# Solution Structure of the Epidermal Growth Factor (EGF)-like Module of Human Complement Protease C1r, an Atypical Member of the EGF Family†

Beate Bersch,*,‡ Jean-François Hernandez,§ Dominique Marion,‡ and Gérard J. Arlaud§

*Laboratoire de Résonance Magnétique Nucléaire and Laboratoire d'Enzymologie Moléculaire, Institut de Biologie Structurale Jean-Pierre Ebel, CNRS-CEA, 41 avenue des Martyrs, 38027 Grenoble, Cedex 1, France*

*Received July 29, 1997; Revised Manuscript Received October 29, 1997*

ABSTRACT: The calcium-dependent interaction between C1r and C1s, the two homologous serine proteases of the first component of human complement C1, is mediated by their N-terminal regions. The latter comprise an epidermal growth factor (EGF)-like module exhibiting the consensus sequence characteristic of $Ca^{2+}$-binding EGF modules, surrounded by two CUB modules. Due to its $Ca^{2+}$ binding ability, the C1r EGF-like module (C1r-EGF) is supposed to participate in the C1r−C1s interaction. An additional interesting feature of C1r-EGF is the unusually large loop connecting the first two conserved cysteine residues. The solution structure of synthetic C1r-EGF (residues 123−175) has been determined using nuclear magnetic resonance and combined simulated annealing−restrained molecular dynamics calculations. The resulting family of 19 structures is characterized by a well-ordered C-terminal part (residues Cys144−Ala174) with a backbone rmsd of 0.7 Å and a disordered N-terminal, including the large loop between the first two cysteines (Cys129 and Cys144). This loop is known to be surface exposed and may be expected to participate in domain−domain or protein−protein interactions. In its C-terminal part, C1r-EGF possesses the characteristic EGF fold with a major and a minor $\beta$-sheet. The latter comprises a $\beta$-bulge, and comparison with other EGF-like modules reveals the existence of two distinct structural and sequential motifs in the bulged part. Additional experiments in the presence of 80 mM $Ca^{2+}$ did not show significant structural variation of C1r-EGF, in keeping with previous observations on blood-clotting factors IX and X.

Human C1 is the multimolecular protease that triggers the classical pathway of complement, a system that participates in innate immunity against various bacteria, parasites, and retroviruses (*1*). Its activation and enzymic activity are mediated by two serine proteases, C1r and C1s,[1] respectively, which share the same type of modular organization, each comprising two CUB modules surrounding a single EGF-like module, a pair of complement control protein (CCP) modules, and a serine protease domain [see reviews by Schumaker et al. (*2*) and Volanakis and Arlaud (*3*)]. A fundamental feature of C1r and C1s is that they exert their catalytic activities within a $Ca^{2+}$-dependent tetramer C1s−C1r−C1r−C1s, which itself associates to a third, nonenzymic protein C1q to form the C1 complex (*4*). The structural determinants responsible for the $Ca^{2+}$-dependent C1r−C1s interactions involved in the assembly of the C1s−C1r−C1r−C1s tetramer are located in the N-terminal α-regions of each protein (*5*). Each of the corresponding fragments (C1rα and C1sα) obtained by limited proteolysis, comprising the first CUB module, the EGF-like module, and the N-terminal disulfide loop of the second CUB module, contains one high-affinity $Ca^{2+}$-binding site and retains the ability to mediate $Ca^{2+}$-dependent protein−protein interaction (*6*). Association of the C1s−C1r−C1r−C1s tetramer with C1q also involves the C1rα and C1sα regions (*7, 8*), which therefore represent key elements of the architecture of the C1 complex. In addition to this structural role, the α-region of C1r has been proposed to participate in the control of the autoactivation of the serine protease domain of the molecule (*7*).

In addition to the usual EGF consensus sequence containing six cysteines forming three disulfide bonds, the EGF-like modules of C1r and C1s exhibit the particular consensus pattern Asp/Asn, Asp/Asn, Asp*/Asn*, Tyr/Phe (where * indicates a $\beta$-hydroxylated residue) that is characteristic of

† This work has been supported by the Centre National de la Recherche Scientifique, the Commissariat à l'Energie Atomique (CEA), the European Union Biotechnology Programme (Contract BIO4-CT96-0662), and Molecular Simulations, Inc. This is publication 460 from the Institut de Biologie Structurale−Jean-Pierre Ebel. A preliminary report of this study was presented at the XVIth International Complement Workshop in June 1996 in Boston, MA.

* To whom correspondence should be addressed.

‡ Laboratoire de Résonance Magnétique Nucléaire.

§ Laboratoire d'Enzymologie Moléculaire.

[1] Abbreviations: CUB, Complement subcomponents C1r/C1s, Uegf, Bone morphogenetic protein-1; EGF, epidermal growth factor; cbEGF, $Ca^{2+}$-binding epidermal growth factor; DQF-COSY, double-quantum filtered correlation spectroscopy; TOCSY, total correlation spectroscopy; HSQC, $^1$H-detected heteronuclear single-quantum coherence; NOESY, nuclear Overhauser spectroscopy; SA, simulated annealing; rMD, restrained molecular dynamics; rmsd, root-mean-square deviation; t-PA, tissue type plasminogen activator; FacIX_1, FacIX_2, FacX_1, and FacX_2, first and second EGF-like modules of coagulation factors IX and X, respectively; hEGF, human epidermal growth factor peptide hormone; TGF-α, transforming growth factor-α; Her-α, heregulin-α; Fib_32, Fib_33, 32nd and 33rd EGF-like module from fibrillin-1; MASP, mannan binding protein-associated serine protease. The nomenclature of complement proteins is that recommended by the World Health Organization. The nomenclature of protein modules is that defined by P. Bork and A. Bairoch [(1995) *Trends Biochem. Sci. 20* (Suppl. March), C03].

S0006-2960(97)01851-5 CCC: $15.00 
Published on Web 02/03/1998

a subset of EGF modules shown to be involved in $Ca^{2+}$ binding (*9*). The hydroxylated residue is an asparagine in both C1r (Asn150) and C1s (Asn134), hydroxylation being complete in C1r and only partial in C1s (*10*, *11*). Given its $Ca^{2+}$ binding characteristics, it was initially assumed that the EGF module of C1r (C1r-EGF) was responsible for the observed $Ca^{2+}$-binding and $Ca^{2+}$-dependent protein−protein interaction properties of the protein. However, deletion of the first CUB module from C1r abolishes its $Ca^{2+}$-dependent interaction properties (*12*). In the same way, the isolated C1r-EGF binds $Ca^{2+}$ with a $K_d$ of 10 mM, which is about 300 times weaker than that measured for fragment C1rα (*13*), indicating the involvement of residues located outside the EGF module.

A very unusual feature of C1r-EGF is the large size of the loop between the two first cysteines, which comprises 14 residues, as opposed to 2−7 residues in other EGF-like modules (*14*). This highly charged loop also contains the single polymorphic site (Ser135/Leu) identified in human C1r (*15*, *16*).

Like other protein modules, EGF modules are found in a wide variety of mostly extracellular proteins, where they occur in different settings. We are hopeful that determination of different high-resolution structures for a specific module type will lead to a better identification of its characteristic structural and functional properties. Only very little structural information has been obtained so far on modules from proteins of the complement system. The particular localization of C1r-EGF between two CUB modules and its unusually large loop as well as its $Ca^{2+}$ binding properties make it an interesting subject for structural studies. With this aim, the 53-residue EGF module of human C1r was synthesized chemically and its three-dimensional structure was determined by NMR spectroscopy. Our study shows that C1r-EGF has a fold similar to that of other EGF-like modules, with the exception of the loop between the first two cysteines which is disordered. Structural comparison with other EGF-like modules leads us to propose a classification of EGF-like modules with regard to structural and sequential characteristics of a β-bulge situated in the minor β-sheet.

## METHODS

*Sample Preparation*. The C1r EGF-like module (residues 123−175) was synthesized chemically by the solid-phase method on an Applied Biosystem 430A synthesizer using the Boc (*tert*-butyloxycarbonyl) chemistry and standard methodology as described elsewhere (*13*). A serine residue was placed at position 135, where a serine/leucine polymorphism occurs in native C1r, and an asparagine residue was placed at position 150, instead of the β-hydroxyasparagine normally found at this position (*10*, *15*−*17*). The latter post-translational modification is not critical for $Ca^{2+}$ binding as deduced from studies on recombinant human C1s (*18*) and blood-clotting factors IX and X (*19*). The peptide was efficiently folded using a mixture of reduced and oxidized glutathione, yielding a single major species (*13*). The correct formation of the three disulfide bridges (1−3, 2−4, and 5−6) characteristic of EGF-like modules was verified by using mass spectrometry and N-terminal sequence analyses of thermolytic fragments (*13*).

For the NMR experiments, approximately 5.4 mg of the lyophilized peptide was dissolved in 500 μL of 95% $H_2O$/5% $D_2O$ to a final concentration of about 1.8 mM. The pH was adjusted to 6.7 by subsequent addition of small amounts of 0.1 M NaOH. For experiments in $D_2O$, the peptide was lyophilized and redissolved in $D_2O$. Additional experiments were recorded in the presence of 80 mM $CaCl_2$, corresponding to a 90% saturation of the $Ca^{2+}$ binding site, as determined previously by one-dimensional NMR (*13*).

*NMR Experiments and Experimental Constraints*. All NMR experiments were performed on a Bruker AMX-600 spectrometer operating at 600 MHz, equipped with a standard 5 mm Bruker probe with an internal pulsed $B_0$ field gradient (PFG) coil. Data were processed using the Bruker XWIN-NMR software (version 1.1.1).

The sequential $^1H$ assignment was performed using the following experiments: DQF-COSY (*20*), clean TOCSY (*21*, *22*), NOESY (*23*), and $^{13}C$-HSQC at natural abundance (*24*). The latter was used for unambiguous assignment of side chain protons using two complementary sets of information; methylene protons were correlated pairwise, and characteristic $^{13}C$ chemical shift ranges were used for $^{13}C$ assignment (*24*). All spectra were recorded in the phase-sensitive mode [States−TPPI (*25*)]. Suppression of the water resonance was achieved either by applying a weak pre-irradiation during the relaxation delay or by applying the WATERGATE pulse sequence in the case of the NOESY experiments (*26*). Mixing times for TOCSY and NOESY spectra were 90 and 200 ms, respectively. Spectra were recorded at 15 °C (NOESY), 20 °C (TOCSY, NOESY, and $^{13}C$-HSQC), and 30 °C (DQF-COSY and TOCSY).

Distance constraints for the structure calculations were obtained from the NOESY experiments acquired in the absence of $Ca^{2+}$ at 15 and 20 °C with 1536 and 800 complex points and spectral widths of 8475 and 6495 Hz in the $t_2$ and $t_1$ dimensions, respectively. An identical NOESY experiment was performed in $D_2O$ at 15 °C in order to discriminate between amide and aromatic resonances. Cross-peaks were assigned by hand and classified in four categories corresponding to upper distance bounds of 2.6, 3.4, 4.5, and 6.0 Å. Pseudo-atom corrections were added for degenerate methylene, methyl, and aromatic protons as proposed by Wüthrich (*27*). In addition to the distance constraints, seven $\chi_1$ dihedral constraints were added on the basis of the relative $NH-C^{\beta}H_{1,2}$ NOESY peak intensities and the $C^{\alpha}H-C^{\beta}H_{1,2}$ cross-peaks in the COSY experiment. No hydrogen bonding distance constraint was used in the structure calculations.

*Structure Calculations*. Structure calculations were performed using DISCOVER (Version 2.9.7) interfaced to the INSIGHT II program (Version 95.0) for structure visualization and result analysis (Molecular Simulations Inc.). A three-stage simulated annealing−restrained molecular dynamics (SA−rMD) protocol was employed, as described in detail elsewhere (*28*). The AMBER4 force field was used in all calculations except the SA protocols, in which a simple quartic nonbond term was employed. All ω dihedrals were forced to trans since all three prolines showed the characteristic strong $C^{\delta}H_2(i)-C^{\alpha}H(i-1)$ NOESY cross-peaks. The three disulfide bonds, whose localization has been determined independently (*13*), were introduced as covalent constraints.

First, the global fold of the peptide was determined using a 60 ps SA protocol starting from randomized coordinates.

The exploratory period of 50 ps was calculated at a nominal temperature of 1000 K. The NOE force constant was scaled to 50 kcal $mol^{-1}$ $Å^{-2}$ during the first 30 ps. After slow cooling to 100 K, the molecule was minimized in the complete AMBER force field. Resolved, but non-stereospecifically assigned protons or methyl groups were allowed to float between the two prochiral positions.

Structures were then refined using a two-stage SA−rMD protocol (*28*). During the SA step, the nominal temperature was set to 2000 K and the NOE force constants were progressively scaled to 50 kcal $mol^{-1}$ $Å^{-2}$. After an exploratory period of 2 ps and subsequent smooth cooling, the molecule was subjected to energy minimization in the full AMBER force field.

In the following rMD calculation, the use of the full force field and an implicit solvent simulation led to physically more viable structures (*28*). After equilibration, the temperature was held to 600 K with the experimental force constant set to 25 kcal $mol^{-1}$ $Å^{-2}$. The molecule was allowed to evolve for a period of 10 ps, and then it was slowly cooled to 300 K where dynamics was continued for 12 ps. Finally, the molecule was minimized using a conjugate gradient algorithm.

As the NOESY spectra acquired in the presence of 80 mM $CaCl_2$ showed no significantly different cross-peak patterns, additional structure calculations were performed in order to test whether the conformations obtained in the absence of $Ca^{2+}$ were compatible with $Ca^{2+}$ binding. In these calculations, a $Ca^{2+}$ ion was added whose ligands were analogous to the structures of the first EGF module of factor IX (*29*) and the EGF module pair from fibrillin-1 (*30*). Five distance constraints with force constants of 200 kcal $mol^{-1}$ $Å^{-2}$, allowing the Ca−O distance to lie between 2.2 and 2.6 Å, were added to the original constraints determined in the absence of $Ca^{2+}$, and the three-step SA−rMD protocol was employed as described above.

The final structural ensembles were selected on the basis of an experimental energy term, including contributions from NOE distance, $\chi_1$, and $\omega$ dihedral as well as chiral constraints. The mean structures of the ensembles were calculated using INSIGHT II by superposing the N, $C^{\alpha}$, and C′ atoms of residues Gln145−Glu175. rmsd values were calculated for the specified atoms in comparison with the mean structure. Hydrogen bonds were considered to be identified if the distance between the hydrogen and the heteroatom or between two heteroatoms was less than 2.5 or 3.0 Å, respectively, and provided that the hydrogen bond angle was greater than 135°.

For structural comparisons, the following coordinates for EGF modules were taken from the Brookhaven Protein Data Bank: E-selectin, crystal structure, 1esl (*31*); human tissue type plasminogen activator (t-PA), minimized average NMR structure, 1tpg (*32*); porcine blood-clotting factor IX (FacIX_1 and FacIX_2), crystal structure, 1pfx (*33*); human FacIX_1, $Ca^{2+}$-loaded, crystal structure, 1edm (*29*); human des(1−45) factor Xa (FacX_2), crystal structure, 1hcg (*34*); and heregulin-α (Her-α), minimized average NMR structure, 1haf (*35*). The coordinates for the minimized average NMR structures for human EGF peptide hormone [hEGF (*36*)], human transforming growth factor-α [TGF-α (*37*)], and a pair of $Ca^{2+}$-binding EGF-like modules from human fibrillin-1 [Fib_32 and Fib_33 (*30*)] were taken from I. D. Campbell's home page. In the case of C1r-EGF, the structure that was the closest to the mean structure in its ordered part was chosen. The coordinates of the structural ensemble have been submitted to the Brookhaven Protein Data Bank (accession code 1apq).

For the search of sequential motifs, the PROSITE tool was used (*38*).

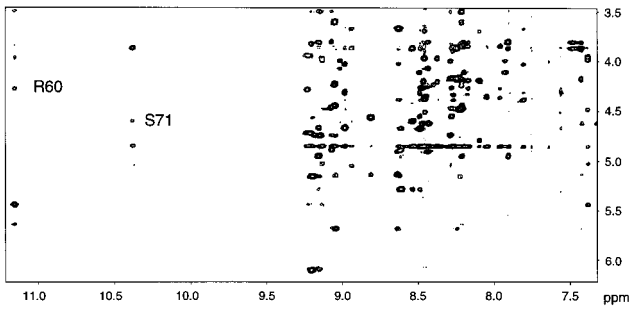


FIGURE 1: Fingerprint region of the NOESY spectrum. This spectrum was recorded at 15 °C in the absence of $Ca^{2+}$ with a sample concentration of 1.8 mM. The $H^N-C^{\alpha}H$ cross-peaks of Arg160 and Ser171 are indicated. Note the very crowded region about an $H^N$ frequency of 8.4 ppm (see the text).

## RESULTS

*Assignment of Proton Resonances*. Using standard techniques, the sequential assignment of the C1r-EGF module was straightforward except for a single residue, Cys144, whose amide proton was absent from the TOCSY spectra. Interestingly, the same phenomenon was recently observed for the homologues Cys9 of the P-selectin EGF module (*39*). Figure 1 shows the fingerprint region of the NOESY spectra. Note the significant downfield shift of the Arg160 and Ser171 amide proton resonances. The same characteristic downfield shift has been observed for a glutamic acid residue in the NMR spectra of several EGF modules [factor IX (*40*), factor X (*41*), TGF-α (*42*), urokinase type plasminogen activator (u-PA) (*43*), and P-selectin (*39*)]. The amide proton resonances of most of the residues in the N-terminal part of the peptide (Ala123−Gln143) are clustered in a region centered around 8.4 ppm, suggesting its conformational flexibility. No line broadening is observed as it would occur in the case of slow chemical exchange. A table listing the proton chemical shifts is available as Supporting Information. Careful analysis of the NOESY spectra recorded in $D_2O$ revealed cross-peaks between the $C^{\beta}H_2$ protons of cysteines linked by a disulfide bond, providing further confirmation of their correct pairing.

*Structure Calculations and Structure Description*. Five hundred forty-three distance constraints (94 intraresidual, 203 sequential, 66 medium-range, and 180 long-range) as well as 6 $\chi_1$ dihedral constraints were used for the calculation of the $Ca^{2+}$-free C1r-EGF module structures. From the sequential distribution of NOESY-derived distances shown in Figure 2a, it became obvious that the N-terminal part of the molecule, particularly the tip and the segment between residues 134 and 142, lacks a well-defined structure, in agreement with the chemical shift values.

After completion of the three-step SA−rMD calculation, an ensemble of 25 structures was obtained, of which 19 conformations were chosen on the basis of their experimental energy violation. The energetic and geometric statistics of

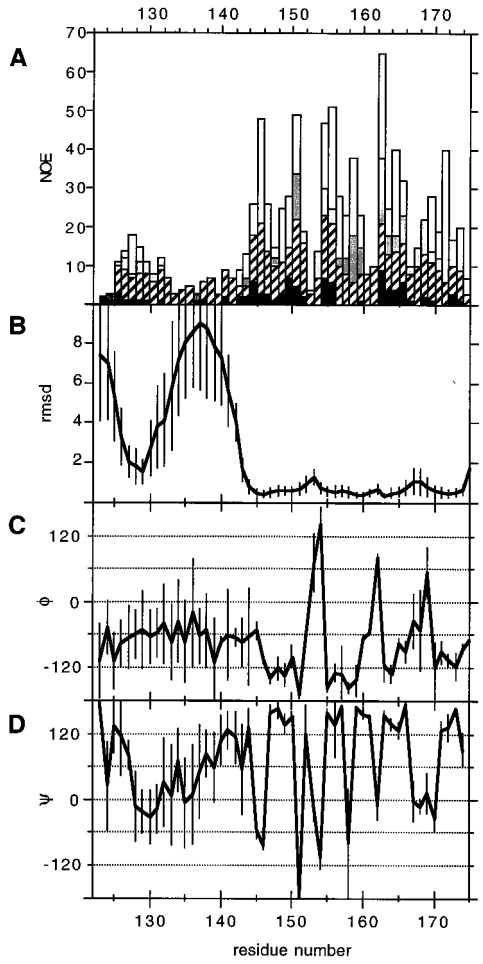


FIGURE 2: Experimental statistics. (A) Sequential distribution and range of NOE constraints: black, intraresidual; gray, sequential; hatched, medium-range (i≤4); and white, long-range constraints. (B) The positional backbone rmsd (N, $C^{\alpha}$, and C′) calculated with respect to the mean structure. (C) Mean values of $\phi$ angles. (D) Mean values of $\psi$ angles. The standard deviation is given by error bars.

the resulting ensemble are summarized in Table 1 and Figure 2, and the structural ensemble is shown in Figure 3. No structure had NOE violations greater than 0.32 Å. Hydrogen bonds occurring in more than 25% of the final structures are listed in Table S2 (Supporting Information). The N-terminal part (Ala123−Gln143) of the molecule is highly disordered, with the exception of residues Asp127−Ala130 which are fixed by the Cys129−Cys148 disulfide bond to the C-terminal part of the module. It is also noteworthy that residues Glu128−Ser131 show a slight preference for an α-helical conformation, as can be seen from the $\phi$ and $\psi$ dihedral means. A similar observation has been made on the corresponding part of human EGF (*36*).

Residues Cys144−Ala174 are structurally well-defined, with a backbone rmsd of 0.7 Å (Table 1), and exhibit a fold characteristic of the EGF modules, consisting of two double-stranded antiparallel $\beta$-sheets. The major sheet is formed by residues Leu147−Tyr151 and Gly154−Ser158. It is stabilized by a regular hydrogen bond network as shown in Figure 4. The two strands of this sheet are linked by a type II $\beta$-turn formed by residues Tyr151−Gly154. The hydrogen bond between Tyr151-O and Gly154-$H^N$ is observed in 14 of the 19 structures.

Table 1: Statistics of C1r-EGF[a]

| | |
|---|---|
| (a) Experimental Statistics | |
| no of experimental violations | |
| >0.05 Å | 26.2 ± 3.5 |
| >0.10 Å | 11.0 ± 2.5 |
| (b) Energetic Statistics | |
| energy term[b] | |
| bond | 8.7 ± 0.6 |
| angle | 73.4 ± 5.4 |
| dihedral | 67.7 ± 7.0 |
| out of plane | 2.3 ± 0.5 |
| H bond | −26.4 ± 2.1 |
| VDW | −146.6 ± 8.2 |
| electrostatic | −651.0 ± 19.7 |
| total | −670.7 ± 20.5 |
| experimental[c] | 8.9 ± 3.0 |
| (c) Structural Statistics | |
| BB[d] (145−174) | 0.70 ± 0.16 |
| heavy (145−174) | 1.39 ± 0.25 |

[a] Final ensemble of 19 structures after SA−rMD calculations. [b] All values are given in kilocalories per mole. [c] Force constants used: $k_{NOE}$ = 25 kcal $mol^{-1}$ $Å^{-2}$, $k_{\chi 1}$ = 60 kcal $mol^{-1}$ $rad^{-2}$. [d] rmsd values (in angstroms) are the average pairwise rmsd relative to the ensemble mean for the residues and atoms shown; backbone atoms correspond to C′, $C^{\alpha}$, and N.

A type II $\beta$-turn between residues Arg160 and Tyr163 connects this major $\beta$-sheet to the second, minor $\beta$-sheet. In that case, the Tyr163-$H^N$···Arg160-O hydrogen bond is found in 5 of the 19 structures. The minor sheet is formed by residues Glu164−Gln166 and Ser171−Gln173. A poorly defined tight turn consisting of residues Gln166−Arg169 is followed by a $\beta$-bulge, involving three residues (Arg169, His170, and Ser171). In this particular area, the normal hydrogen bond pattern of a $\beta$-sheet is interrupted by a hydrogen bond between Gln166-O and Arg169-$H^N$. This bulge can therefore be classified as S3 according to Chan et al. (*44*). A more detailed description of this region is given in the Discussion.

No unique geometry could be determined for the three disulfide bonds, even though $\chi_1$ dihedral constraints were used for Cys148, Cys157, Cys159, and Cys172. According to the $\chi_3$ dihedral, all three disulfides are mixtures of left- and right-handed conformations.

*Calcium Binding*. To determine whether $Ca^{2+}$ binding induces structural changes, particularly in the disordered N-terminal part of the molecule, a set of NMR spectra was acquired in the presence of 80 mM $CaCl_2$ [i.e. a concentration corresponding to 90% saturation of the binding site (*13*)], under otherwise unchanged experimental conditions. The complete peak assignment was performed as in the case of the $Ca^{2+}$-free sample. Chemical shift changes induced by $Ca^{2+}$ ligation are summarized in Figure 5 for $H^N$, $C^{\alpha}H$, and $C^{\beta}H$ protons. Significant chemical shift changes occur in two regions, comprising residues Asp125−Ala130 and Tyr151−Tyr155. Note that the Tyr155-$C^{\alpha}H$ chemical shift variation has previously been used to determine the apparent $K_d$ for $Ca^{2+}$ binding by C1r-EGF (*13*). Both of the regions exhibiting chemical shift changes include or are contiguous to those residues identified as $Ca^{2+}$-binding ligands in other EGF modules (Asp125, Leu126, Glu128, Asn150, and Tyr151 in C1r) (*9*, *29*). Therefore, on the basis of sequence comparison and the chemical shift variations observed upon $Ca^{2+}$ binding, it may be postulated that C1r-EGF binds $Ca^{2+}$

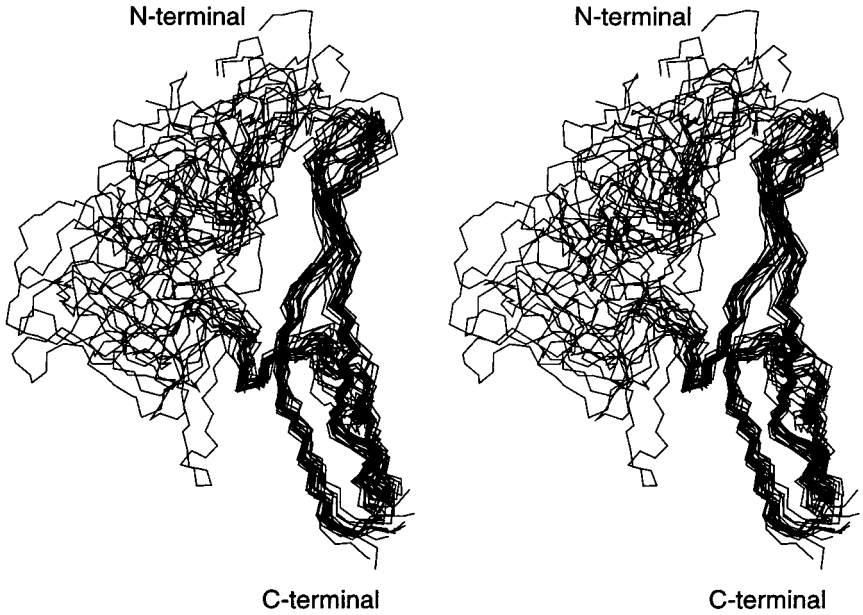


FIGURE 3: Backbone superposition of 19 C1r-EGF structures in a stereorepresentation. The backbone atoms (N, $C^{\alpha}$, and C′) of residues Cys144−Ala174 were superposed on the mean structure.

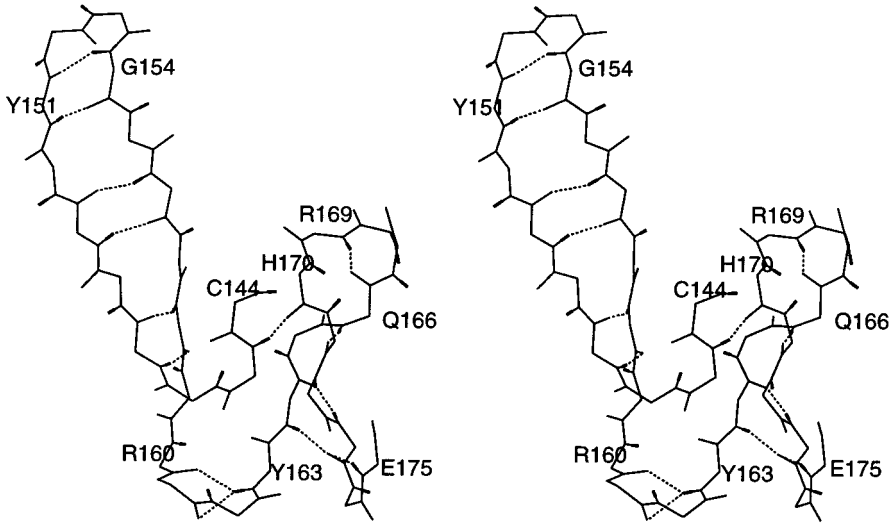


FIGURE 4: Secondary structure and hydrogen bond network of residues Cys144−Glu175. Hydrogen bonds are given by dotted lines, and the amide proton is represented by a broad tick. Side chains have been omitted for clarity.

in a manner similar to that observed for EGF modules from blood-clotting factors IX (*29*) and X (*45*) as well as from human fibrillin-1 (*30*).

Comparison of the NOESY spectra acquired under identical experimental conditions either in the absence or in the presence of $Ca^{2+}$ did not reveal significant change in the cross-peak patterns. However, the quality of the NOESY spectra decreased with added $Ca^{2+}$, probably because of the dielectric losses in the NMR probe due to increased ionic strength and exchange effects due to incomplete calcium saturation. No additional NOEs between the two amino acid stretches involved in $Ca^{2+}$ binding were observed. These two regions are already in contact in the apo form as deduced from NOEs between residues Asp127, Glu128, and Cys129 and the Asn150 side chain and are linked by the Cys129−Cys148 disulfide bond. From these NOE data, we could not deduce significant $Ca^{2+}$-induced structural stabilization of the N-terminal residues involved in $Ca^{2+}$ binding, or of the disordered loop between Cys129 and Cys144. This confirms previous observations made on the first EGF module of factor X that $Ca^{2+}$ binding does not induce significant structural changes (*41*, *45*).

This prompted us to test whether our experimental data set obtained in the absence of $Ca^{2+}$ was compatible with $Ca^{2+}$ binding by analogy with the crystal structure of the first $Ca^{2+}$-bound EGF module of human factor IX (*29*). Five additional distance constraints (from $Ca^{2+}$ to Asp125-$O^{\delta}$, Glu128-$O^{\epsilon}$, Asn150-$O^{\delta}$, Leu126-O, and Tyr151-O) were added according to this structure (*29*). A sixth ligand is probably provided by the Glu128 side chain, as factor IX Asp64 is replaced by Asn150 in C1r-EGF (*9*; see also the sequence alignment presented in Figure 7). Addition of $Ca^{2+}$ distance constraints did not lead to systematic violations of NOE distance constraints, and both the experimental and physical energy terms were found to be almost identical with those obtained in the absence of $Ca^{2+}$ (23.5 and −701 kcal $mol^{-1}$, respectively). A final ensemble of 26 conformations selected on the basis of their experimental energy was

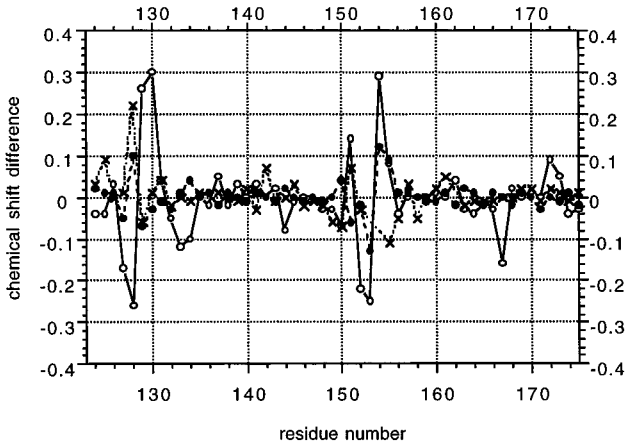


FIGURE 5: Chemical shift variation induced by $Ca^{2+}$ binding. The chemical shift difference ($\delta_{Ca} - \delta_{apo}$) is given in parts per million: $H^N$, open circles and continuous line; $C^\alpha H$, closed circles and broken line; and $C^\beta H$, crosses and dotted line. In the case of two different values ($C^\beta H_2$, and $C^\alpha H_2$), the most important one has been taken.

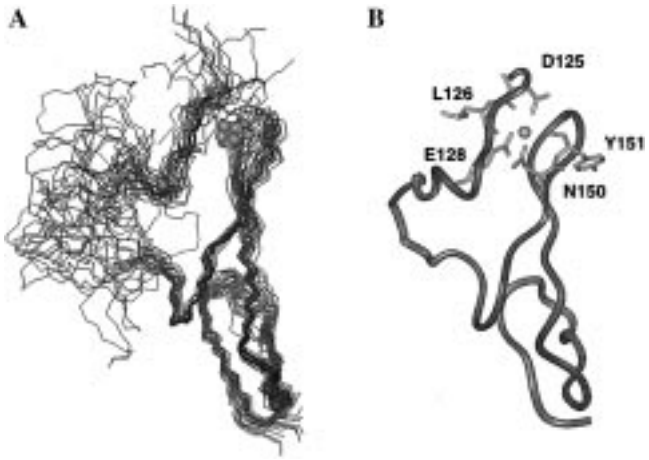


FIGURE 6: Model of the $Ca^{2+}$-bound form of C1r-EGF. (A) Backbone representation of the 23 selected structures shown in the same orientation as in Figure 3. The backbone atoms of residues Cys144−Ala174 were superposed on the mean structure. (B) Ribbon representation of the mean structure. The heavy atoms of the $Ca^{2+}$-binding residues are shown for one structure. $Ca^{2+}$ is represented by a sphere.

obtained after the three-stage SA−rMD protocol, of which three were further excluded because the orientation of the large loop was not consistent with the experimental constraints. Figure 6A shows the resulting 23 structures in the same orientation as the apo form (see Figure 3), and Figure 6B shows the $Ca^{2+}$ binding ligands on a ribbon representation. The ordered C-terminal part of the molecule adopts the same conformation as that in the apo form as deduced from the rmsd between the two ensembles (0.42 Å for the $C^\alpha$, N, and C′ atoms of residues 145−174 of the corresponding ensemble means), the distribution of $\phi$ and $\psi$ dihedrals, and the hydrogen bond networks (data not shown). On the other hand, the order in the N-terminal part, particularly from residue Val124 to Ala130, is significantly increased due to the additional $Ca^{2+}$ distance constraints. However, this observation is only based on the structure modeling and could not be related to any increase in the number of NOEs in the corresponding spectra (see above).

Table 2: Comparison of Atomic Coordinates for Different EGF-like Modules[a]

| | C1r-EGF (148−151, 154−163)[b] | hEGF (20−23, 28−37)[b] | TGF-α (21−24, 29−38)[b] |
|---|---|---|---|
| C1r-EGF | − | 1.21 | 1.62 |
| FacIX_2 | 0.76 | 1.16 | 1.68 |
| FaxX_2 | 0.85 | 0.95 | 1.53 |
| FacIX_1 | 1.71 | 0.97 | 1.29 |
| E-sel | 1.55 | 0.79 | 0.85 |
| hEGF | 1.21 | − | 1.27 |
| TGF-α | 1.62 | 1.27 | − |
| t-PA | 1.85 | 1.25 | 1.59 |
| Her-α | 1.96 | 1.02 | 1.31 |

[a] rmsd calculated in angstroms over the trace atoms of the reduced EGF core as defined in the text. [b] Residue numbers corresponding to the reduced EGF core.

## DISCUSSION

The family of EGF modules is extremely widespread, and probably more than 600 different versions of the EGF module have so far been identified from sequence analysis (*14*). As sequence comparisons of various EGF modules reveal that the number of amino acids located between the six conserved cysteines varies consistently (*14*, *46*), EGF-like modules may be best characterized by the presence of two double-stranded antiparallel $\beta$-sheets, connected by loops of varying lengths.

Figure 7 shows a sequence alignment for selected EGF-like modules with known three-dimensional structures. Besides the six cysteines involved in the three disulfide bonds, there are only very few residues which are conserved among the different EGF-like modules. Indeed, the glycine at position 162 in C1r-EGF seems to be the only strictly conserved residue, this being always followed by an aromatic amino acid. In addition, the subclass of $Ca^{2+}$-binding EGF-like modules (cbEGF) contains the characteristic $Ca^{2+}$-binding consensus sequence highlighted in Figure 7.

Table 2 resumes the structural comparison of different EGF-like modules, for which the trace atoms of the reduced core have been superimposed. The core of the EGF-like modules is defined as the region between the third cysteine and the conserved aromatic amino acid (residues Cys148−Tyr163 in C1r-EGF) and includes the major $\beta$-sheet and the adjacent turn (*30*). However, in several EGF-like modules, residue insertions at the top of the major $\beta$-sheet lead to the replacement of the tight turn by an Ω-loop. Therefore, we introduce a reduced core which excludes the central residues of the turn or Ω-loop. It becomes obvious that the structure of C1r-EGF is very similar to FacIX_2 and FacX_2 (rmsd < 0.9 Å), whereas the rmsd with the other EGF-like modules varies between 1.2 and 1.9 Å. The three-dimensional structures of some representative EGF modules superposed on that of C1r-EGF are shown in Figure 8, the N-terminal part of the molecules being discarded for clarity. It can be seen that the backbone conformations of C1r-EGF, FaxIX_2, and FacX_2 are nearly identical, whereas the folds of TGF-α, E-selectin, and FacIX_1 differ from that of C1r-EGF, mainly in the shape and orientation of the minor $\beta$-sheet.

Indeed, sequence comparison shows that in these particular cases the two C-terminal cysteines [Cys(5) and Cys(6)] are separated either by 12 residues, as in the case of C1r, or by 8 residues, as for example in the human EGF peptide

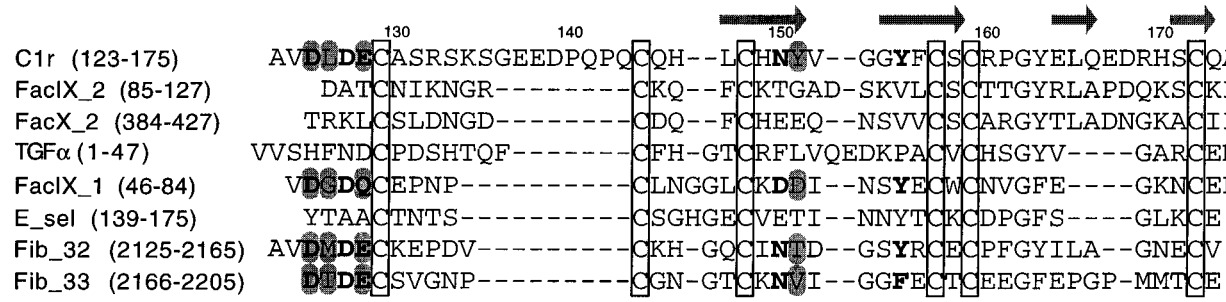


FIGURE 7: Sequence alignment of selected EGF-like modules. The six conserved cysteines are boxed. Residues belonging to the $Ca^{2+}$ consensus sequence are in bold letters, and $Ca^{2+}$-binding residues are shaded. Arrows illustrate the $\beta$-sheet regions observed in C1r, and the numbering corresponds to the C1r sequence. All sequences are from human species except for the Fac_IX sequences which are from porcine factor IX.

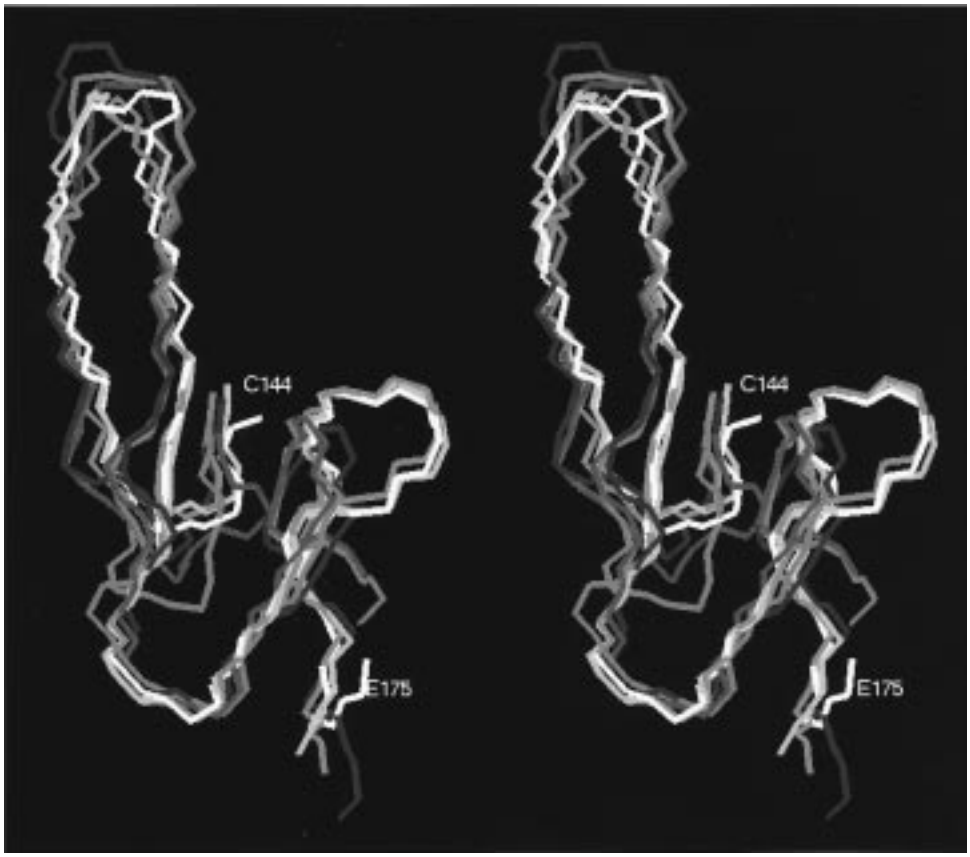


FIGURE 8: Superposition of the backbone atoms of several EGF-like modules in a stereorepresentation. C1r-EGF is shown in white, and the second EGF-like modules of factors IX and X are shown in yellow and orange, respectively; the first EGF-like module of factor IX is shown in pink, E-selectin in blue, and TGF-α in violet. This representation results from the superposition of the reduced core on the C1r-EGF coordinates, as defined in the text.

hormone (see Figure 7). In what follows, we will discuss conserved structural motifs for each of these two groups, referred to as C1r and hEGF homologues for differentiation. Other examples of C1r homologues are the second EGF-like modules of factors IX (FacIX_2) and X (FacX_2), that do not possess the $Ca^{2+}$ binding consensus sequence, whereas human transforming growth factor-α (TGF-α), heregulin-α, the first EGF-like module of factors IX (FacIX_1) and X (FacX_1), and the EGF-like modules from E- and P-selectin as well as that from human tissue type plasminogen activator (t-PA) are hEGF homologues.

Figure 9 shows a schematic representation of the minor $\beta$-sheets of three C1r and three hEGF homologues. Comparison of the individual structures within each group reveals that long-range hydrogen bond networks are well-conserved. In both groups, a $\beta$-bulge immediately precedes the tight turn. The nomenclature of Richardson et al. (*47*) has been adopted, in which residues on the bulged side are labeled 1, 2, 3, ..., whereas the residue on the opposite strand is called X. The bulge in the C1r homologues belongs to the S3 class (*44*) in which the amide of residue X is linked to the carbonyl of residue 1 and the amide of residue 3 to the residue X carbonyl. In the three C1r homologues, residue 1 is characterized by a positive $\phi$-angle (*33*, *34*). In most EGF-like modules containing the Cys(5)−$Xxx_{12}$−Cys(6) pattern, there are two residues which are highly conserved. A residue with a long and bulky side chain is found at the position before residue X, which most often corresponds to a leucine (Leu165 in C1r-EGF), but can be replaced by methionine, valine, glutamine, or isoleucine. As in our structures, this

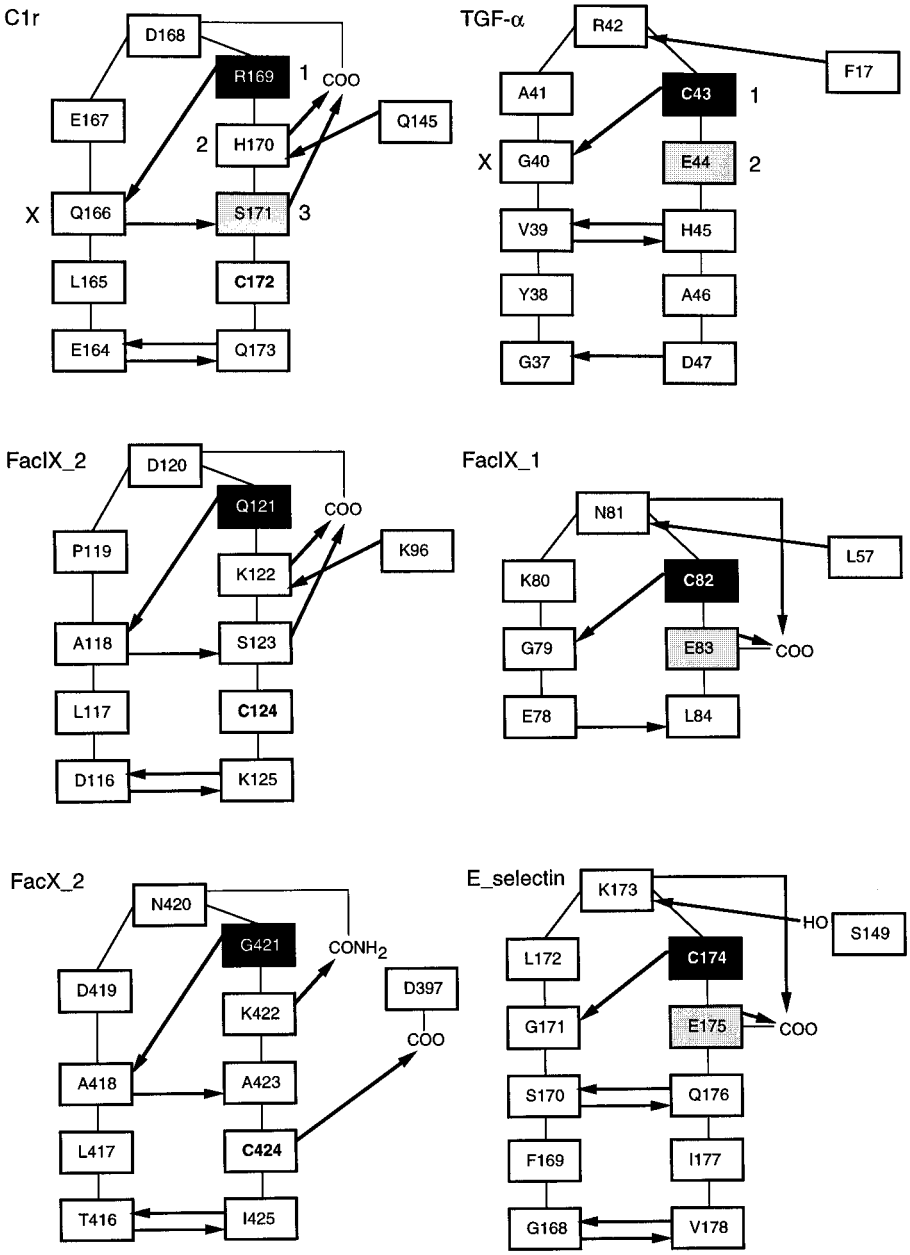


FIGURE 9: Schematic representation of the minor $\beta$-sheet of three C1r (left) and three hEGF homologues (right). Hydrogen bonds are indicated by arrows from the donor to the acceptor. The functional groups of side chains involved in such hydrogen bonds are shown. Residues with a positive $\phi$ angle are shaded in dark gray, and those whose amide proton is characterized by a strong downfield shift are marked in light gray. The nomenclature used in the text is made explicit for the two structures at the top.

residue is found to be in van der Waals contact with Cys157, and its possible function might be the stabilization of the minor $\beta$-sheet. The second conserved residue corresponds to C1r Asp168, situated in the tight turn. A residue with a polar side chain is conserved in most cases. It may be that the corresponding side chain is involved in hydrogen bonds, as was found in the case of C1r-EGF, FacIX_2, and FacX_2 (see Figure 9). Note that most EGF-like modules that do not possess a bulky residue (L, M, I, V, and Q) at position X − 1 do not have the second conserved residue in the tight turn. This strongly suggests that these two residues are characteristic of the structural motif found in the C1r homologues.

On the other hand, the bulges of the hEGF homologues do not fall into one of the classes described by Chan et al. (*44*). In this group, residue X always corresponds to a glycine. It receives a hydrogen bond from residue 1, which corresponds to Cys(6), and has a positive $\phi$-angle, as in the C1r homologues. The same hydrogen bond pattern also occurs in heregulin-α and in t-PA. As in the C1r family, two residues are highly conserved in the $\beta$-bulge structure of the hEGF homologues. The first one is the above-mentioned glycine, the second one being a glutamic acid, located at position 2. As in the C1r homologues, this residue can be replaced by residues with polar side chains (Q, D, N, T, and S), able to form hydrogen bonds. Again, a correlation exists between the conservation of the glycine and this polar side chain, indicating their importance for the corresponding structural motif.

An interesting consequence of the occurrence of these two different structural motifs is the position of Cys(6) relative to the $\beta$-bulge; in both cases, this lies on the bulged strand,

but corresponds to residue 1 in the hEGF homologues, whereas it is located immediately after the bulge structure in the C1r homologues. It is clear therefore that there is sequential but no structural homology for this cysteine between these two types of EGF-like modules.

Two other striking features are the conservation of a hydrogen bond between the minor $\beta$-sheet and the residue following the second cysteine (see Figure 9), as well as the downfield shift observed for an amide proton in the $\beta$-bulge structure. In the hEGF homologues, it corresponds to the conserved glutamic acid following Cys(6) (E44 in TGF-$\alpha$). The chemical shift of its amide proton has been found above 9.6 ppm in several hEGF homologues (*36*, *39*−*43*, *48*). There is no chemical shift information available on the C1r homologues except C1r, in which the Ser171 amide proton has been detected at 10.4 ppm. A possible reason might be a hydrogen bond with a carboxyl group, as described by Freedman et al. (*39*) in the case of P-selectin. Such hydrogen bonds involving the conserved glutamic acid have been found in at least two hEGF homologues, FacIX_1 (*33*) and E-selectin (*31*), whereas in the solution structures of hEGF, TGF-$\alpha$, heregulin-$\alpha$, and u-PA, the glutamic acid side chain seems to be unconstrained. In the three C1r homologues with known three-dimensional structures, analogous hydrogen bonds exist which are formed with the side chain of the conserved polar residue preceding the $\beta$-bulge (Asp168 in C1r-EGF; see above).

C1r-EGF contains an unusual loop connecting the first and second cysteines (Cys129 and Cys144). With its 14 amino acids, it is much larger than the corresponding segment of most EGF-like modules identified so far [maximum of 7 residues (*14*)], with known exceptions being the 29th EGF-like module of the notch gene product from *Drosophila melanogaster* (*49*) and the recently identified tunicate MASP protein from *Halocynthia roretzi* (*50*). From the structural studies presented here, it is clear that this loop does not possess a unique structure, at least in the isolated EGF-like module. The high number of charged residues (two basic and three acidic amino acids) and the absence of hydrophobic residues strongly suggest that this loop is surface-exposed, in agreement with the observation that the arginyl bond at position 134 is cleaved upon limited proteolysis of intact C1r by trypsin (*6*). A likely hypothesis is that this loop participates via electrostatic contacts in domain−domain interactions within C1r, or in protein−protein interactions within the C1 complex. Indeed, various data provide support for the occurrence of a $Ca^{2+}$-dependent interaction between the $\alpha$-region of C1r and its C-terminal catalytic region (*7*, *51*). Another possibility is that the extended loop of C1r-EGF participates in the interaction between the C1s−C1r−C1r−C1s tetramer and C1q (*7*). Whatever the precise interaction role of this loop, its large size and apparent flexibility make it particularly well adapted for such a function.

EGF modules with the consensus sequence for $Ca^{2+}$ binding occur in a number of extracellular proteins with diverse functions. To understand the structural and functional consequences of $Ca^{2+}$ binding, numerous structural investigations have been performed on single modules or module pairs in the presence and absence of $Ca^{2+}$ (*29*, *30*, *33*, *40*, *41*, *45*, *52*−*55*). In this study, we have shown that the binding of $Ca^{2+}$ to the EGF-like module from C1r does not induce major structural modification of its solution structure. This observation is in accordance with previous results obtained on the single cbEGF modules from blood coagulation factors IX (*29*, *52*) and X (*41*) as well as those from human fibrillin (*55*). From the NMR investigation of $Ca^{2+}$-bound factor X EGF, there was some indication that $Ca^{2+}$ might stabilize the peptide N-terminal end, but the authors discussed the possibility that the lower resolution of the apo form might be due to the lack of experimental constraints (*41*). In our case, there is no experimental evidence for $Ca^{2+}$-induced stabilization of the N-terminal end, but modeling studies with additional $Ca^{2+}$-binding constraints suggest an increased order in this region. Therefore, the present study does not allow us to draw a firm conclusion on this question. As C1r-EGF possesses the unusually large loop in its N-terminal part, one other aim of our study was to determine whether $Ca^{2+}$ binding would have some influence on the stability of this loop. This hypothesis can be ruled out on the basis of both the experimental data and modeling studies.

Nevertheless, chemical shift changes are observed upon $Ca^{2+}$ addition, and their profile (see Figure 5) suggests that C1r-EGF binds $Ca^{2+}$ in the same way as blood coagulation factors IX and X or fibrillin EGF modules (*41*, *45*, *52*). Interestingly, all studies performed on isolated $Ca^{2+}$-binding EGF modules show $Ca^{2+}$ binding constants that are much weaker than those determined on the entire molecule or functional parts of it (*13*, *56*−*59*). It was initially assumed that the reason for this may be the incomplete ligation of the $Ca^{2+}$ ion by the EGF-like module on its own, as this provides only six of the ideally seven $Ca^{2+}$ ligands (*29*, *41*), and that the additional ligand could be provided by the preceding module. Experimental support for this hypothesis came from the crystal structure of the first factor IX cbEGF module in which an asparagine belonging to a second molecule in the crystal lattice participates in $Ca^{2+}$ ligation (*29*). More recently, NMR structures of two module pairs, $\gamma$-carboxyglutamate (GLA)−cbEGF from factor X (*45*) and cbEGF−cbEGF (Fib_32,Fib_33) from human fibrillin-1 (*30*), became available. In these studies, the donation of an additional $Ca^{2+}$ ligand from the N-terminal module could not be confirmed. Indeed, the difference in $Ca^{2+}$ binding affinity, also observed on the two $Ca^{2+}$ sites of the Fib_32,-Fib_33 double module (*59*), seems to reflect the stability of the $Ca^{2+}$ binding site (*30*). The module−module interface would thus contribute to the maintenance of the correct conformation for high-affinity $Ca^{2+}$ binding (*30*). In addition, it may be assumed from these two studies that $Ca^{2+}$ intervenes at the module−module interface by inducing a relative reorientation of the module pair as seen in factor X (*45*) and/or by the stabilization of the module−module contacts. The observation that $Ca^{2+}$ binding stabilizes fibrillin-1 against proteolytic degradation underlies this hypothesis (*60*). With respect to C1r, recent experiments on the recombinant N-terminal CUB−cbEGF module pair (N. Thielens, K. Enrié, M. Lacroix, A. F. Esser, and G. J. Arlaud, unpublished data) show that this assembly exhibits a high-affinity $Ca^{2+}$ binding site and is able to interact with C1s in a $Ca^{2+}$-dependent fashion. It is clear therefore that the N-terminal CUB module contributes in some way to $Ca^{2+}$ binding. Additional structural studies on this particular module pair are required to elucidate the mechanism of $Ca^{2+}$

binding and its functional role in human C1r. These would also provide a further model for the study of $Ca^{2+}$ binding by an EGF module within a module pair.

## ACKNOWLEDGMENT

We are very grateful to Dr. M. J. Blackledge for interesting discussions and his help with the structure calculations. We thank F. Hatt for her participation in NOE cross-peak assignment and initial structure calculations.

## SUPPORTING INFORMATION AVAILABLE

Two tables listing the $^1H$ chemical shifts and the hydrogen bonds occurring in more than 20% of the final structures (4 pages). Ordering information is given on any current masthead page.

## REFERENCES

1. Cooper, N. R. (1985) *Adv. Immunol. 37*, 151−216.
2. Schumaker, V. N., Zavodszky, P., and Poon, P. H. (1987) *Annu. Rev. Immunol. 5*, 21−42.
3. Volanakis, J. E., and Arlaud, G. J. (1997) in *Human Complement System in Health and Disease* (Frank, M., and Volanakis, J. E., Eds.) Marcel-Dekker, New York (in press).
4. Arlaud, G. J., Colomb, M. G., and Gagnon, J. (1987) *Immunol. Today 8*, 106−111.
5. Villiers, C. L., Arlaud, G. J., and Colomb, M. G. (1985) *Proc. Natl. Acad. Sci. U.S.A. 82*, 4477−4481.
6. Thielens, N. M., Aude, C. A., Lacroix, M. B., Gagnon, J., and Arlaud, G. J. (1990) *J. Biol. Chem. 265*, 14469−14475.
7. Thielens, N. M., Illy, C., Bally, I. M., and Arlaud, G. J. (1994) *Biochem. J. 301*, 509−516.
8. Busby, T. F., and Ingham, K. C. (1990) *Biochemistry 29*, 4613−4618.
9. Handford, P. A., Mayhew, M., Baron, M., Winship, P. R., Campbell, I. D., and Brownlee, G. G. (1991) *Nature 351*, 164−167.
10. Arlaud, G. J., van Dorsselaer, A., Bell, A., Mancini, M., Aude, C., and Gagnon, J. (1987) *FEBS Lett. 222*, 129−134.
11. Thielens, N. M., van Dorsselaer, A., Gagnon, J., and Arlaud, G. J. (1990) *Biochemistry 29*, 3570−3578.
12. Cseh, S., Gal, P., Sarvari, M., Dobo, J., Lorincz, Z., Schumaker, V. N., and Zavodszky, P. (1996) *Mol. Immunol. 33*, 351−359.
13. Hernandez, J.-F., Bersch, B., Pétillot, Y., Gagnon, J., and Arlaud, G. J. (1997) *J. Pept. Res. 49*, 221−231.
14. Campbell, I. D., and Bork, P. (1993) *Curr. Opin. Struct. Biol. 3*, 385−392.
15. Leytus, S. P., Kurachi, K., Sakariassen, K. S., and Davie, E. W. (1986) *Biochemistry 25*, 4855−4863.
16. Journet, A., and Tosi, M. (1986) *Biochem. J. 240*, 783−787.
17. Arlaud, G. J., and Gagnon, J. (1987) *Biochem. J. 241*, 711−720.
18. Luo, C., Thielens, N. M., Gagnon, J., Gal, P., Sarvari, M., Tseng, Y., Tosi, M., Zavodszky, P., Arlaud, G. J., and Schumaker, V. N. (1992) *Biochemistry 31*, 4254−4262.
19. Selander-Sunnerhagen, M., Persson, E., Dahlqvist, I., Drakenberg, T., Stenflo, J., Mayhew, M., Robin, M., Handford, P., Tilley, J. W., Campbell, I. D., and Brownlee, G. G. (1993) *J. Biol. Chem. 268*, 23339−23344.
20. Rance, M., Sørensen, O. W., Bodenhausen, G., Wagner, G., Ernst, R. R., and Wüthrich, K. (1983) *Biochem. Biophys. Res. Commun. 117*, 479−485.
21. Davis, D. G., and Bax, A. (1985) *J. Am. Chem. Soc. 107*, 2820−2821.
22. Griesinger, C., Otting, G., Wüthrich, K., and Ernst, R. R. (1988) *J. Am. Chem. Soc. 110*, 7870−7872.
23. Macura, S., Huang, Y., Suter, D., and Ernst, R. R. (1981) *J. Magn. Reson. 43*, 259−281.
24. Medvedeva, S., Simorre, J.-P., Brutscher, B., Guerlesquin, F., and Marion, D. (1993) *FEBS Lett. 333*, 251−256.
25. Marion, D., Ikura, M., Tschudin, R., and Bax, A. (1989) *J. Magn. Reson. 85*, 393−399.
26. Piotto, M., Saudek, V., and Sklenár, V. (1992) *J. Biomol. NMR 2*, 661−665.
27. Wüthrich, K. (1986) *NMR of proteins and nucleic acids*, John Wiley, New York.
28. Blackledge, M. J., Medvedeva, S., Poncin, M., Guerlesquin, F., Bruschi, M., and Marion, D. (1995) *J. Mol. Biol. 245*, 661−681.
29. Rao, Z., Handford, P. A., Mayhew, M., Knott, V., Brownlee, G. G., and Stuart, D. (1995) *Cell 82*, 131−141.
30. Downing, A. K., Knott, V., Werner, J. M., Cardy, C. M., Campbell, I. D., and Handford, P. A. (1996) *Cell 85*, 597−605.
31. Graves, B. J., Crowther, R. L., Chandran, C., Rumberger, J. M., Li, S., Huang, K.-S., Presky, D. H., Familletti, P., Wolitzky, B. A., and Burns, D. K. (1994) *Nature 367*, 532−538.
32. Smith, B. O., Downing, A. K., Driscoll, P. C., Dudgeon, T. J., and Campbell, I. D. (1995) *Structure 3*, 823−833.
33. Brandstetter, H., Bauer, M., Huber, R., Lollar, P., and Bode, W. (1995) *Proc. Natl. Acad. Sci. U.S.A. 92*, 9796−9800.
34. Padmanabhan, K., Padmanabhan, K. P., Tullinsky, A., Park, C. H., Bode, W., Huber, R., Blankenship, D. T., Cardin, A. D., and Kisiel, W. (1993) *J. Mol. Biol. 232*, 947−966.
35. Jacobsen, N. E., Abadi, N., Siwkowski, M. X., Reilly, D., Skelton, N. J., and Fairbrother, W. J. (1996) *Biochemistry 35*, 3402−3417.
36. Hommel, U., Harvey, T. S., Driscoll, P. C., and Campbell, I. D. (1992) *J. Mol. Biol. 227*, 271−282.
37. Harvey, T. S., Wilkinson, A. J., Tappin, M. J., Cooke, R. M., and Campbell, I. D. (1991) *Eur. J. Biochem. 198*, 555−562.
38. Bairoch, A., Bucher, P., and Hofman, K. (1995) *Nucleic Acids Res. 24*, 189−196.
39. Freedman, S. J., Sanford, D. G., Bachovchin, W. W., Furie, B., and Baleja, J. D. (1996) *Biochemistry 35*, 13733−13744.
40. Baron, M., Norman, D. G., Harvey, T. S., Handford, P. A., Mayhew, M., Tse, A. G. D., Brownlee, G. G., and Campbell, I. D. (1992) *Protein Sci. 1*, 81−90.
41. Selander-Sunnerhagen, M., Ullner, M., Persson, E., Teleman, O., Stenflo, J., and Drakenberg, T. (1992) *J. Biol. Chem. 267*, 19642−19649.
42. Tappin, M. J., Cooke, R. M., Fitton, J. E., and Campbell, I. D. (1989) *Eur. J. Biochem. 179*, 629−637.
43. Hansen, A. P., Petros, A. M., Meadows, R. P., Nettesheim, D. G., Mazar, A. P., Olejniczak, E. T., Xu, R. X., Pederson, T. M., Henkin, J., and Fesik, S. W. (1994) *Biochemistry 33*, 4847−4864.
44. Chan, A. W. E., Hutchinson, E. G., Harris, D., and Thornton, J. M. (1993) *Protein Sci. 2*, 1574−1590.
45. Sunnerhagen, M., Olah, G. A., Stenflo, J., Forsén, S., Drakenberg, T., and Trewhella, J. (1996) *Biochemistry 35*, 11547−11559.
46. Bork, P., Downing, A. K., Kieffer, B., and Campbell, I. D. (1996) *Q. Rev. Biophys. 29*, 119−167.
47. Richardson, J. S., Getzoff, E. D., and Richardson, D. C. (1978) *Proc. Natl. Acad. Sci. U.S.A. 75*, 2574−2578.
48. Smith, B. O., Downing, A. K., Dudgeon, T. J., Cunningham, M., Driscoll, P. C., and Campbell, I. D. (1994) *Biochemistry 33*, 2422−2429.
49. Wharton, K. A., Johansen, K. M., Xu, T., and Artavanis-Tsakonas, S. (1985) *Cell 43*, 567−581.
50. Ji, X., Azumi, K., Sasaki, M., and Nonaka, M. (1997) *Proc. Natl. Acad. Sci. U.S.A. 94*, 6340−6345.
51. Pelloux, S., Thielens, N. M., Hudry-Clergeon, G., Pétillot, Y., Filhol, O., and Arlaud, G. J. (1996) *FEBS Lett. 386*, 15−20.
52. Huang, L. H., Cheng, H., Pardi, A., Tam, J. P., and Sweeney, W. V. (1991) *Biochemistry 30*, 7402−7409.
53. Ullner, M., Selander, M., Persson, E., Stenflo, J., and Drakenberg, T. (1992) *Biochemistry 31*, 5974−5983.

54. Sunnerhagen, M., Forsén, S., Hoffrén, A.-M., Drakenberg, T., Teleman, O., and Stenflo, J. (1995) *Nat. Struct. Biol. 2*, 504−509.
55. Wu, Y.-S., Bevilacqua, V. L. H., and Berg, J. M. (1995) *Curr. Biol. 2*, 91−97.
56. Handford, P., Downing, A. K., Rao, Z., Hewett, D. R., Sykes, B. C., and Kielty, C. M. (1995) *J. Biol. Chem. 270*, 6751−6756.
57. Astermak, J., Björk, I., Öhlin, A. K., and Stenflo, J. (1991) *J. Biol. Chem. 266*, 2430−2437.
58. Valcarce, C., Selander-Sunnerhagen, M., Tämlitz, A. M., Drakenberg, T., Björk, I., and Stenflo, J. (1993) *J. Biol. Chem. 268*, 26673−26678.
59. Knott, V., Downing, A. K., Cardy, C. M., and Handford, P. (1996) *J. Mol. Biol. 255*, 22−27.
60. Reinhardt, D. P., Ono, R. N., and Sakai, L. Y. (1997) *J. Biol. Chem. 2*, 1231−1236.

BI971851V