin (1.58–1.50 Å) are in parentheses.

<sup>d</sup> Statistics for high resolution bin (3.0–3.15 Å) are in parentheses.

---

Loop 10, which is disordered in the C1r EGF module (32), is much shorter in C1s (Fig. 4B) and structurally well defined, except Phe$^{125}$ in monomer A. The remainder of the C1r and C1s EGF modules shows a root mean square deviation value of 0.90 Å. As predicted from the amino acid sequence (3, 6), a Ca$^{2+}$ ion (site I) is bound to both EGF modules of the CUB1-EGF homodimer (Fig. 2A). The Ca$^{2+}$ ion is coordinated by seven oxygen ligands, including a water molecule (W3) and six ligands provided by the EGF module itself, namely one of the side chain oxygens of Asp$^{116}$ and Glu$^{119}$, the side chain carboxyl of Asn$^{134}$, and the main chain carbonyl oxygen of Ile$^{117}$, Phe$^{118}$, and Gly$^{138}$ (Fig. 4A). As observed in site II, the average bond distance is 2.4 Å. In agreement with analyses performed on recombinant C1s expressed in a baculovirus/insect cell system (21), Asn$^{134}$ lacks $\beta$-hydroxylsine, indicating that post-translational modification of this residue to erythro-$\beta$-hydroxyasparagine, as observed in human serum C1s (37), is not essential for Ca$^{2+}$ binding.

Remarkably, W3 forms hydrogen bonds with three of the Ca$^{2+}$ ligands (Asp$^{116}$, Glu$^{119}$, and Gly$^{138}$) and with Gly$^{32}$ in loop 4 of the CUB1 module (Fig. 4A). This molecule is therefore a key element of a network of interactions connecting together loops 4 and 10 and the main body of the EGF module, thereby deeply stabilizing the intra-monomer CUB1-EGF interface. Further stabilization of this interface is achieved by a second network of hydrogen bonds involving two other water molecules that bridge Glu$^{31}$ and Tyr$^{33}$ (in loop 4 of the CUB1 module) to Gly$^{137}$ (in loop 12 of the EGF module).

#### The Inter-monomer Interface—This interface involves a combination of hydrophobic interactions and hydrogen bonds evenly distributed on the surface (Fig. 2A): (i) About the 2-fold symmetry axis of the dimer lies a central hydrophobic pocket formed by Met$^{4}$, Tyr$^{33}$, and Ile$^{118}$ from each monomer. (ii) At each inter-monomer CUB1-EGF interface, further hydrophobic interactions are achieved by an aromatic triad comprising Tyr$^{3}$ (from the CUB1 module), Phe$^{135}$, and Phe$^{140}$ (from the EGF module). Interestingly, Phe$^{135}$ also coordinates Ca$^{2+}$ through its main chain carbonyl oxygen, thereby providing a link between Ca$^{2+}$-binding site I and the inter-monomer interface. (iii) The distal end of each CUB1-EGF interface is stabilized by an additional hydrophobic pocket involving Tyr$^{13}$, Pro$^{14}$, and His$^{13}$ from the CUB1 module and Pro$^{145}$ and Glu$^{146}$ from the EGF module. (iv) In between the latter two hydrophobic clusters, Pro$^{19}$ from loop 7 of the CUB1 module makes a hydrophobic contact with Phe131 of the EGF module. (v) Although water molecules are excluded from the above hydrophobic interfaces, they are abundant at the periphery of the pockets, where they participate in two networks of hydrogen bonds providing indirect connections between residues Glu$^{31}$, Tyr$^{33}$, His$^{36}$, Tyr$^{38}$, Thr$^{40}$, Ser$^{78}$, and Pro$^{79}$ of a CUB1 module and Asn$^{133}$, Asn$^{134}$, Ser$^{142}$, and Cys$^{143}$ of its partner EGF module. Asn$^{134}$ plays a key part in this concert, because it coordinates Ca$^{2+}$ in site I, forms a hydrogen bond with another Ca$^{2+}$-ligand Asp$^{116}$, and participates in the hydrogen bond network. In the vicinity of the aromatic triad, a direct hydrogen bond takes place between the side chains of Tyr$^{38}$ in the CUB1 module and Asn$^{133}$ in the EGF module. Among the residues involved in the hydrophobic interactions and hydrogen bonds, Tyr$^{5}$, Tyr$^{13}$, Pro$^{14}$, Tyr$^{17}$, Tyr$^{38}$, His$^{41}$, Asn$^{134}$, Gly$^{138}$, and Phe$^{140}$ are either conserved or substituted by similar residues within the C1r/C1s/MASP family (Figs. 3C and 4B), indicating that these proteins likely share the ability to form Ca$^{2+}$-dependent dimers with a head-to-tail configuration similar to that observed for the C1s homodimer.
intra- and inter-monomer CUB1-EGF interfaces, whereas site II provides extensive stabilization of the distal part of the CUB1 module. These findings provide a structural basis for the observations that the interaction domain of C1s (39, 40), and the corresponding region of C1r (41) both exhibit low temperature transitions that are abolished or shifted to higher temperatures in the presence of Ca\textsuperscript{2+} ions. In the same way, it is known that Ca\textsuperscript{2+} ions protect the interaction domain of C1s against proteolysis with plasmin (10). Interestingly, the major sites of plasmin cleavage are Lys\textsuperscript{96} and Arg\textsuperscript{104} (37). The fact that these are located in the vicinity of Ca\textsuperscript{2+}-binding site II (see Fig. 3) suggests a major role of this site in the protective effect of Ca\textsuperscript{2+}.

Ca\textsuperscript{2+}-binding site I was predictable from the occurrence in the C1s EGF module of the consensus sequence characteristic of the Ca\textsuperscript{2+}-binding subset (6). Comparison of our structure with the one determined for the EGF\textsubscript{1} module of blood clotting factor IX (42) reveals that Ca\textsuperscript{2+} ions are coordinated in both cases by seven ligands that form a pentagonal bipyramid. Five of these ligands are contributed by homologous residues in both proteins, corresponding to Asp\textsuperscript{116}, Ile\textsuperscript{117}, Glu\textsuperscript{119}, Asn\textsuperscript{134}, and Phe\textsuperscript{135} in C1s. However, there are slight differences in the coordination of Ca\textsuperscript{2+}: (i) The residue equivalent to Asn\textsuperscript{134} is an Asp in factor IX and contributes two ligands instead of one. (ii) Whereas Gly\textsuperscript{138} of C1s and the equivalent residue of factor IX is not involved in Ca\textsuperscript{2+} binding. (iii) The seventh ligand is a water molecule in C1s, whereas it is supplied by a neighboring EGF molecule in the factor IX structure. Compared with the sites in C1s and factor IX, the Ca\textsuperscript{2+}-binding site observed in the CUB1-EGF-CUB2 domain of rat MASP-2 (38) exhibits more significant differences: (i) Only five coordination ligands, contributed by a water molecule and four residues equivalent to Asp\textsuperscript{116}, Ile\textsuperscript{117}, Asn\textsuperscript{134}, and Phe\textsuperscript{135} of C1s, are observed. (ii) In contrast to Glu\textsuperscript{119} of C1s and Gln\textsuperscript{120} of factor IX, the equivalent residue of rat MASP-2 (Glu\textsuperscript{122}) does not interact with Ca\textsuperscript{2+}. The fact that the C1s and factor IX structures were both refined at a higher resolution (1.5 Å) than the MASP-2 structure (2.7 Å) may explain, at least in part, the above discrepancies. Nevertheless, the subtle differences observed between the C1s and factor IX structures strongly suggest that

![Figure 3. Structure of the C1s CUB1 module.](http://www.jbc.org/)

A, superposition of the C1s CUB1 structure (red) on that of aSFP (47) (blue) (stereo view). The Ca\textsuperscript{2+} ion bound to the C1s CUB1 module is represented as a golden sphere. B, stereo view of the Ca\textsuperscript{2+}-binding site. Oxygen atoms are shown in red, and nitrogen atoms are in blue. Water molecules are represented as light blue spheres. Ionic and hydrogen bonds are represented by dotted black and blue lines, respectively. C, sequence alignment of various CUB modules including the CUB1 and CUB2 modules of C1s, C1r, MASP-1 and MASP-2, the CUB modules of the spermathecin family, and selected CUB modules from PCPE, TSG6, BMP-1, and cubilin. All of the sequences are from human proteins, except PSP-I and -II (porcine) and aSFP (bovine). The secondary structure elements (except strands β1 and β2 and loop L1) and the numbering are those of the C1s CUB1 module. Conserved residues defining the CUB domain signature and cysteines are colored blue, those involved in Ca\textsuperscript{2+} binding in C1s are pink, and those involved in the inter-monomer interface are orange.
Ca\textsuperscript{2+}-binding EGF-like modules may slightly adapt their coordination mode depending on their particular protein context. The second Ca\textsuperscript{2+}-binding site observed at the distal end of the CUB1 modules was totally unexpected, because it represents the first example of such a site in a module of the CUB family. In this respect, the observation that the three residues involved in the coordination of Ca\textsuperscript{2+} (Glu\textsuperscript{45}, Asp\textsuperscript{53}, and Asp\textsuperscript{98}), as well as Tyr\textsuperscript{17}, which is closely associated to the Ca\textsuperscript{2+}-binding site, are conserved in a large proportion of the CUB module repertoire (Fig. 3C), strongly supports the hypothesis that these residues define a particular CUB module subset with the specific ability to bind Ca\textsuperscript{2+}. The fact that the spermadhesins lack the corresponding consensus sequence is consistent with the absence of Ca\textsuperscript{2+} in their structures (35). In contrast, these residues are strictly conserved in the CUB1 modules of the proteins of the C1r/C1s/MASP family as well as in the CUB2 module of MASP-2 (Fig. 3C). Because rat MASP-2 also fulfills this criterion, it is surprising therefore that no Ca\textsuperscript{2+} was seen in the recently published rat CUB1-EGF-CUB2 structure (38). A plausible explanation lies in the fact that, because of the particular location of site II (Fig. 3), Ca\textsuperscript{2+} is exposed to the solvent and therefore readily exchangeable. This is reflected in the fact that the Ca\textsuperscript{2+} ion in site II of C1s could be replaced by Tb\textsuperscript{3+}, whereas substitution of the other Ca\textsuperscript{2+} in site I was not possible, because of its position at the interface between the CUB1 and EGF modules. A likely hypothesis would be that the initial Ca\textsuperscript{2+} concentration used for crystallization of the MASP-2 CUB1-EGF-CUB2 fragment (38) was not sufficient to maintain full occupancy of site II during the crystallization process. Indeed, analysis by these authors of Ca\textsuperscript{2+} binding by isothermal titration calorimetry provided evidence for the occurrence of two Ca\textsuperscript{2+}-binding sites on each monomer: a high affinity site and a lower affinity site (\(K_D < 40\ \mu\text{M}\)) with an occupancy of less than 0.1. A further relevant feature is that CUB1 residues 103–106, and CUB2 residues 218–221 and 223–224 are disordered in the MASP-2 structure (38). Because these residues belong to two of the loops that are stabilized by Ca\textsuperscript{2+} in the C1s structure (Fig. 3), this observation strengthens the hypothesis that Ca\textsuperscript{2+}-binding site II is present in MASP-2 but was not occupied to a significant extent in the structure solved by Feinberg et al. (38).

The residues engaged in hydrophobic interactions and hydrogen bonds at the inter-monomer interface of the C1s CUB1-EGF homodimer are highly conserved in the whole C1r/C1s/MASP family. Based on this observation, it is tempting to speculate that these proteins all have the ability to associate as
head-to-tail dimers with a configuration similar to that observed in the case of C1s. The MASp-2 CUB1-EGF-CUB2 structure reported by Feinberg et al. (38) provides strong support to this hypothesis, because the "compact dimer" considered by these authors as the physiological configuration reproduces, at the level of the CUB1-EGF region, the head-to-tail assembly observed in C1s. Further comparative analysis reveals a number of common features between the C1s and MASp-2 structures, both at the intra-monomer and at the inter-monomer CUB1-EGF interfaces.

Attempts to express the CUB1-EGF domain of C1r in various eucaryotic systems have resulted in the production of low amounts of material, with a marked tendency to aggregation (12), hence precluding analysis by x-ray crystallography. This amounts of material, with a marked tendency to aggregation of the C1 complex, connects C1r to C1s and mediates interaction with the collagen-like triple helices of C1q (9–11, 14, 15). Modeling of the C1r CUB1 module on the coordinates of its C1s counterpart was facilitated by the fact that these modules share 55% sequence homology, with only two extensions at the N-terminal end and in loop 9 of the C1r module (Fig. 3C). The C1r CUB1-EGF model was completed using the average C1r EGF structure determined by NMR (32). Assembly of the C1r-C1s heterodimer was achieved by superimposing the C1r CUB1 and EGF models onto one of the C1s monomers. Remarkably, both the intra- and inter-monomer CUB1-EGF interfaces appear to be maintained in the C1r-C1s heterodimer, notably the hydrophobic pockets at the inter-monomer interface, with only subtle modifications such as the lack of one of the two direct hydrogen bonds observed in the C1s homodimer and no steric hindrance. These observations support the hypothesis that the C1s-C1s homodimer reproduces, for the most part, the interactions occurring in the C1r-C1s heterodimer. Conservation of the dimer interface in the MASp-2 structure (38) also validates the model.

As observed in the C1s structure (Fig. 2C), one side of the C1r-C1s heterodimer forms a groove in the region where the the C1q/C1r/C1s interface using the segment of the C1q triple helix, which is not only a key element of the architecture of the C1 complex but also plays a crucial role in the transduction of the activating signal.

Acknowledgments—We are grateful to Dr. Philippe data for providing assistance and helpful discussion with regard to the single-wavelength anomalous dispersion method.

REFERENCES
1. Cooper, N. R. (1985) Adv. Immunol. 37, 151–216
2. Arlaud, G. J., Colomb, M. G., and Gagnon, J. (1987) Immunol. Today 8, 106–111
3. Arlaud, G. J., Gaboriaud, C., Thielens, N. M., Rossi, V., Berro, H., Hernandez, X., and Fontecilla-Camps, J. C. (2001) Immunol. Rev. 180, 136–145
4. Kishore, U., and Reid, K. B. (2000) Immunopharmacology 49, 159–170
5. Bork, P., and Beckmann, G. (1993) J. Mol. Biol. 231, 539–545
6. Campbell, I. D., and Bork, P. (1996) Curr. Opin. Struct. Biol. 6, 385–392
7. Reid, K. B. M., Bentley, D. R., Campbell, R. D., Chung, L. P., Sim, R. B., Kristensen, T., and Tack, B. F. (1986) Immunol. Today 7, 230–234
8. Fujita, T. (2002) Nat. Rev. Immunol. 2, 346–353
9. Thielens, N. M., Aude, C. A., Lacroix, M. B., Gagnon, J., and Arlaud, G. J. (1990) J. Biol. Chem. 265, 14469–14475
10. Busby, T. F., and Ingham, K. C. (1990) Biochemistry 29, 4613–4618
11. Tsai, S.-W., Poon, P. H., and Schumaker, V. N. (1997) Mol. Immunol. 34, 1273–1280
12. Thielens, N. M., Enrie, K., Lacroix, M., Jaquiniq, M., Hernandez, J.-F., Esser, T. F., and Arlaud, G. J. (1997) J. Biol. Chem. 272, 9131–9134
13. Thielens, N. M., Ily, C., Bally, I. M., and Arlaud, G. J. (1994) Biochem. J. 301, 378–384
14. Siegel, R. C., and Schumaker, V. N. (1983) Mol. Immunol. 20, 53–66
15. Streng, C. J., Siegel, R. C., Phillips, M. L., Poon, P. H., and Schumaker, V. N. (1982) Proc. Natl. Acad. Sci. U. S. A. 79, 886–890
16. Schumaker, V. N., Hansen, W., Kuncher, E., Phillips, M. L., and Poon, P. H. (1986) Mol. Immunol. 23, 557–565
17. Weiss, V., Fauser, C., and Engel, J. (1989) J. Mol. Biol. 218, 573–581
18. Budovsky-Spano, M., Lacroix, M., Thielens, N. M., Arolaud, G. J., Fontecilla-Camps, J. C., and Gaboriaud, C. (2002) EMBO J. 21, 231–239
19. Budovsky-Spano, M., Grabarske, W., Thielens, N. M., Hillen, H., Lacroix, M., Schmidt, M., Fontecilla-Camps, J. C., Arlaud, G. J., and Gaboriaud, C. (2002) Structure 10, 1509–1519
20. Gaboriaud, C., Rossi, V., Bally, I., Arlaud, G. J., and Fontecilla-Camps, J. C. (2000) EMBO J. 19, 1755–1763
21. Luo, C., Thielens, N. M., Gaboriaud, C., Gal, P., Sarvay, M., Tseng, Y., Tosi, M., Zavodszky, P., Arlaud, G. J., and Schumaker, V. N. (1992) Biochemistry 31, 4254–4262
22. Thielens, N. M., Cash, S., Thiel, S., Vorup-Jensen, T., Rossi, V., Jensenius, J. C., and Arlaud, G. J. (2001) J. Immunol. 166, 5068–5077
23. Lacroix, M., Rossi, V., Gaboriaud, C., Chevallier, S., Jaquiniq, M., Thielens, N. M., Gagnon, J., and Arlaud, G. J. (1997) Biochemistry 36, 6270–6282
24. Arlaud, G. J., Sim, R. B., Duplaa, A. M., and Colomb, M. G. (1979) Mol. Immunol. 16, 445–450
25. Leslie, A. G. W. (1991) in Crystallographic Computing (Moras, D., Podjarny, A. D., and Thierry, D. C., eds) pp. 50–61, Oxford University Press, Oxford
26. Collaborative Computational Project 4 (1994) Acta Crystallogr. Sect. D Biol. Crystallogr. 50, 760–763
27. Keough, W. (2001) in International Tables for Crystallography (Rossmann, M. G., and Arnold, E., eds) pp. 730–734, Kluwer Academic Publishers, Dordrecht, The Netherlands
28. Bruguier, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grose-Kustenke, R. W., Jiang, J.-S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) Acta Crystallogr. Sect. D Biol. Crystallogr. 54, 905–921
X-ray Structure of the Ca^{2+}-Binding Interaction Domain of C1s: Insights Into The Assembly of the C1 Complex of Complement
Lynn A. Gregory, Nicole M. Thielens, Gérard J. Arlaud, Juan Carlos Fontecilla-Camps and Christine Gaboriaud

J. Biol. Chem. 2003, 278:32157-32164.
doi: 10.1074/jbc.M305175200 originally published online June 4, 2003

Access the most updated version of this article at doi: 10.1074/jbc.M305175200

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 44 references, 8 of which can be accessed free at http://www.jbc.org/content/278/34/32157.full.html#ref-list-1