Structure and Assembly of the Catalytic Region of Human Complement Protease C1r: A Three-Dimensional Model Based on Chemical Cross-Linking and Homology Modeling†

Monique Lacroix,‡ Véronique Rossi,‡ Christine Gaboriaud,§ Sylvie Chevallier,‡ Michel Jaquinod,|| Nicole M. Thielens,‡ Jean Gagnon,‡ and Gérard J. Arlaud*,‡

Laboratoire d'Enzymologie Moléculaire, Laboratoire de Cristallogénèse et Cristallographie des Protéines, and Laboratoire de Spectrométrie de Masse des Protéines, Institut de Biologie Structurale Jean-Pierre Ebel (CEA-CNRS), 41, avenue des Martyrs, 38027 Grenoble Cedex 1, France

Received October 30, 1996; Revised Manuscript Received February 18, 1997⊗

ABSTRACT: C1r is the modular serine protease responsible for autocatalytic activation of C1, the first component of the complement classical pathway. Its catalytic region is a noncovalent homodimer of two γ-B monomers, each comprising two contiguous complement control protein (CCP) modules, IV and V [also known as short consensus repeats (SCRs)], a 15-residue intermediary segment, and the serine protease B domain. With a view to gain insight into domain−domain interactions within this region, fragment C1r (γ-B)$_2$, obtained by autolytic proteolysis of the active protease, was cross-linked with the water-soluble reagent 1-ethyl-3-[3-(dimethylamino)propyl]carbodiimide. Cross-linked species γ-B *intra* and γ-B *inter*, containing intra- and intermonomer cross-links, respectively, were isolated and then fragmented by CNBr cleavage and trypsin digestion. N-Terminal sequence and mass spectrometry analyses of the resulting cross-linked peptides allowed us to identify one intramonomer cross-link between Lys$_{426}$ of module V and the C-terminal Asp$_{688}$ of the serine protease B domain and one intermonomer cross-link between the N-terminal Gly$_{280}$ of fragment γ and Glu$_{493}$ of the B domain. Three-dimensional homology modeling of the CCP modules IV and V and of the B domain was also performed. The complementary information provided by chemical cross-linking and homology modeling studies was used to construct a three-dimensional model of the γ-B monomer, in which module V interacts with the serine protease on the side opposite to both the active site and the Arg$_{446}$−Ile$_{447}$ activation site. Also, a tentative three-dimensional model of the (γ-B)$_2$ dimer was built, indicating a loose "head to tail" association of the monomers, with the active sites facing opposite directions toward the outside of the dimer. The latter model is compared with available low-resolution structural data, and its functional implications are discussed in terms of the conformational changes occurring during C1r activation.

Human C1 is the complex modular protease that initiates the classical pathway of complement in response to the formation of immune complexes and to infection by various nonimmune activators, such as certain bacteria, parasites, and retroviruses, including HIV. Initial recognition of the target by C1 is mediated by C1q and triggers activation of its catalytic subunit C1s−C1r−C1r−C1s, a calcium-dependent assembly of two serine proteases.[1] This two-step activation process involves autolytic activation of C1r and then C1r-mediated conversion of proenzyme C1s into the active enzyme responsible for limited proteolysis of C4 and C2, the protein substrates of C1 [see reviews by Cooper (1985), Schumaker et al. (1987), and Arlaud et al. (1987a)].

Human C1r is a noncovalent homodimer of M_r 172 600, each monomer comprising a series of protein modules, including two CUB modules, a single EGF-like module, two CCP (or SCR) modules, and a serine protease domain. This same type of modular organization is also shared by C1s and by MASP, a protease which, in association with mannan binding lectin, forms a complex responsible for the initiation of the "lectin pathway" of complement activation (Takayama et al., 1994; Volanakis & Arlaud, 1997). Both C1r autoactivation and subsequent activation of C1s by C1r are mediated by its catalytic (γ-B)$_2$ region (Lacroix et al., 1989), a noncovalent dimer that forms the core of the C1r−C1r dimer and C1s−C1r−C1r−C1s tetramer (Villiers et al., 1985). The various studies performed on C1r autoactivation are consistent with the occurrence of an intramolecular cross-

† This is Publication No. 424 from the Institut de Biologie Structurale Jean-Pierre Ebel. This work was supported by the Commissariat à l'Energie Atomique and the Centre National de la Recherche Scientifique. A preliminary report of this study was presented at the XVIth International Complement Workshop in Boston, MA, June 1996.

* To whom correspondence should be addressed.
‡ Laboratoire d'Enzymologie Moléculaire.
§ Laboratoire de Cristallogénèse et Cristallographie des Protéines.
|| Laboratoire de Spectrométrie de Masse de Protéines.

⊗ Abstract published in *Advance ACS Abstracts,* May 1, 1997.

[1] Abbreviations: CCP modules, complement control protein modules [also known as short consensus repeats (SCRs) or sushi modules]; CUB, complement C1r, C1s, *u*egf, *b*one morphogenetic protein; EDC, 1-ethyl-3-[3-(dimethylamino)propyl]carbodiimide; EDTA, ethylenediaminetetraacetic acid; HPLC, high-pressure liquid chromatography; MALDI, matrix-assisted laser desorption ionization; MASP, mannan binding lectin-associated serine protease; PTH, phenylthiohydantoin; SDS−PAGE, sodium dodecyl sulfate−polyacrylamide gel electrophoresis. The nomenclature of complement components is that recommended by the World Health Organization; activated components are indicated by an overbar, e.g., C1r. The nomenclature of protein modules is that defined by Bork and Bairoch (1995) following the International Workshop on Extracellular Protein Modules held in September 1994 in Margretetorp, Sweden.

S0006-2960(96)02719-5 CCC: $14.00

mechanism involving mutual cleavage of each γ-B monomer by its neighbor (Arlaud et al., 1985; Cooper, 1985). In the proenzyme form, each γ-B monomer, derived from the C-terminal part of C1r, is a single-chain polypeptide comprising a tandem repeat of CCP modules (IV and V), a 15-residue intermediary segment, and the serine protease (B) domain. Activation occurs through cleavage of the Arg_{446}−Ile_{447} bond, thereby splitting the molecule in between the intermediary segment and the B domain and yielding two polypeptides (γ and B) connected by a single disulfide bridge (Arlaud et al., 1986). $C\bar{1}r$ autolytic cleavage, as well as limited proteolysis by proteases of various specificities, yields $(\gamma\text{-B})_2$ dimers of slightly different lengths, depending on the position of the cleavage sites, which all occur within a 13-residue segment of $C\bar{1}r$ (positions 274−286), at the N-terminal end of the γ segment (Arlaud et al., 1986). Although these $(\gamma\text{-B})_2$ dimers are seen by electron microscopy after rotary shadowing as twin domains of globular appearance (Villiers et al., 1985), the negative staining technique suggests the occurrence of discrete domains in each γ-B monomer (Weiss et al., 1986), in agreement with differential scanning calorimetry studies performed on the homologous γ-B fragment of C1s (Medved et al., 1989). Apart from preliminary data indicating that each $C\bar{1}r$ γ-B monomer interacts with its neighbor in a "head to tail" configuration (Arlaud et al., 1986), no precise information has been obtained yet on the structure and assembly of these regions, which play a central role in C1 activation.

Recent studies based on the combined use of chemical cross-linking with EDC, a reagent able to convert ionic bonds between carboxyl and amino groups into covalent pseudopeptide bonds (Means & Feeney, 1971), and homology modeling, have provided insights into the three-dimensional structure of the γ-B region of the homologous protease $C\bar{1}s$ (Rossi et al., 1995). This same approach was applied in the present study to the $C\bar{1}r$ $(\gamma\text{-B})_2$ region. The complementary information provided by cross-linking and modeling studies has led to three-dimensional models of the γ-B monomer and $(\gamma\text{-B})_2$ dimer, allowing insight into the assembly of the domains involved in C1r activation and activity.

EXPERIMENTAL PROCEDURES

Materials. Human plasmin was obtained from Kabi Vitrum, Stockholm, Sweden. Thermolysin from *Bacillus thermoproteolyticus* and trypsin (sequencing grade) were from Boehringer Mannheim, France. S. A. EDC and diisopropyl fluorophosphate were purchased from Sigma.

Isolation of Fragment $C\bar{1}r$ $(\gamma\text{-B})_2$. $C\bar{1}r$ was purified from human plasma as described previously (Arlaud et al., 1979). Limited proteolysis of $C\bar{1}r$ by plasmin and thermolysin was performed as reported by Arlaud et al. (1986). Autolytic cleavage was realized as described by Arlaud et al. (1986), and the reaction mixture was incubated twice in the presence of 2 mM diisopropyl fluorophosphate for 30 min at 30 °C and dialyzed against 50 mM triethanolamine hydrochloride (pH 7.4). Purification of the autolytic $C\bar{1}r$ $(\gamma\text{-B})_2$ fragment was achieved by ion-exchange chromatography on a Mono Q HR 5/5 column (Pharmacia), using the same buffer and a linear NaCl gradient from 0 to 0.25 M in 20 min. The flow rate was 0.5 mL/min, and fragment $(\gamma\text{-B})_2$ was detected from its absorbance at 280 nm. The concentrations of purified $C\bar{1}r$ and its γ-B fragment were determined from A_{280} measurements by using respective values of $E(1\%, 1\text{ cm}) = 12.4$ and 14.5 and $M_r = 86\,300$ and 50 400 (Thielens et al., 1990a; Zaccaï et al., 1990; this study).

SDS−PAGE Analysis. Proteins were analyzed on 10% or 7.5% polyacrylamide gels as described by Laemmli (1970). Phosphorylase *b* (M_r 94 000), bovine serum albumin (M_r 67 000), ovalbumin (M_r 43 000), carbonic anhydrase (M_r 30 000), soybean trypsin inhibitor (M_r 20 100), and α-lactalbumin (M_r 14 400) were used as molecular weight markers.

Chemical Cross-Linking and Isolation of the Cross-Linked γ-B Intra and γ-B Inter Species. The autolytically cleaved $C\bar{1}r$ $(\gamma\text{-B})_2$ fragment (213 nmol) was dialyzed against 150 mM NaCl/20 mM 2-(*N*-morpholino)ethanesulfonic acid (pH 6.5), diluted to 0.5 mg/mL, and then incubated for 1 h at 30 °C in the presence of 10 mM EDC. The reaction mixture was dialyzed exhaustively against the same buffer without EDC and then against 0.5% acetic acid and freeze-dried. The cross-linking reaction mixture was dissolved in 6 M urea/0.2 M formic acid (2.6 mL) and fractionated by high-pressure gel permeation on two successive Ultraspherogel SEC 3000 columns (7.5 mm × 300 mm) (Beckman) equilibrated in the same medium. Fractions corresponding to γ-B monomers and $(\gamma\text{-B})_2$ dimers were dialyzed against 0.5% acetic acid, freeze-dried, and then dissolved in 5 mL of 6 M guanidine hydrochloride/2 mM EDTA/0.4 M Tris-HCl (pH 8.5). Each species was reduced by incubation in the presence of 0.5% (v/v) 2-mercaptoethanol for 2 h at 25 °C and alkylated by a further incubation for 2 h at 25 °C in the presence of 4% (v/v) 4-vinylpyridine. The reduced and S-pyridylethylated samples were dialyzed against 0.5% acetic acid, freeze-dried, and fractionated by high-pressure gel permeation on Ultraspherogel SEC 3000 columns as described above. Fractions corresponding to cross-linked γ-B species originating from γ-B monomers (γ-B *intra*) and $(\gamma\text{-B})_2$ dimers (γ-B *inter*) were collected, dialyzed against 0.5% acetic acid, and freeze-dried.

CNBr Cleavage of γ-B Intra and γ-B Inter Species. Cross-linked samples γ-B *intra* (40 nmol) and γ-B *inter* (31 nmol), along with a control, un-cross-linked, reduced, and S-pyridylethylated γ-B sample (58 nmol), were dissolved in 70% trifluoroacetic acid (1 mL) and incubated in the presence of CNBr (CNBr/protein = 30/1 w/w) for 24 h at 4 °C in the dark. After 1:15 dilution with water, the CNBr-cleavage peptides from each γ-B sample were freeze-dried, dissolved in 6 M urea/20% trifluoroacetic acid, and separated by reverse-phase HPLC on a C4 Deltapak column (3.9 mm × 150 mm) (Waters Associates). The column was equilibrated with a mixture of solutions A (0.1% trifluoroacetic acid) and B (acetonitrile/methanol/propan-2-ol, 1/1/1 v/v/v) in the ratio 95:5 (v/v) and eluted with a linear gradient to give a final ratio of 20:80 (v/v) in 60 min. The flow rate was 0.8 mL/min, and peptides were detected from absorbance at 215 nm.

Isolation of the Cross-Linked Peptides. The cross-linked CNBr peptides contained in pool γ-B *intra*-5 (approximately 5 nmol) were dissolved in 0.5 mL of 0.1 M NH_4HCO_3/2 mM $CaCl_2$ (pH 8.0) and digested by 4% (w/w) trypsin for 2 h at 37 °C. The resulting tryptic peptides were separated by reverse-phase HPLC on a Novapak C18 column (3.9 mm × 150 mm), using the same elution system as described above, by means of a linear gradient from 5% to 60% solution B in 30 min.

The CNBr peptides contained in pool γ-B *inter*-9 were further separated by high-pressure gel permeation on a TSK

G-2000 SW column (7.5 mm × 600 mm) (LKB) equilibrated in 0.1% trifluoroacetic acid and eluted at a flow rate of 1 mL/min. The cross-linked material eluting in the first peak (approximately 6 nmol) was dissolved in 0.4 mL of 1 M urea/20 mM methylamine/0.1 M NH_4HCO_3 (pH 8.0) and cleaved with 4% (w/w) trypsin for 3 h at 37 °C and then further digested overnight at 37 °C after a second addition of 4% (w/w) trypsin. Tryptic peptides were separated by reverse-phase HPLC on a Vydac C18 column (4.6 mm × 250 mm) (Hesperia), using the elution system described above, by means of a linear gradient from 5% to 65% in 50 min.

N-Terminal Sequence Analysis. N-Terminal sequence analyses were performed as described previously (Rossi et al., 1995).

Mass Spectrometry Analyses. Fast atom bombardment mass spectrometry analyses were carried out on small, nonglycosylated peptides as described previously (Thielens et al., 1990b).

The matrix-assisted laser desorption ionization (MALDI) technique was used in the case of large and/or glycosylated peptides, and spectra were recorded on a Voyager Elite XL instrument (Perseptive Biosystems, Cambridge, MA). α-Cyano-4-hydroxycinnamic acid or 2,5-dihydroxybenzoic acid dissolved in a 70% acetonitrile solution containing 0.1% TFA was used as a matrix. One microliter of the peptide sample (1 μg/μL) was mixed with an equal volume of saturated matrix solution directly on the steel sample plate; then the mixture was allowed to dry at room temperature. Spectra were recorded from 64 to 128 laser shots (nitrogen laser, 337 nm) with an accelerating voltage of 20 keV in the linear mode. In order to improve both the accuracy and sensitivity, measurements were carried out using the pulsed source delay mode. Calibration was performed with human insulin and horse hemoglobin.

Computer-Assisted Three-Dimensional Homology Modeling. The program O (Jones et al., 1991) was used on an ESV/20 graphics station (Evans & Sutherland) to build homology-based three-dimensional models of the protein modules of the $C\bar{1}r$ γ-B segment and to assemble the modules interactively on the basis of the information yielded by chemical cross-linking. The protocol used for modeling the individual modules was similar to that described by Greer (1990) for mammalian serine proteases and applied to complement serine proteases by Perkins and Smith (1993) and is detailed in our previous study on the $C\bar{1}s$ γ-B region (Rossi et al., 1995).

The serine proteases used as reference structures for modeling the $C\bar{1}r$ B chain were the following: bovine chymotrypsin (Birktoft & Blow, 1972); bovine trypsin (Bartunik et al., 1989); porcine elastase (Sawyer et al., 1978); porcine kallikrein (Bode et al., 1983); rat mast cell protease (Remington et al., 1988); *Streptomyces griseus* trypsin (Read & James, 1988); rat tonin (Fujinaga & James, 1987); human leucocyte elastase (Navia et al., 1989); bovine chymotrypsinogen (Wang et al., 1985). All coordinates were obtained from the Protein Data Bank (Abola et al., 1987; Bernstein et al., 1977). The CCP modules IV and V were modeled from the coordinates of the homologous 5th, 15th, and 16th modules from human complement factor H (Norman et al., 1991; Barlow et al., 1992, 1993).

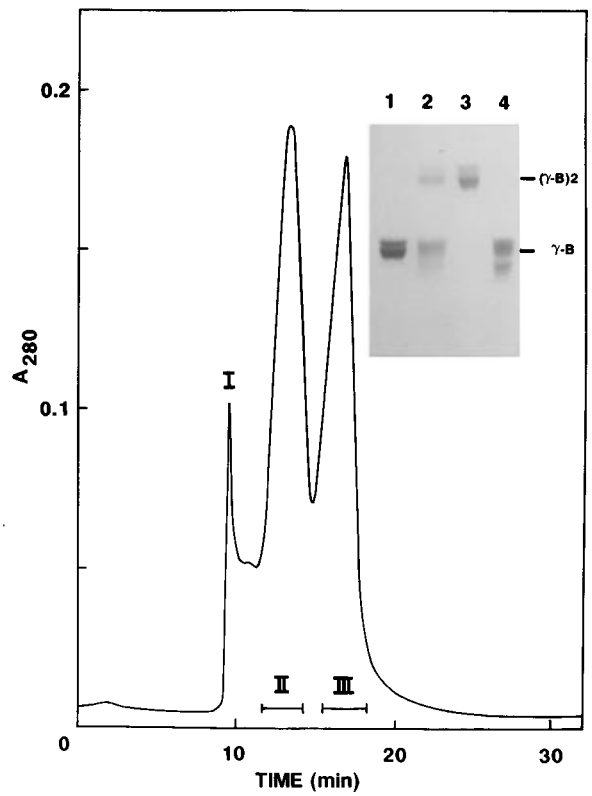

FIGURE 1: Separation of γ-B monomers and $(\gamma\text{-B})_2$ dimers by high-pressure gel permeation. Fragment $C\bar{1}r$ $(\gamma\text{-B})_2$ was cross-linked with EDC, and the reaction mixture was fractionated on two Ultraspherogel SEC 3000 columns (7.5 mm × 300 mm) equilibrated in 6 M urea/0.2 M formic acid. Proteins were detected by absorbance at 280 nm, and pools II and III were made as indicated by bars. The insert shows SDS−PAGE analysis under nonreducing conditions of (lane 1) the native $(\gamma\text{-B})_2$ fragment, (lane 2) the cross-linking mixture, (lane 3) the material contained in fraction II, and (lane 4) the material contained in fraction III.

RESULTS

Chemical Cross-Linking of Fragment $C\bar{1}r$ $(\gamma\text{-B})_2$ and Isolation of γ-B Intra and γ-B Inter Species. Human $C\bar{1}r$ was submitted to autolytic cleavage, and the resulting noncovalent $(\gamma\text{-B})_2$ dimer was isolated by ion-exchange chromatography, as described under Experimental Procedures. Analysis of the purified material by SDS−PAGE under nonreducing conditions indicated that it migrated as a diffuse band of M_r 53 000−56 000, corresponding to the monomer (Figure 1, lane 1). Reduction and alkylation yielded the B chain (M_r 34 000) and the γ fragment (M_r 18 500) (Figure 2, lane 1). N-Terminal sequence analysis of the γ fragment after electrotransfer onto a polyvinylidene difluoride membrane showed a single major sequence Gly-Trp-Lys-Leu-Arg..., indicating that autolytic proteolysis resulted in cleavage of the Arg_{279}−Gly_{280} bond, in agreement with previous findings (Arlaud et al., 1986).

Two hundred thirteen nanomoles of $C\bar{1}r$ $(\gamma\text{-B})_2$ was submitted to chemical cross-linking with EDC, and the reaction mixture was fractionated by high-pressure gel permeation under denaturing conditions, as illustrated in Figure 1. The early eluting peak I contained aggregated material. As shown by SDS−PAGE analysis, the major peak II contained molecules that remained dimeric under denaturing conditions (apparent M_r 95 000) and therefore contained at least one intermonomer cross-link. In contrast, those molecules eluting in peak III were monomerized under

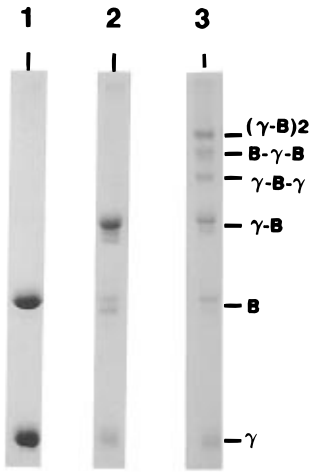

FIGURE 2: SDS−PAGE analysis of the reduced and S-pyridylethylated γ-B monomers and (γ-B)$_2$ dimers. (γ-B)$_2$ dimers and γ-B monomers obtained from pools II and III (Figure 1) were reduced and S-pyridylethylated and analyzed by SDS−PAGE. Lanes: 1, control, un-cross-linked γ-B; 2, γ-B monomers; 3, (γ-B)$_2$ dimers.

denaturing conditions and therefore contained no intermonomer cross-link. As judged from the elution profile of the reaction mixture (Figure 1), the yield of the cross-linked (γ-B)$_2$ species reached typically 50−55%.

In order to disrupt the disulfide bridge connecting γ and B, and thereby to discriminate between chemically cross-linked and disulfide-bonded γ-B molecules, (γ-B)$_2$ dimers and γ-B monomers were both reduced and then alkylated with 4-vinylpyridine. As shown by SDS−PAGE analysis, reduction of γ-B monomers yielded fragment γ (M_r 18 500), the B chain (M_r 33 000−34 000), and a cross-linked γ-B species (M_r 53 000−55 000) (Figure 2, lane 2). Reduction of (γ-B)$_2$ dimers also yielded γ, B, and a cross-linked γ-B species, as well as three higher molecular weight cross-linked species which were identified as γ-B-γ, B-γ-B, and (γ-B)$_2$ (Figure 2, lane 3). Identification of the latter three species was based on the following grounds: (i) their apparent M_r (72 000, 84 000, and 92 000, respectively) were consistent with the values predicted from the apparent M_r of γ and B (71 000, 86 500, and 105 000, respectively); (ii) N-terminal sequence analysis of the three species after electrotransfer onto a membrane showed that each contained two sequences, $_{280}$Gly-Trp-Lys-Leu-Arg... and $_{447}$Ile-Ile-Gly-Gly-Gln..., corresponding to the γ fragment and the B chain, respectively. Interestingly, for the three species, the relative yield of the former sequence was significantly lower than that of the latter (B chain/γ fragment sequence ratios for γ-B−γ, B−γ-B, and (γ-B)$_2$ = 2.5, 10.3, and 5.0, respectively), indicating that most of the γ segments had their N-termini blocked. Also, with respect to the γ sequences, it was noteworthy that the observed yield of the PTH derivative of Lys$_{282}$ was consistently low compared to the other residues (relative values 0.18−0.46), indicating that, in those molecules which did not have their N-termini blocked, the ϵ-amino group of Lys$_{282}$ was not entirely free. SDS−PAGE analysis showed no evidence for the occurrence of homologous B−B or γ−γ cross-links.

Taken together, the above data were consistent with the occurrence of cross-links only between γ and B, both inside γ-B monomers and between monomers, in agreement with previous observations (Arlaud et al., 1986). Considering that

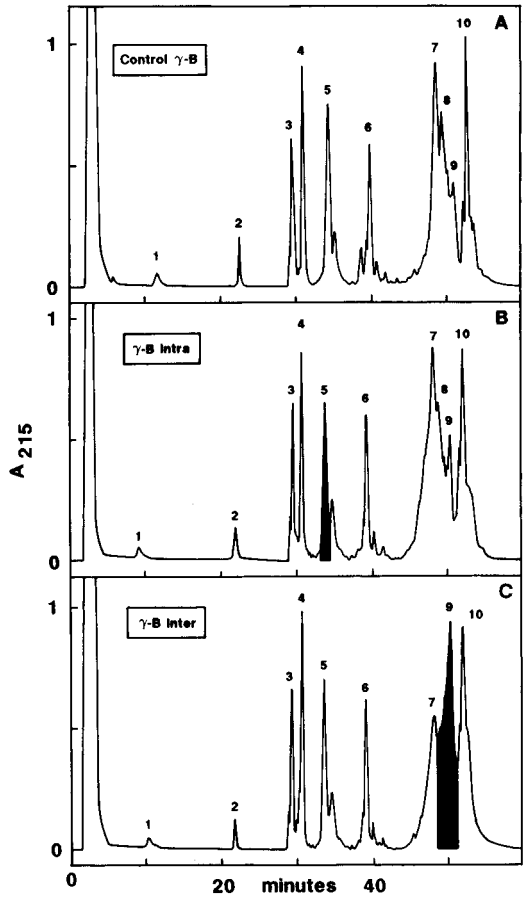

FIGURE 3: Fractionation of the CNBr-cleavage peptides from the cross-linked γ-B *intra* and γ-B *inter* species and from control γ-B. The CNBr-cleavage peptides from control γ-B (panel A), γ-B *intra* (panel B), and γ-B *inter* (panel C) were separated by reverse-phase HPLC as described under Experimental Procedures. Peptides were detected by absorbance at 215 nm. Fractions γ-B *intra*-5 and γ-B *inter*-9, containing cross-linked peptides, are shown in black.

those molecules that were monomerized under the dissociating conditions used for gel permeation had no intermonomer cross-link, it was clear that the cross-linked γ-B species yielded upon reduction of γ-B monomers only contained intramonomer γ-B cross-link(s). Conversely, given that those molecules that remained dimeric under these conditions all had at least one intermonomer cross-link, it was concluded that the cross-linked γ-B species yielded upon reduction of (γ-B)$_2$ dimers only contained intermonomer cross-links. We therefore concentrated our efforts on these two species (thereafter called γ-B *intra* and γ-B *inter*, respectively), which were purified to homogeneity by high-pressure gel permeation as described under Experimental Procedures (not shown) and submitted to CNBr cleavage, with a view to identify the cross-linking sites.

CNBr Cleavage of γ-B Intra and γ-B Inter and Isolation of Cross-Linked Peptides. The reduced and S-pyridylethylated γ-B *intra* and γ-B *inter* cross-linked samples were submitted to CNBr cleavage, and the same treatment was also applied to a reduced and S-pyridylethylated, un-cross-linked control γ-B sample. The CNBr-cleavage peptides from all γ-B samples were initially fractionated by reverse-phase HPLC as illustrated in Figure 3. From the number of methionine residues present in fragment γ and the B chain, CNBr cleavage was expected to yield 12 peptides (Figure 4). As shown in Figure 3A, the digest from the control γ-B sample was resolved into 10 major peaks, the contents of

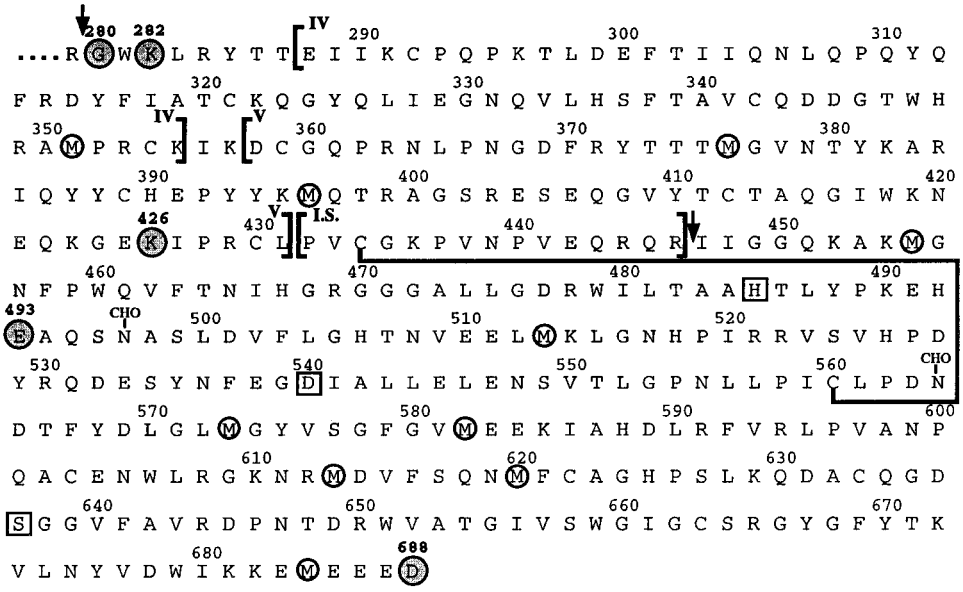

FIGURE 4: Amino acid sequence of the C1̄r γ-B fragment. The amino acid numbering is that of intact C1r (Arlaud & Gagnon, 1983; Journet & Tosi, 1986; Leytus et al., 1986; Arlaud et al., 1987b). The site of autolytic cleavage generating the γ-B fragment and the activation site of C1r defining fragment γ and the B chain are indicated by arrows. Methionine residues are circled, and active site residues are boxed. Amino acids involved in the cross-links are circled and shaded. The boundaries of CCP modules IV and V and of the intermediary segment (I.S.) are indicated as defined in this study. The only disulfide bridge shown is that connecting, through half-cystine residues 434 and 560, the γ fragment to the B chain. CHO = N-linked oligosaccharide.

Table 1: Analysis by MALDI Mass Spectrometry of the Glycopeptides from the C1̄r B Chain

peptide	position in sequence	calcd mass of peptide (Da)[a]	exptl mass (Da)	relative occurrence (%)	nature of carbohydrate	expected mass of glycopeptide (Da)
control γ-B-8	$G_{456}-M_{513}$	6388	8732 ± 5	10.8	bisialylated, fucosylated	8740
			8601 ± 5	56.8	bisialylated	8594
			8446 ± 5	10.8	monosialylated, fucosylated	8449
			8309 ± 5	21.5	monosialylated	8303
			8017 ± 5	traces	asialylated	8012
control γ-B-10	$K_{514}-M_{573}$	6904	9253 ± 5	17.8	bisialylated, fucosylated	9256
			9114 ± 5	55.0	bisialylated	9110
			8971 ± 5	7.6	monosialylated, fucosylated	8965
			8823 ± 5	19.5	monosialylated	8819

[a] Calculated mass of the homoserine form of the peptides.

which were determined by N-terminal sequence and mass spectrometry analyses, allowing identification of all peptides. Peaks 1−3 were found to contain peptides $Ile_{447}-Met_{455}$, $Asp_{614}-Met_{620}$, and $Pro_{352}-Met_{376}$, respectively. Two peptides, $Gly_{377}-Met_{396}$, and $Gly_{574}-Met_{582}$, were present in peak 4. Peaks 5 and 6 contained respectively peptides $Gln_{397}-Arg_{446}$ and $Glu_{583}-Met_{613}$. The poorly resolved peaks 7−9 each contained one large peptide, namely $Gly_{280}-Met_{351}$ (peak 7), $Gly_{456}-Met_{513}$ (peak 8), and $Phe_{621}-Met_{684}$ (peak 9), whereas another large peptide, $Lys_{514}-Met_{573}$, was present in peak 10. The short and highly charged C-terminal peptide from the B chain, $Glu_{685}-Asp_{688}$, eluted in the injection peak. Analysis by MALDI mass spectrometry of the two glycopeptides from the B chain provided precise information on their carbohydrate moieties (Table 1). Thus, both peptides $Gly_{456}-Met_{513}$ and $Lys_{514}-Met_{573}$, each containing a carbohydrate attachment site (at Asn_{497} and Asn_{564}, respectively), yielded four major species, with average mass increments of 2347 ± 5, 2211 ± 5, 2062 ± 5, and 1920 ± 5 Da, consistent with the occurrence at each site of complex-type, biantennary oligosaccharides, either bisialylated and fucosylated (calculated mass = 2352 Da), bisialylated (calculated mass = 2206 Da), monosialylated and fucosylated (calculated mass = 2061 Da), or monosialylated (calculated mass = 1915 Da). Traces of an asialylated species were also detected in the case of peptide $Gly_{456}-Met_{513}$. Interestingly, the relative ratios of the four major species were comparable in both peptides, with a predominance of the bisialylated form (Table 1). The other two large peptides each yielded a single species, with mass values of 8820 ± 5 Da for $Gly_{280}-Met_{351}$ (calculated mass = 8812 Da) and 7504 ± 4 Da for $Phe_{621}-Met_{684}$ (calculated mass = 7508 Da, including an increment of 164 Da accounting for the diisopropyl phosphate group linked to the active site Ser_{637}). All other peptides had mass values consistent with the predicted ones.

The CNBr digest from γ-B *intra* (Figure 3B) showed a pattern very similar to that of control γ-B, with no striking differences at first sight. Systematic analysis of all peaks by N-terminal sequence and mass spectrometry revealed a single difference in pool γ-B *intra*-5. First, in addition to the N-terminal sequence of the C-terminal peptide $Gln_{397}-Arg_{446}$ from γ also found in the corresponding peak from control γ-B, Edman degradation showed a second sequence Glu-Glu-Glu-(Asp) corresponding to peptide $Glu_{685}-Asp_{688}$, the C-terminal peptide from the B chain, normally eluting in the injection peak. Analysis by FAB mass spectrometry also indicated that peptide γ-B *intra*-5 had a mass of 6459 ± 1 Da, consistent with the occurrence of a cross-link between peptides $Gln_{397}-Arg_{446}$ and $Glu_{685}-Asp_{688}$ (calcu-

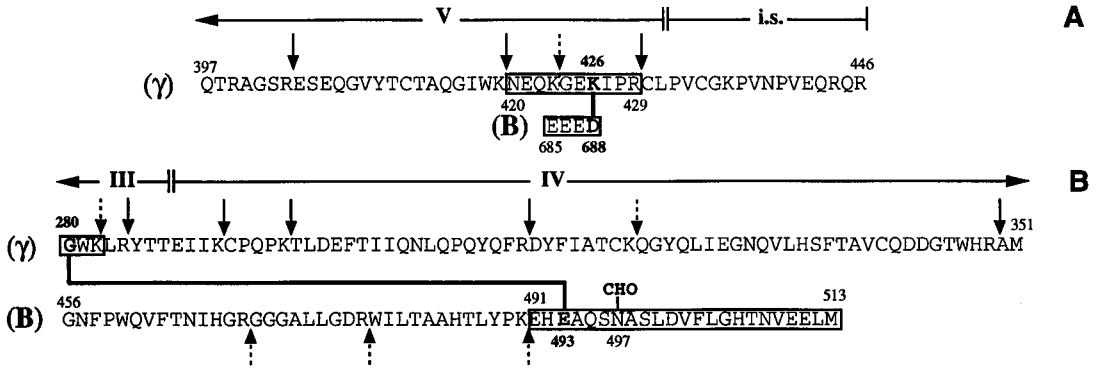

FIGURE 5: Proposed structures of the cross-linked peptides isolated from CNBr cleavage of γ-B *intra* (A) and γ-B *inter* (B). (A) Peptides Gln_{397}−Arg_{446} and Glu_{685}−Asp_{688} originate from the C-terminal ends of fragment γ and the B chain, respectively. Those portions of peptide Gln_{397}−Arg_{446} corresponding to CCP module V and the intermediary segment (i.s.) are indicated. (B) Peptides Gly_{280}−Met_{351} and Gly_{456}−Met_{513} originate from the N-terminal end of fragment γ and from the B chain, respectively. The portions of peptide Gly_{280}−Met_{351} that belong to modules CUB III and CCP IV are indicated. Sites of complete and partial cleavage by trypsin are shown by solid and dotted arrows, respectively. The cross-linked peptides isolated after tryptic cleavage are boxed, and the proposed cross-linking sites are represented by a solid line. CHO = N-linked oligosaccharide.

lated mass values = 5955 and 521 Da, respectively; expected mass of the cross-linked peptide = 6458 Da).

Reverse-phase HPLC analysis of the CNBr digest from γ-B *inter* (Figure 3C) revealed a striking difference at the level of the late eluting peaks, compared to both the control γ-B and γ-B *intra* profiles. Analysis of pools γ-B *inter*-7 and -10 by Edman degradation and mass spectrometry indicated that they contained residual amounts of peptides Gly_{280}−Met_{351} and Gly_{456}−Met_{513} and peptide Lys_{514}−Met_{573}, respectively. In contrast, mass spectrometry analysis of peak γ-B *inter*-9 revealed that, in addition to the unmodified peptide Phe_{621}−Met_{684}, it contained heterogeneous high molecular weight species centered on mass values of 17205 ± 9 and 17511 ± 9 Da, consistent with the occurrence of a cross-link between peptides Gly_{280}−Met_{351} and Gly_{456}−Met_{513} (expected mass values of the cross-linked species = 17 097, 17 243, 17 388, and 17 534 Da, depending on the nature of the oligosaccharide moiety of peptide Gly_{456}−Met_{513}; see Table 1). Further separation by high-pressure gel permeation (see Experimental Procedures) allowed isolation of the high molecular weight cross-linked species, which was found to elute well ahead of the unmodified peptides Gly_{280}−Met_{351} and Gly_{456}−Met_{513} used as standards (not shown). N-Terminal sequence analysis of the purified material showed a single major sequence, Gly-Asn-Phe-Pro-Trp..., corresponding to peptide Gly_{456}−Met_{513}, with only low amounts of the sequence corresponding to peptide Gly_{280}−Met_{351}. In light of the observation that most of the γ segments in the cross-linked species isolated from $(γ\text{-B})_2$ dimers had their N-termini blocked, and given that peptide Gly_{280}−Met_{351} originates from the N-terminal part of fragment γ, it was hypothesized that this high molecular weight species resulted from a cross-link between peptides Gly_{280}−Met_{351} and Gly_{456}−Met_{513} involving the N-terminus of the former peptide. No such high molecular weight species was detected in peaks 7−9 from γ-B *intra*. Conversely, peak γ-B *inter*-5 only contained the unmodified peptide Gln_{397}−Arg_{446} (experimental mass = 5955 ± 1 Da). It became clear therefore that the materials isolated from peaks γ-B *intra*-5 and γ-B *inter*-9 contained intra- and intermonomer cross-links, respectively. Both samples were therefore submitted to enzymic digestion with a view to identify the cross-links.

Identification of the Intramonomer Cross-Linking Site. Peptide γ-B *intra*-5, containing peptides Gln_{397}−Arg_{446} and

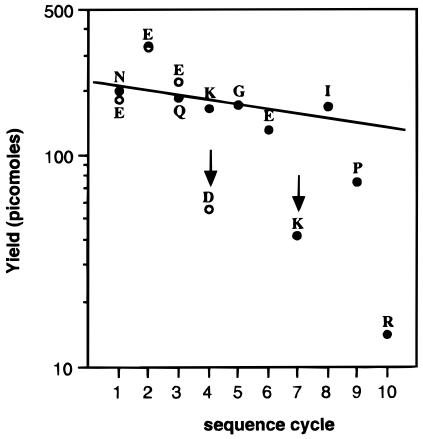

FIGURE 6: Quantitative analysis of the Edman degradation of the cross-linked peptides Asn_{420}−Arg_{429}/Glu_{685}−Asp_{688}: (●) sequence of peptide Asn_{420}−Arg_{429}; (○) sequence of peptide Glu_{685}−Asp_{688}. The PTH-Glu derivative at cycle 2 is common to both sequences. Arrows indicate residues Asp_{688} and Lys_{426} involved in the cross-link.

Glu_{685}−Asp_{688} cross-linked to each other, was cleaved with trypsin, and the resulting peptides were separated by reverse-phase HPLC (not shown) and identified by Edman degradation and mass spectrometry. As schematized in Figure 5A, complete trypsin cleavage occurred at bonds Arg_{403}−Glu_{404}, Lys_{419}−Asn_{420}, and Arg_{429}−Cys_{430}, whereas only partial cleavage was observed at Lys_{423}−Gly_{424}. Due to this partial cleavage, two cross-linked peptides were obtained, each containing segment Glu_{685}−Asp_{688}, either cross-linked to segment $_{420}$Asn-Glu-Gln-Lys-Gly-Glu-Lys-Ile-Pro-Arg_{429} or to the shorter segment $_{424}$Gly-Glu-Lys-Ile-Pro-Arg_{429}, strongly suggesting that Lys_{426} was involved in the cross-link. FAB mass spectrometry analysis of each species yielded mass values of 1700.8 ± 0.1 and 1201.4 ± 0.1 Da, respectively (expected values = 1700.8 and 1201.3 Da). As expected, Edman degradation of the larger cross-linked species yielded two sequences corresponding to peptides Asn_{420}−Arg_{429} and Glu_{685}−Asp_{688}, obtained in equivalent amounts. Quantitative analysis of the sequence data (Figure 6) revealed low recovery (20−30% of the expected yield) of the PTH derivatives of a single acidic residue (Asp_{688}) and of a single Lys residue (Lys_{426}), showing unambiguously that peptides Asn_{420}−Arg_{429} and Glu_{685}−Asp_{688} were linked to each other through a pseudopeptide bond involving the ε-amino group

Table 2: Analysis by MALDI Mass Spectrometry of the Cross-Linked Glycopeptide $Gly_{280}-Lys_{282}/Glu_{491}-Met_{513}$

peptide	calcd mass of peptide (Da)[a]	exptl mass (Da)	relative occurrence (%)	nature of carbohydrate	expected mass of glycopeptide (Da)
cross-linked peptide $G_{280}-K_{282}/E_{491}-M_{513}$	2912	5272 ± 3	8.6	bisialylated, fucosylated	5265
		5123 ± 3	48.9	bisialylated	5118
		4979 ± 3	7.9	monosialylated, fucosylated	4973
		4833 ± 3	29.5	monosialylated	4827
		4539 ± 3	5.0	asialylated	4536
control peptide $E_{491}-M_{513}$	2541	4898 ± 3	6.3	bisialylated, fucosylated	4893
		4753 ± 3	51.3	bisialylated	4747
		4607 ± 3	6.3	monosialylated, fucosylated	4601
		4461 ± 3	30.1	monosialylated	4456
		4170 ± 3	5.8	asialylated	4165

[a] Calculated mass of the homoserine form of the peptides.

of Lys_{426} and either of the two carboxyl groups of Asp_{688}.[2] It was concluded therefore that, in each γ-B monomer, Lys_{426} in CCP module V interacted through an ionic bond with the C-terminal residue Asp_{688} of the serine protease B domain.

Identification of the Intermonomer Cross-Linking Site. The high molecular weight cross-linked species isolated from pool γ-B *inter*-9 was submitted to prolonged trypsin cleavage in the presence of 1 M urea (see Experimental Procedures) in order to allow efficient proteolysis, which could not be achieved under usual cleavage conditions. As a control, the same tryptic cleavage procedure was also applied to a mixture of the unmodified peptides $Gly_{280}-Met_{351}$ and $Gly_{456}-Met_{513}$, isolated from peaks 7 and 8 from control γ-B. In both cases, the resulting tryptic peptides were separated by reverse-phase HPLC as described under Experimental Procedures (not shown). As summarized in Figure 5B, trypsin cleavage occurred as expected at arginyl and lysyl bonds in both the cross-linked material and the corresponding unmodified peptides, although it was clear that some of these bonds, e.g., $Lys_{282}-Leu_{283}$, $Lys_{322}-Gln_{323}$, $Arg_{469}-Gly_{470}$, $Arg_{478}-Trp_{479}$, and $Lys_{490}-Glu_{491}$, were cleaved less efficiently in the former case. Comparison of the chromatographic profiles revealed that the unmodified N-terminal peptide from $Gly_{280}-Met_{351}$, $Gly_{280}-Lys_{282}$, present in the control digest, was lacking in the digest from the cross-linked material. Also, the peak corresponding to the C-terminal peptide of $Gly_{456}-Met_{513}$, $Glu_{491}-Met_{513}$, was slightly retarded in the digest from the cross-linked material. MALDI mass spectrometry analysis of the unmodified peptide $Glu_{491}-Met_{513}$ isolated from the control digest (Table 2) showed a series of peaks with mass values of 4898, 4753, 4607, 4461, and 4170 Da, corresponding to the various isoforms of glycopeptide $Glu_{491}-Met_{513}$, as already determined (see Table 1). Comparative analysis of the corresponding peptide isolated from the digest of the cross-linked material (Table 2) showed a series of peaks homologous to the previous ones, but with mass values differing by an average increment of 371.4 Da, fully consistent with cross-linking of peptide $Gly_{280}-Trp-Lys_{282}$ (expected additional mass = 371.4 Da). The relative amounts of the various oligosaccharide isoforms were very close in each peptide and only differed from the ratios determined previously for peptide $Gly_{456}-Met_{513}$ (see Table 1) by increased proportions of the mono- and asialylated forms. N-Terminal sequence analysis of the cross-linked peptide (Figure 7) showed a single detectable sequence

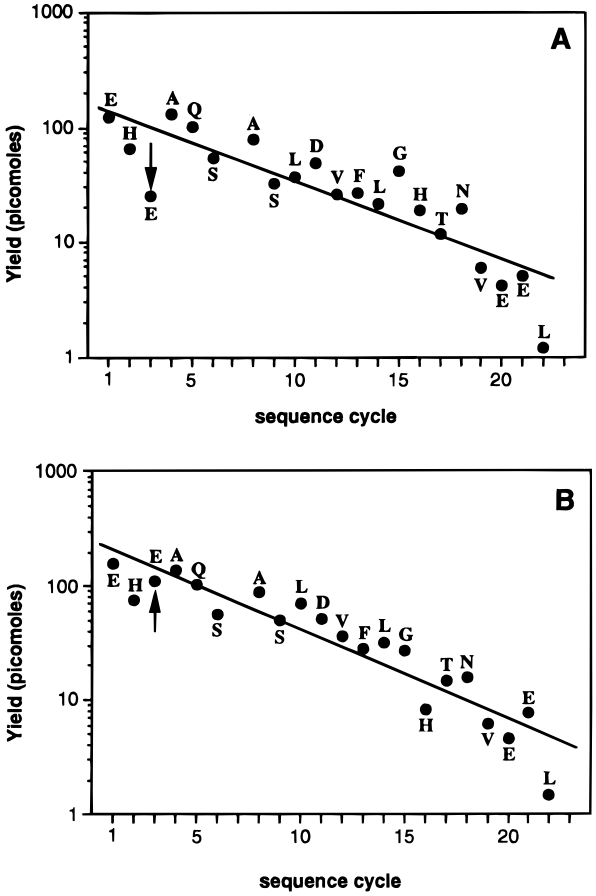

FIGURE 7: Quantitative analysis of the Edman degradation of the cross-linked peptides $Gly_{280}-Lys_{282}/Glu_{491}-Met_{513}$. (A) Sequence analysis of the cross-linked peptides. The only observed sequence is that of peptide $Glu_{491}-Met_{513}$, that of peptide $Gly_{280}-Lys_{282}$ being silent due to the involvement of its N-terminus in the cross-link. (B) Comparative sequence analysis of the unmodified peptide $Glu_{491}-Met_{513}$ isolated from tryptic cleavage of a mixture of control peptides $Gly_{280}-Met_{351}$ and $Gly_{456}-Met_{513}$. The PTH-Asn derivative expected at cycle 7 was not observed due to the presence of a carbohydrate moiety, and the PTH-Hse derivative at cycle 23 was identified but not quantitated. The arrows indicate residue Glu_{493} involved in the cross-link.

corresponding to peptide $Glu_{491}-Met_{513}$, and quantitative analysis of the sequence data revealed a low yield of the PTH derivative of a single acidic residue, Glu_{493} (approximately 25%) compared to the corresponding residue in the unmodified peptide $Glu_{491}-Met_{513}$, isolated from the control tryptic digest. In agreement with previous observations (see above), it became clear that peptides $Gly_{280}-Met_{351}$ and $Gly_{456}-Met_{513}$ were cross-linked to each other through a major pseudopeptide bond involving the N-terminus of the

[2] That significant amounts of Asp_{688} and Lys_{426} are released upon sequence analysis strongly suggests that the cross-link partly occurs through the α-carboxyl group of Asp_{688}.

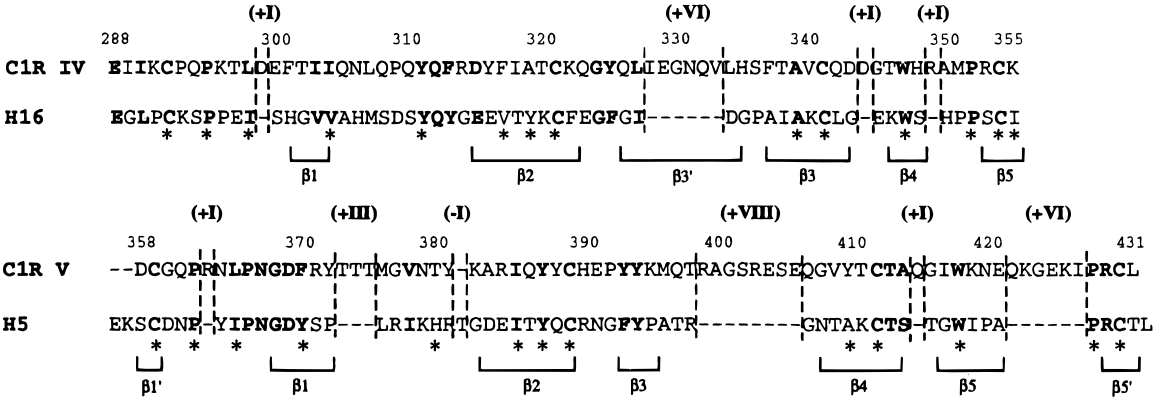

FIGURE 8: Sequence alignments used for modeling CCP modules IV and V of C1r. The amino acid numbering used is that of C1r. Identical and conservatively replaced residues are shown in bold letters. $\beta1-\beta5$ are the β-strands of the 16th and 5th CCP modules of factor H, and residues considered to form the hydrophobic core of these modules (Norman et al., 1991; Barlow et al., 1992) are indicated by asterisks. Insertions and deletions in the C1r CCP modules are shown by Roman numerals between parentheses.

former and the carboxyl group of residue Glu_{493} of the latter.[3] It was concluded, therefore, that assembly of the $(\gamma\text{-B})_2$ dimers involved formation of a major ionic bond between the N-terminal Gly_{280} of the γ-region of one monomer and Glu_{493} of the serine protease domain of the other monomer. In view of our previous observation that cross-linked molecules which did not have the N-terminus of their γ fragment(s) blocked showed low sequence yield of Lys_{282} (see above), it is likely that a minor cross-link also occurred through Lys_{282}. Indeed, this cross-link is expected to prevent trypsin cleavage at $Lys_{282}-Leu_{283}$ and thereby to yield a slightly different cross-linked tryptic peptide $Gly_{280}-Arg_{284}/Glu_{491}-Met_{513}$ (see Figure 5), likely separated by reverse-phase HPLC from the major peptide $Gly_{280}-Lys_{282}/Glu_{491}-Met_{513}$ isolated in this study. It is therefore not surprising that sequence analysis of the latter showed no detectable sequence $Gly_{280}-Trp-(Lys_{282})$....

Homology Modeling of the Protein Modules of the C1r γ-B Region. As in the case of C1s (Rossi et al., 1995), bovine chymotrypsin was used as the primary template for modeling the serine protease B domain of C1r, and the sequence alignment was similar to that defined by Greer (1990). Whereas the sequence identity between C1r and C1s amounts to 47%, that between C1r and chymotrypsin is 35% and 76% of their residues are superimposed within the limit of 1.5 Å. Compared with chymotrypsin, C1r shows four deletions at positions 466−467 (three residues), 511−512 (one residue), 584−585 (one residue), and 665−666 (two residues). There are three major insertions of eight, four, and three residues ($Glu_{491}-Ala_{498}$, $Trp_{606}-Gly_{609}$, $Ser_{534}-Asn_{536}$), a few minor insertions (Ser_{617}, $Ser_{627}-Leu_{628}$, $Asn_{647}-Thr_{648}$), and a one-residue extension at the C-terminal end of the domain, in contrast with the corresponding region of C1s, which exhibits a five-residue extension (Rossi et al., 1995). In keeping with previous observations (Fothergill et al., 1989; Perkins & Smith, 1993), the major insertions and all deletions occur in the vicinity of the active site cleft entrance, the former and latter being located on opposite sides (not shown). From a functional point of view, as already mentioned by others (Fothergill et al., 1989; Perkins & Smith, 1993), the model is consistent with a classical activation mechanism involving formation of a salt bridge between the N-terminal Ile_{447} and Asp_{636}, whereas the S1 subsite (Asp_{631}), conferring trypsin-like specificity, is located in a well-conserved region, as are the amino acids of the catalytic triad. Also, our previous hypothesis that the increase in intrinsic tryptophan fluorescence occurring upon C1r autoactivation results from the closeness of Trp residues 460 and 651 (Villiers et al., 1983) appeared plausible, in view of the proximity of these residues ($C\beta-C\beta$ distance approximately 7 Å) in the active form of the serine protease domain of C1r (see Figure 9).

The 15-residue intermediary segment ($Pro_{432}-Arg_{446}$), which is C-terminal to the γ region in activated C1r and covalently linked to the serine protease B domain through a disulfide bridge involving Cys_{434} and Cys_{560} (Figure 4), was modeled by homology with the activation peptide of bovine chymotrypsinogen, and the disulfide bond was used as a constraint to position the intermediary segment with respect to the B domain (see Figure 9), in the same way as previously described for C1s (Rossi et al., 1995). As already observed for C1s, it should be mentioned that Cys_{560} of C1r and its counterpart in chymotrypsinogen (Cys_{122}) were found to be exactly superimposed, whereas Cys_{434} of Clr was slightly displaced with respect to Cys_1 of chymotrypsinogen. However, the $Cys_{434}-Cys_{560}$ $C\alpha-C\alpha$ distance in the C1r model was compatible with the formation of a disulfide bridge.

Despite the fact that the structures of only three CCP modules are presently available, models for the CCP modules IV and V of C1r were constructed. Sequence alignments were facilitated by the use of the hydrophobic cluster analysis (HCA) method (Gaboriaud et al., 1987) and the fact that the amino acids considered to form the hydrophobic core, as defined for modules H16 and H5 (Norman et al., 1991; Barlow et al., 1992), are well conserved in C1r. These latter modules were used previously as primary templates to model CCP modules IV and V of C1r, on the basis of the alignments shown in Figure 8. Sequence identities and conservative replacements amount to 30% and 28% for the C1rIV/H16 and C1rV/H5 pairs, respectively, and the homology increases to 78% and 85%, respectively, if one only considers the core residues, clearly indicating that both pairs of modules have a common framework and thereby validating the three-dimensional models. For both C1r IV and V, the models exhibit a five-β-stranded globular structure characteristic of the CCP modules, with modifications mostly located in the

[3] The significant amount of Glu_{493} released upon sequence analysis (Figure 7A) arises mainly from a carryover from Glu_{491} and possibly also from the occurrence of a minor cross-link through Glu_{491}.

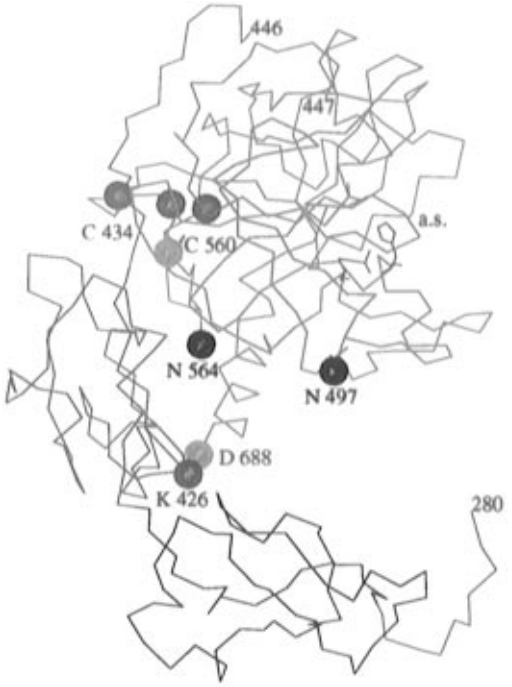

FIGURE 9: Three-dimensional Cα model of the catalytic γ-B region of C1̄r. The N-terminal extension Gly_{280}−Thr_{287}, the CCP modules IV and V, the intermediary segment (IS), and the serine protease domain (B) are shown in pink, black, red, green, and blue, respectively. The side chains of the active site amino acids (His_{485}, Asp_{540}, Ser_{637}) are shown in pink. The pink spheres represent the Cβ of Trp_{460} and Trp_{651}. The other spheres represent the Nϵ of Lys_{426}, Oδ of Asp_{688}, Cα of Cys_{434} and Cys_{560}, and Nδ of Asn_{564} and Asn_{497}.

C-terminal region of the modules. Thus, compared to H16, C1rIV shows three one-residue insertions (Asp_{299}, Asp_{344}, Arg_{349}) and a major, six-residue insertion (Ile_{328}−Val_{333}) forming a bulge within the third β-strand, exactly as previously observed in the homologous CCP module IV of C1s (Rossi et al., 1995). Compared to H5, C1rV exhibits more modifications, including a single one-residue deletion, two minor insertions (Arg_{363}, Gln_{415}), and three major insertions (Thr_{373}−Thr_{375}, Arg_{399}−Glu_{406}, Gln_{422}−Ile_{427}). Again, the latter of these insertions, extending a preexisting bulge within the fifth β-strand, was homologous to that previously described in CCP module V of C1s.

Assembly of the C1̄r γ-B Region. The assembly process was initially applied to CCP module V, the intermediary segment, and the serine protease domain. CCP module V was connected to the intermediary segment through the Leu_{431}−Pro_{432} bond and then positioned with respect to the serine protease domain by means of the intramolecular cross-link identified between residues Lys_{426}, in module V, and Asp_{688}, at the C-terminus of the serine protease domain (Figure 9). In contrast with C1s, which exhibits a five-residue extension at its C-terminal end, C1r only contains one additional residue (Asp_{688}) in this region, and therefore modeling the C-terminal 11-residue segment entirely as an α-helix introduced no uncertainty. However, some inaccuracy in the relative positioning of module V and the protease domain arose from the fact that Lys_{426} is located in the C-terminal six-residue insertion of module V. Despite these limitations, the constraints brought about by the cross-link and the Cys_{434}−Cys_{560} disulfide bond clearly indicate that module V interacts with the serine protease on the side opposite to both the active site cleft and the Arg_{446}−Ile_{447} bond cleaved upon activation (Figure 9).

As chemical cross-linking provided no information about the relative positioning of the N-terminal CCP module IV with respect to the remainder of the γ-B region, this module was fitted to the following module V on the basis of the average three-dimensional orientation defined for the H15/H16 CCP modules pair from human factor H (Barlow et al., 1993). The resulting assembly depicted on Figure 9 must be considered as tentative, as it assumes that the orthogonal angles δ and θ that describe the "tilt" of the modules are similar in both C1r and factor H. Also, this assembly should be seen as featuring a mean configuration of the pair of modules, in view of the likely occurrence in C1r of a significant range of angles of "twist" (ω), allowing rotation about the long axis of one module with respect to the other, as described in factor H (Barlow et al., 1993; Molina et al., 1995).

Assembly of the C1̄r (γ-B)$_2$ Dimer. Chemical cross-linking provided evidence that assembly of the C1̄r (γ-B)$_2$ dimer involves a major intermonomer ionic interaction between the carboxyl group of Glu_{493}, in the serine protease domain, and the free amino group of the N-terminal Gly_{280} of the γ segment. However, it is clear that the amino group of Gly_{280} is freed only upon C1̄r autolytic cleavage and therefore that Gly_{280} cannot participate in the ionic bond in the intact activated C1̄r−C1̄r dimer. In light of our observation suggesting partial involvement of Lys_{282} in the intermonomer cross-link detected in the autolytic (γ-B)$_2$ fragment (see above), the most likely hypothesis was therefore that, in intact C1̄r, the intermonomer ionic bond mainly occurs through Lys_{282} and that this bond is, for a large part, shifted upon autolytic cleavage to the newly freed amino group of Gly_{280}. With a view to further test this hypothesis, activated C1̄r was submitted to autolytic cleavage, as well as to limited proteolysis with plasmin and thermolysin. The resulting C1̄r (γ-B)$_2$ fragments obtained in each case (Arlaud et al., 1986) were then submitted to cross-linking with EDC, and the relative amounts of cross-linked (γ-B)$_2$ dimers produced were measured. Compared to the autolytic fragment, the cross-linking yield dropped to 38% and 30% for the plasmic and thermolytic fragments, respectively. Given that plasmin and thermolysin cleave C1̄r at Lys_{282}−Leu_{283} and Tyr_{285}−Thr_{286}, respectively (Arlaud et al., 1986; see Figure 4), the observed decrease was clearly correlated with the removal of Lys_{282}, providing further support to the hypothesis of a major involvement of Lys_{282} in the intermonomer ionic interaction within the intact C1̄r−C1̄r dimer. Further attempts to build a three-dimensional model of the assembly of the (γ-B)$_2$ dimer were therefore based on the hypothesis of converse salt bridges between Lys_{282} of one monomer and Glu_{493} of the other monomer.

Given that Lys_{282} is located outside CCP module V, in the preceding CUB module III (see Figure 4), a three-dimensional model of segment Gly_{280}−Ile_{289} was constructed by homology with the homologous segment $_{30}$Gly−Ala−Lys−Leu−Ser−Ala−Asp−Thr−Glu−Val_{39} from chicken triose-phosphate isomerase (Zhang et al., 1994) and then fitted to the N-terminal end of module IV, providing a model for the entire Gly_{280}−Asp_{688} C-terminal moiety of the C1̄r monomer (Figure 9). The final step consisted in fitting together two γ-B monomers by means of mutual intermonomer ionic interactions between residue Glu_{493} of the serine

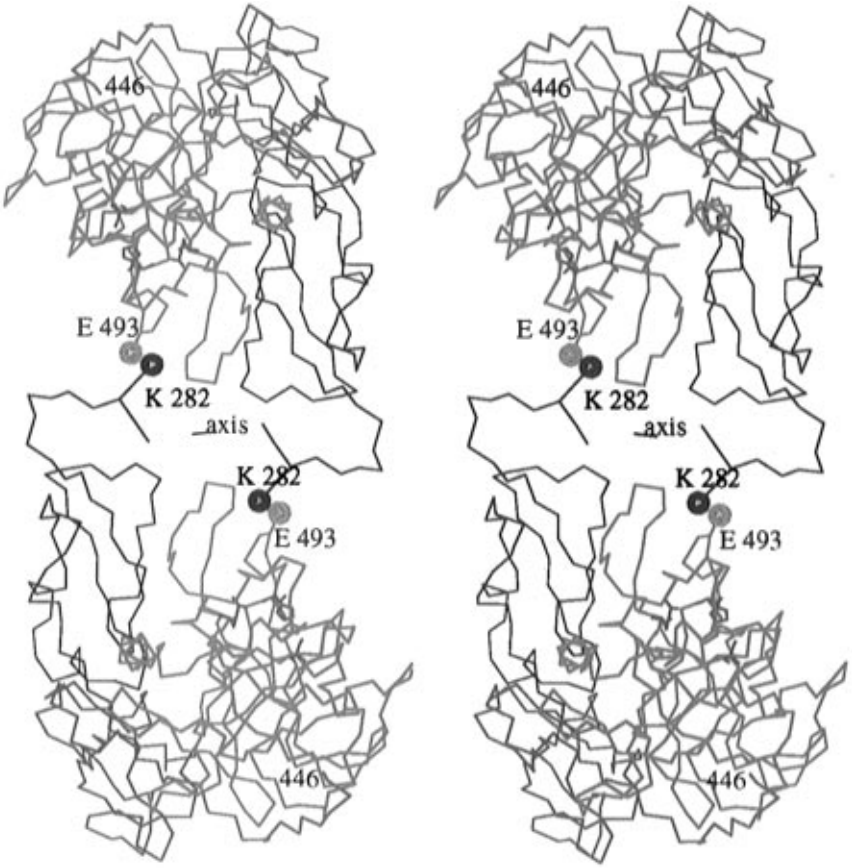

FIGURE 10: Stereoview of the three-dimensional Cα model of the assembly of the $C\bar{1}r$ (γ-B)$_2$ dimer, illustrating the head to tail configuration. The axis of symmetry is perpendicular to the plane of the figure. CCP module IV and its N-terminal extension $Gly_{280}-Thr_{287}$ are both shown in black. CCP module V, the intermediary segment, and the serine protease domain are shown in red, green, and blue, respectively. Positions Nϵ of Lys_{282} and Oδ of Glu_{493} are represented by black and blue spheres, respectively. The side chains of the active site amino acids are shown in pink.

protease domain of one monomer and residue Lys_{282} of the other monomer. The resulting tentative model (Figure 10) suffers from various uncertainties, arising in particular from the structure and orientation of the $Gly_{280}-Ile_{289}$ segment, from the fact that the relative orientation of CCP modules IV and V within each monomer is the average orientation defined for modules H15−H16, and from Glu_{493} being located within the major insertion ($Glu_{491}-Ala_{498}$) of the serine protease domain. Despite these local uncertainties, the model shows that the monomers are loosely interacting and positioned with respect to each other in a head to tail configuration, with a 2-fold symmetry about an axis parallel to the C-terminal α helix of both serine protease domains (Figure 10). Also, the active sites of both monomers clearly face opposite directions, toward the outside of the dimer.

DISCUSSION

Chemical cross-linking with EDC was used to gain information on the structure and assembly of the dimeric (γ-B)$_2$ catalytic region of human $C\bar{1}r$. Our previous study by the same approach performed on the homologous protease $C\bar{1}s$ proved somewhat difficult, due to the occurrence of various side reactions, such as formylations and "abortive" reactions of EDC with carboxyl groups (Rossi et al., 1995). Comparatively, the present analysis performed on $C\bar{1}r$ showed no particular problem with respect to the identification of the two cross-links through the combined use of N-terminal sequence and mass spectrometry analyses. Most of the experiments reported in this study were performed several times, and identical results were obtained in each case. However, compared to $C\bar{1}s$ γ-B, which is a monomer, analysis of the data obtained on $C\bar{1}r$ (γ-B)$_2$ was more complex, due to the necessity to discriminate intra- and intermonomer cross-links. This was achieved by using a two-step strategy involving (i) separation of cross-linked dimers, all containing at least one intermonomer cross-link, from residual monomers, containing no intermonomer cross-link and (ii) isolation, from these species, of cross-linked γ-B molecules containing only inter- and intramonomer cross-link(s), respectively. The fact that the cross-link identified as intramonomer in the present study is homologous to that previously shown to occur in $C\bar{1}s$ γ-B (see further discussion below) validates the strategy employed and thereby the data obtained in the present study. As judged from the separation of dimeric and monomeric molecules by gel filtration after cross-linking, the overall reaction yield was consistently higher than 50%. This value should be relativized, however, considering that cross-linked dimers represent in fact a heterogeneous family of molecules bearing various numbers and combinations of inter- and intramonomer cross-links (see Figure 2). With respect to the location of the residues involved in the cross-links identified in this study, it is noteworthy that three of them (Gly_{280}, Lys_{426}, Asp_{688}) occur in the N- or C-terminal parts of the γ fragment and B chain (see Figure 5), the fourth one (Glu_{493}) lying within a surface-exposed loop also bearing an N-linked oligosaccharide. This suggests that, due to its water solubility, EDC only reacts with well-exposed amino acids and

therefore that other ionic bonds located in more hydrophobic environments are not likely to be detected by this reagent. Also, it should be emphasized that most of the residues involved in the ionic interactions appear as specific of C1r, two of them (Lys_{426}, Glu_{493}) being located within major sequence insertions, and Asp_{688} representing the C-terminal extension of the serine protease domains of $C\bar{1}r$. It should be mentioned that intramolecular cross-links, not identified in this study, very likely occurred within individual modules, as suggested, e.g., by the duplication of the band corresponding to the B chain after cross-linking (see Figure 2).

A three-dimensional homology model of the serine protease B domain of $C\bar{1}r$ was built using bovine chymotrypsin as a primary template, according to the procedure defined by Greer (1990), previously applied to complement proteases by Perkins and Smith (1993). Indeed, our alignment is very similar to that reported by Perkins and Smith (1993). In contrast, compared to that of Carter et al. (1984), our model shows major differences with respect to the lengths and locations of some of the insertions and deletions. In particular, whereas the major insertion ($Glu_{491}-Ala_{498}$) is comparable in both models, two extra insertions reported by Carter et al. do not appear in our model. These differences likely arise from the fact that these authors based their approach solely on a linear sequence alignment with chymotrypsin, whereas the method designed by Greer, taking into account a series of proteases, is based on the superimposition of three-dimensional structures.

Our data also provide, by means of mass spectrometry analyses of the corresponding glycopeptides, the first precise information on the structure of the carbohydrates attached to Asn residues 497 and 564 of the serine protease domain of $C\bar{1}r$, indicating that each position is occupied by heterogeneous complex-type biantennary species bearing either a single or two terminal sialic acid residues and one or no fucose residue, with a major species (NeuAc2 Gal2 GlcNAc4 Man3) in both cases. This information is fully consistent with previous carbohydrate composition analyses (Arlaud & Gagnon, 1983) and in agreement with identification of carbohydrates in the homologous protease $C\bar{1}s$ (Pétillot et al., 1995), also indicating the presence of complex-type oligosaccharides. As judged from mass analyses of the glycopeptides, the total mass of the carbohydrate moiety of $C\bar{1}r$ γ-B can be estimated at 4120 Da. It should be mentioned that all other mass analyses performed on the CNBr-cleavage peptides from $C\bar{1}r$ γ-B were found consistent with sequence data, indicating that glycosylation is the only posttranslational modification occurring in this region.

As previously observed in the case of C1s (Rossi et al., 1995), CCP modules IV and V of C1r were found to exhibit closer resemblance with modules 16 and 5 of human factor H, respectively, with respect to both the structural cores and the overall shapes of the modules. The three-dimensional homology models derived for modules IV and V of C1r mainly differ from their respective templates by one and two major insertions, all located in the C-terminal moiety of the modules, and may therefore be considered reliable for the major part of their structure.

The data provided by our cross-linking and modeling studies, along with available structural information on the conformation of a pair of CCP modules (Barlow et al., 1993), have allowed us to construct a three-dimensional model of the assembly of CCP modules IV and V, the intermediary segment, and the serine protease domain, within the $C\bar{1}r$ γ-B region. With respect to the assembly of the latter entities, the model exhibits a striking resemblance with the corresponding part of $C\bar{1}s$ (Rossi et al., 1995). Indeed, as also observed in $C\bar{1}s$, CCP module V and the intermediary segment of $C\bar{1}r$ are associated with the serine protease in such a way that module V is closely interacting with the protease on the side opposite to both the active site and the $Arg_{446}-Ile_{447}$ bond cleaved upon C1r activation. As in the case of C1s, this configuration allows free access of the active site of each C1r monomer to its substrates, i.e., first, the activation site of the neighboring C1r monomer and, then, the $Arg_{422}-Ile_{423}$ bond cleaved upon C1s activation by $C\bar{1}r$. The two intramonomer constraints used to position the C1r modules, namely, the $Cys_{434}-Cys_{560}$ disulfide bond connecting the intermediary segment to the protease domain and the ionic bond between Lys_{426} of module V and the C-terminal Asp_{688} of the protease, are also present in C1s, at homologous positions ($Cys_{410}-Cys_{534}$ and $Lys_{405}-Glu_{672}$) (Rossi et al., 1995). Indeed, detailed sequence comparison of human C1r and C1s with hamster C1s (Kinoshita et al., 1989) and human mannan binding lectin-associated serine protease (MASP), a modular protein belonging to the C1r/C1s family (Takayama et al., 1994), reveals that the latter two proteins also contain both the homologous disulfide bond and the appropriately positioned basic and acidic residues to form the above-mentioned ionic bond (not shown). Thus, it may be predicted that the CCP module V and serine protease domain of hamster C1s and human MASP are connected through ionic interactions involving residues $Lys_{406}-Asp_{674}$, and $Arg_{430}-Asn_{699}$, respectively. On the basis of the above findings, we propose that C1r, C1s, and MASP constitute a family of structurally related proteins, not only from the sequence point of view but also with respect to their three-dimensional modular assembly.

Chemical cross-linking of the $C\bar{1}r$ $(\gamma\text{-B})_2$ dimer provided evidence for the occurrence of a major ionic interaction between the N-terminal Gly_{280} of one monomer and Asp_{493} of the serine protease domain of the other monomer. However, as discussed in the Results section, it is highly likely that involvement of the free amino group of Gly_{280} results from a local rearrangement in the N-terminal part of the γ region occurring upon $C\bar{1}r$ autolytic cleavage and that the intermonomer ionic bond mainly occurs through the ϵ-amino group of the neighboring Lys_{282} in the intact $C\bar{1}r$ molecule. A model of the whole γ-B monomer, comprising residues 282−688 of $C\bar{1}r$, was constructed, and then assembly of the $(\gamma\text{-B})_2$ dimer was realized on the basis of the above hypothesis. The resulting model (Figure 10) features a loose assembly of the dimer, with the two monomers virtually interacting only through the ionic bonds and with empty spaces between monomers, as well as between the serine protease domain and CCP module IV of each monomer. The latter space may be filled in part by the carbohydrate linked to Asn_{497}, whereas that linked to Asn_{564} lies at the interface between the serine protease and CCP module V (see Figure 9). Such a location of the carbohydrates appears consistent with the observation that both are resistent to cleavage by peptide:*N*-glycosidase F (Aude et al., 1988). A loose assembly of the dimer is fully consistent with electron microscopy pictures, which surprisingly show a virtually empty interface between the monomers (Villiers et al., 1985; Weiss et al., 1986). According to our model,

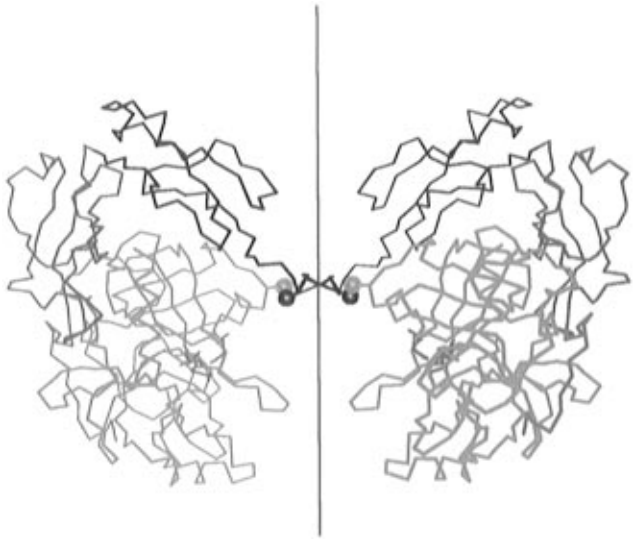

FIGURE 11: Three-dimensional Cα model of the assembly of the $C\bar{1}r$ (γ-B)$_2$ dimer illustrating the butterfly-like shape. The axis of symmetry lies within the plane of the figure. Colors and symbols are the same as in Figure 10. The relative thickness of the peptide chain is a function of its position with respect to the plane of the figure (the thicker the closer to the viewer).

the distance between the centers of gravity of the monomers is 62 Å, which compares well with the value of about 70 Å estimated from electron microscopy pictures (Weiss et al., 1986). Also, our model is consistent with neutron diffraction studies (Zaccaï et al., 1990), which concluded to a "loose packing of the γ-B subunits". The neutron scattering amplitude radius of gyration calculated from the atomic coordinates of our model (not taking into account the carbohydrates) is 38.5 Å, whereas a value of 34 ± 1 Å was determined experimentally (Zaccaï et al., 1990). These two values may be considered reasonably consistent in view of the multidomain structure and flexibility of the assembly and considering the uncertainties inherent to the model. Indeed, the difference may suggest a slightly less loose packing of the monomers than featured in the model. Comparison with the hydrodynamic sphere models described by Perkins and Nealis (1989) shows that the overall length of our γ-B model (70 Å) is significantly less than the value of 90 Å determined for their sphere model featuring a more extended structure. As for the (γ-B)$_2$ dimer, our head to tail assembly has an overall length of 110 Å, identical to that predicted by Perkins and Nealis (1989) for their X-shaped sphere model featuring a "head to head" association of the monomers.

A direct consequence of the intermonomer salt bridge identified in the present study is that the two monomers interact in a head to tail configuration, with a 2-fold symmetry (Figure 10). This type of assembly is in keeping with earlier data (Arlaud et al., 1986) but disagrees with the X-shaped C1r model (Weiss et al., 1986; Perkins & Nealis, 1989) based initially on electron microscopy studies and featuring a head to head interaction of the γ-B monomers. In this respect, it is noteworthy that certain views of our model, such as that shown on Figure 11, give an overall impression of a head to head, rather than head to tail, assembly. Indeed, such views appear very similar to most of the pictures seen by electron microscopy after negative staining (Weiss et al., 1986), some of them showing molecules with a "butterfly-like" shape, comparable to that of Figure 11. It is clear, therefore, that there is no incompatibility between our head to tail model and the structures seen by electron microscopy, the latter possibly resulting from a preferential adsorption of the protein onto the grid.

From a functional point of view, the head to tail configuration places the active site of each monomer well apart from the Arg_{446}−Ile_{447} activation site of the other monomer, with a Cα−Cα distance of 77 Å between Ser_{637} and Arg_{446}. Also, the two active site clefts clearly face the outside of the dimer and point toward opposite directions (Figure 10). Although the above value is likely incorrect due to the uncertainties inherent in the assembly, obviously the model is valid with respect to the overall location and relative orientation of the active sites, which are a direct consequence of the symmetry of the assembly. As expected from the fact that cross-linking was performed on the activated $C\bar{1}r$ (γ-B)$_2$ fragment, it is clear, therefore, that the assembly depicted in Figure 10 corresponds to the activated state of the molecule, i.e., a configuration in which the $C\bar{1}r$ active sites have access to the Arg_{422}−Ile_{423} activation sites in the proenzyme C1s γ-B regions, which, according to current C1 models (Weiss et al., 1986; Schumaker et al., 1986; Arlaud et al., 1987a; Perkins, 1989), are supposed to interact with opposite sides of the $C\bar{1}r$ (γ-B)$_2$ dimer.

In contrast, the proenzyme state of the C1r (γ-B)$_2$ dimer should allow cross-activation of the monomers, which requires a configuration in which the activation site of each monomer fits into the pro site of the other monomer. Obviously, this implies that a dramatic conformational change must take place between the proenzyme and active states. Such a change may occur, at least in part, through a rotation about its long axis of one CCP module with respect to the other, considering the range of angles of twist (ω = 131 ± 46°) described for the H15−H16 pair of CCP modules (Barlow et al., 1993; Molina et al., 1995). However, it is clear that the constraints brought about by the intermonomer ionic bonds lock the dimer in such a way that precludes this kind of conformational change. The most likely hypothesis is therefore that these bonds only form in the active state, as a result of activation-induced local conformational changes. Indeed, this hypothesis appears consistent with the observation that the stability of the C1r−C1r dimer with respect to acidic pH increases upon activation (Arlaud et al., 1980).

Despite the local uncertainties of our structural model of the $C\bar{1}r$ (γ-B)$_2$ dimer, we believe that it represents a further step toward elucidation of the sophisticated mechanisms involved in C1 activation and provides a reliable basis for further studies at the atomic level through the use of X-ray crystallography and site-directed mutagenesis.

ACKNOWLEDGMENT

We are indebted to Pr. I. D. Campbell and Dr. P. N. Barlow, who kindly sent us the coordinates of the 5th, 15th, and 16th CCP modules of human factor H. We thank Y. Pétillot for mass spectrometry measurements, J.-F. Hernandez for help with HPLC analysis, G. Zaccaï for calculation of the radius of gyration, and J. C. Fontecilla-Camps for helpful discussions.

REFERENCES

Abola, E. E., Bernstein, F. C. Bryant, S. H., Koetzle, T. F., & Weng, J. (1987) in *Crystallographic Databases—Information Content, Software Systems, Scientific Applications* (Allen, F. H., Berger-

hoff, G., & Sievers, R., Eds.) pp 107−132, Data Commission of the International Union of Crystallography, Bonn, Cambridge, and Chester.
Arlaud, G. J., & Gagnon, J. (1983) *Biochemistry 22*, 1758−1764.
Arlaud, G. J., Sim, R. B., Duplaa, A.-M., & Colomb, M. G. (1979) *Mol. Immunol. 16*, 445−450.
Arlaud, G. J., Chesne, S., Villiers, C. L., & Colomb, M. G. (1980) *Biochim. Biophys. Acta 616*, 105−115.
Arlaud, G. J., Colomb, M. G., & Villiers, C. L. (1985) *Biosci. Rep. 5*, 831−837.
Arlaud, G. J., Gagnon, J., Villiers, C. L., & Colomb, M. G. (1986) *Biochemistry 25*, 5177−5182.
Arlaud, G. J., Colomb, M. G., & Gagnon, J. (1987a) *Immunol. Today 8*, 106−111.
Arlaud, G. J., Willis, A. C., & Gagnon, J. (1987b) *Biochem. J. 241*, 711−720.
Aude, C. A., Lacroix, M. B., Arlaud, G. J., Gagnon, J., & Colomb, M. G. (1988) *Biochemistry 27*, 8641−8648.
Barlow, P. N., Steinkasserer, A., Horne, T. J., Pearce, J., Driscoll, P. C., Sim, R. B., & Campbell, I. D. (1992) *Biochemistry 31*, 3626−3634.
Barlow, P. N., Steinkasserer, A., Norman, D. G., Kieffer, B., Wiles, A. P., Sim, R. B., & Campbell, I. D. (1993) *J. Mol. Biol. 232*, 268−284.
Bartunik, H. D., Summers, L. J., & Bartsch, H. H. (1989) *J. Mol. Biol. 210*, 813−828.
Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, E. F., Jr., Brice, M. D., Rodgers, J. R., Kennard, O., Shomanouchi, T., & Tasumi, M. (1977) *J. Mol. Biol. 112*, 535−542.
Birktoft, J. J., & Blow, D. M. (1972) *J. Mol. Biol. 68*, 187−240.
Bode, W., Chen, Z., Bartels, K., Kutzbach, C., Schmidt, G., & Bartunik, H. (1983) *J. Mol. Biol. 164*, 237−282.
Bork, P., & Bairoch, A. (1995) *Trends Biochem. Sci. 20* (Suppl. March), C03.
Carter, P. E., Dunbar, B., & Fothergill, J. E. (1984) *Philos. Trans. R. Soc. London, B 306*, 293−299.
Cooper, N. R. (1985) *Adv. Immunol. 37*, 151−216.
Fothergill, J., Kemp, G., Paton, N., Carter, P., & Gray, P. (1989) *Behring Inst. Mitt. 84*, 72−79.
Fujinaga, M., & James, M. N. G. (1987) *J. Mol. Biol. 195*, 373−396.
Gaboriaud, C., Bissery, V., Benchetrit, T., & Mornon, J. P. (1987) *FEBS Lett. 224*, 149−155.
Greer, J. (1990) *Proteins 7*, 317−334.
Jones, T. A., Zou, J. Y., Lowan, S. W., & Kjeldgaard, M. (1991) *Acta Crystallogr., Sect. A 47*, 110−119.
Journet, A., & Tosi, M. (1986) *Biochem. J. 240*, 783−787.
Kinoshita, H., Sakiyama, H., Tokunaga, K., Imajoh-Ohmi, S., Hamada, Y., Isono, K., & Sakiyama, S. (1989) *FEBS Lett. 250*, 411−415.
Lacroix, M. B., Aude, C. A., Arlaud, G. J., & Colomb, M. G. (1989) *Biochem. J. 257*, 885−891.
Laemmli, U. K. (1970) *Nature 227*, 680−685.
Leytus, S. P., Kurachi, K., Sakariassen, K. S., & Davie, E. W. (1986) *Biochemistry 25*, 4855−4863.
Means, G. E., & Feeney, R. E. (1971) in *Chemical Modification of Proteins*, pp 144−148, Holden-Day, Inc., San Francisco.
Medved, L. V., Busby, T. F., & Ingham, K. C. (1989) *Biochemistry 28*, 5408−5414.
Molina, H., Perkins, S. J., Guthridge, J., Gorka, J., Kinoshita, T., & Holers, V. M. (1995) *J. Immunol. 154*, 5426−5435.
Navia, M. A., Mc Keever, B. M., Springer, J. P., Lin, T.-Y., Williams, H. R., Fluder, E. M., Dorn, C. P., & Hoogsteen, K. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 7−11.
Norman, D. G., Barlow, P. N., Baron, M., Day, A. J., Sim, R. B., & Campbell, I. D. (1991) *J. Mol. Biol. 219*, 717−725.
Perkins, S. J. (1989) *Behring Inst. Mitt. 84*, 129−141.
Perkins, S. J., & Nealis, A. S. (1989) *Biochem. J. 263*, 463−469.
Perkins, S. J., & Smith, K. F. (1993) *Biochem. J. 293*, 109−114.
Pétillot, Y., Thibault, P., Thielens, N. M., Rossi, V., Lacroix, M., Coddeville, B., Spik, G., Schumaker, V. N., Gagnon, J., & Arlaud, G. J. (1995) *FEBS Lett. 358*, 323−328.
Read, R. J., & James, M. N. G. (1988) *J. Mol. Biol. 200*, 523−551.
Remington, S. J., Woodbury, R. G., Reynolds, R. A., Matthews, B. W., & Neurath, H. (1988) *Biochemistry 27*, 8097−8105.
Rossi, V., Gaboriaud, C., Lacroix, M., Ulrich, J., Fontecilla-Camps, J. C., Gagnon, J., & Arlaud, G. J. (1995) *Biochemistry 34*, 7311−7321.
Sawyer, L., Shotton, D. M., Campbell, J. W., Wendell, P. L., Muirhead, H., Watson, H. C., Diamond, R., & Ladner, R. C. (1978) *J. Mol. Biol. 118*, 137−208.
Schumaker, V. N., Zavodszky, P., & Poon, P. H. (1987) *Annu. Rev. Immunol. 5*, 21−42.
Takayama, Y., Takada, F., Takahashi, A., & Kawakami, M. (1994) *J. Immunol. 152*, 2308−2316.
Thielens, N. M., Aude, C. A., Lacroix, M. B., Gagnon, J., & Arlaud, G. J. (1990a) *J. Biol. Chem. 265*, 14469−14475.
Thielens, N. M., Van Dorsselaer, A., Gagnon, J., & Arlaud, G. J. (1990b) *Biochemistry 29*, 3570−3578.
Villiers, C. L., Arlaud, G. J., & Colomb, M. G. (1983) *Biochem. J. 215*, 369−375.
Villiers, C. L., Arlaud, G. J., & Colomb, M. G. (1985) *Proc. Natl. Acad. Sci. U.S.A. 82*, 4477−4481.
Volanakis, J. E., & Arlaud, G. J. (1997) in *Human Complement System in Health and Disease* (Frank, M., & Volanakis, J. E., Eds.) Marcel Dekker (in press).
Wang, D., Bode, W., & Huber, R. (1985) *J. Mol. Biol. 185*, 595−624.
Weiss, V., Fauser, C., & Engel, J. (1986) *J. Mol. Biol. 189*, 573−581.
Zaccaï, G., Aude, C. A., Thielens, N. M., & Arlaud, G. J. (1990) *FEBS Lett. 269*, 19−22.
Zhang, Z., Sugio, S., Komives, E. A., Liu, K. D., Knowles, J. R., Petsko, G. A., & Ringe, D. (1994) *Biochemistry 33*, 2830−2837.

BI962719I