Ciechanover, A., Heller, H., Elias, S., Haas, A., & Hershko, A. (1980) *Proc. Natl. Acad. Sci. U.S.A. 77*, 1365-1368.
Cox, M. J., Shapira, R., & Wilkinson, K. D. (1986) *Anal. Biochem. 154*, 345-352.
Duerksen, P. J., & Wilkinson, K. D. (1987) *Anal. Biochem. 160*, 444-454.
Evans, A. C., Jr., & Wilkinson, K. D. (1985) *Biochemistry 24*, 2915-2923.
Finley, D., Ozkaynak, E., & Varshavsky, A. (1987) *Cell (Cambridge, Mass.) 48*, 1035-1046.
Goldknopf, I. A., Sudhakar, W., Rosenbaum, F., & Busch, H. (1980) *Biochem. Biophys. Res. Commun. 95*, 1253-1260.
Haas, A. L., & Bright, P. M. (1985) *J. Biol. Chem. 260*, 12464-12473.
Haas, A. L., & Wilkinson, K. D. (1985) *Prep. Biochem. 15*, 49-60.
Haas, A. L., Warms, J. V. B., Hershko, A., & Rose, I. A. (1982) *J. Biol. Chem. 257*, 2543-2548.
Hershko, A., & Rose, I. A. (1987) *Proc. Natl. Acad. Sci. U.S.A. 84*, 1829-1833.
Hershko, A., Ciechanover, A., Heller, H., Haas, A. L., & Rose, I. A. (1980) *Proc. Natl. Acad. Sci. U.S.A. 77*, 1783-1786.
Hershko, A., Heller, H., Elias, S., & Ciechanover, A. (1983) *J. Biol. Chem. 258*, 8206-8214.
Hershko, A., Heller, H., Eytan, E., Kaklij, G., & Rose, I. A. (1984) *Proc. Natl. Acad. Sci. U.S.A. 81*, 7021-7025.
Hough, R., Pratt, G., & Rechsteiner, M. (1986) *J. Biol. Chem. 261*, 2400-2408.
Kanda, F., Sykes, D. E., Yasuda, H., Sandberg, A., & Matsui, S. (1986) *Biochim. Biophys. Acta 870*, 64-75.
Klausner, Y. S., & Bodanszky, M. (1974) *Synthesis 1974*, 549-555.
Laemmli, U. K. (1970) *Nature (London) 227*, 680-685.
Lowry, O. H., Rosenbrough, N. J., Farr, A. L., & Randall, R. J. (1951) *J. Biol. Chem. 226*, 265-275.
Lund, P. K., Moats-Staats, B. M., Simmons, J. G., Hoyt, E., D'Ercole, A. J., Martin, F., & VanWyk, J. J. (1985) *J. Biol. Chem. 260*, 7609-7613.
Matsui, S., Seon, B. K., & Sandberg, A. A. (1979) *Proc. Natl. Acad. Sci. U.S.A. 76*, 6386-6390.
Matsui, S., Sandberg, A. A., Negoro, S., Seon, B. K., & Goldstein, G. (1982) *Proc. Natl. Acad. Sci. U.S.A. 79*, 1535-1539.
Mueller, R. D., Yasuda, H., Hatch, C. L., Bonner, W. M., & Bradbury, E. M. (1985) *J. Biol. Chem. 260*, 5147-5153.
Ozkaynak, E., Finley, D., Solomon, M. J., & Varshavsky, A. (1987) *EMBO J. 6*, 1429-1439.
Pickart, C. M., & Rose, I. A. (1985) *J. Biol. Chem. 260*, 7903-7910.
Pickart, C. M., & Rose, I. A. (1986) *J. Biol. Chem. 261*, 10210-10217.
Wilkinson, K. D., Urban, M. K., & Haas, A. L. (1980) *J. Biol. Chem. 255*, 7529-7532.
Wilkinson, K. D., Cox, M. J., Mayer, A. N., & Frey, T. (1986a) *Biochemistry 25*, 6644-6649.
Wilkinson, K. D., Cox, M. J., O'Connor, L. B., & Shapira, R. (1986b) *Biochemistry 25*, 4999-5004.

Tertiary Structure of Human Complement Component C5a in Solution from Nuclear Magnetic Resonance Data

Erik R. P. Zuiderweg,*,‡ David G. Nettesheim,‡ Karl W. Mollison,§ and George W. Carter§

Research NMR Group and Immunosciences Area of the Pharmaceutical Discovery Division, Abbott Laboratories, Abbott Park, Illinois 60064

Received June 8, 1988; Revised Manuscript Received August 12, 1988

ABSTRACT: The tertiary structure for the region 1-63 of the 74 amino acid human complement protein C5a in solution was calculated from a large number of distance constraints derived from nuclear Overhauser effects with an angular distance geometry algorithm. The protein consists of four helices juxtaposed in an approximately antiparallel topology connected by peptide loops located at the surface of the molecule. The structures obtained for the helices are compatible with α-helical hydrogen-bonding patterns, which provides an explanation for the observed slow solvent exchange kinetics of the amide protons in these peptide regions. In contrast to the peptide region 1-63, no defined structure could be assigned to the C-terminal region 64-74, which increasingly acquires dynamic random coil characteristics as the end of the peptide chain is approached. An average root-mean-square deviation of 1.6 Å was obtained for the α-carbons of the first 63 residues in the calculated ensemble of C5a structures, while the α-helices were determined with an average root-mean-square deviation of 0.8 Å for the α-carbons. A comparison between the solution structure of C5a and the crystal structure of the functionally related C3a protein, as well as inferences for the interaction of C5a with its receptor on polymorphonuclear leukocytes, is discussed.

The complement system consists of a set of regulatory factors and proteolytic enzymes (C1-C9) which aid in the recognition and elimination of foreign substances. In classical activation, a cascade of proteolytic cleavages of complement proteins C1-C5 is triggered by the formation of immune complexes. This ultimately results in the formation of a molecular lytic complex between C5b, C6, C7, C8, and C9 which assists in the lysis of bacterial cells [reviewed in Mayer (1979)]. The

* To whom correspondence should be addressed.
‡ Research NMR Group.
§ Immunosciences Area.

0006-2960/89/0428-0172$01.50/0

byproducts of the cleavages of C3, C4, and C5 are known as anaphylatoxins (C3a, C4a, and C5a, respectively) and facilitate the immune response by causing vasodilation, increasing vascular permeability, inducing the contraction of smooth muscle, and triggering the release of histamine from mast cells and basophilic leukocytes. Among the anaphylatoxins, the 74 amino acid C5a protein is the principal inflammatory molecule derived from the complement system. In addition to the biological actions common to all anaphylatoxins, this potent mediator stimulates the locomotion (chemokinesis) and recruitment (chemotaxis) of polymorphonuclear leukocytes to sites of inflammation and triggers these cells to release tissue-digesting enzymes and other damaging substances [for a review, see Hugli (1981)]. C5a has also been shown to augment antibody production in vitro by inducing release of monokines (Weigle et al., 1983). Because of this broad range of biological properties, C5a has been implicated as a causative or aggravating agent in a variety of inflammatory and allergic diseases (Ward, 1970). Compounds which inhibit the inflammatory actions of C5a would therefore be powerful therapeutic agents for the treatment of such diseases.

The interaction of C5a with polymorphonuclear leukocyte receptors has been studied by several groups. As yet, however, the understanding of the C5a regions which interact with the receptor is incomplete. The C-terminal residues (Fernandez et al., 1978; Gerard et al., 1979; Chenoweth & Hugli 1978; Edalji et al., 1987) and N-terminal residues (Damerau et al., 1983; Gerard et al., 1985), as well as the disulfide linked core (Johnson et al., 1987; Johnson & Chenoweth, 1985), have all been implicated to play an important role in binding to the C5a receptor. Site-directed mutagenis studies have also suggested involvement of both core and C-terminal residues (Mollison et al., 1988a). The interaction of C5a with its receptor appears to be far more complicated than for C3a, where synthetic peptides of the 21 C-terminal residues have been shown to elicit a full biological response (Lu et al., 1984); only negligible affinity was obtained for similar C-terminal peptides of C5a (Chenoweth et al., 1979; Khan et al., 1985).

Compared to the number of studies of the biological functions of C5a, relatively few experiments have been carried out to investigate its structure. The sequence of C5a has been determined (Fernandez & Hugli, 1978), the secondary structure was characterized as predominantly α-helical from optical studies (Hugli, 1981), and the disulfide linkages between six of the seven cysteine residues have been obtained from biochemical studies (Zimmerman & Vogt, 1984). No detailed experimental three-dimensional structure of C5a, e.g., from X-ray crystallography, has been reported. From comparative computer modeling however, a model structure has been proposed for human C5a (Greer, 1985) on the basis of the crystal structure of C3a (Huber et al., 1980). C3a has 35% amino acid homology to C5a, and its disulfide linkages are located in homologous positions.

Recently, we have described the sequential assignment of the ^{1}H nuclear magnetic resonance (NMR)[1] spectrum of C5a (Zuiderweg et al., 1988a). The backbone protons for all residues and the side-chain protons of 57 residues were fully assigned while for the remaining 18 side chains partial assignments were obtained. On the basis of these results, we have identified for this protein a large number of interresidue nuclear Overhauser effects (NOEs) in the two-dimensional NOE spectrum (NOESY) (Jeener et al., 1979). In a preliminary report, we have recently presented an ensemble of tertiary structures calculated from constraints derived from these NOEs (Zuiderweg et al., 1988b) for purposes of comparison with the structure of C5a based on comparative computer modeling (Greer, 1985).

Here, we give a full account of the methods and procedures followed for the determination of the three-dimensional structure of C5a in solution. On the basis of a slightly larger NOE data set than that used in the preliminary communication (Zuiderweg et al., 1988b), we calculated an ensemble of nine equally well converged tertiary structures for C5a in which only NOE constraints were used as the target distances in the DISMAN program (Braun & Go, 1985). Another 11 structures were computed in which the distance list was augmented with constraints for the disulfide linkages (Zimmerman & Vogt, 1984) and for hydrogen bonding as determined from a careful investigation of the C5a secondary structure. The resulting ensembles of structures are mutually compared and compared with the crystal structure for C3a (Huber et al., 1980). Possible locations of receptor binding sites on the C5a molecule are discussed on the basis of the experimentally determined solution structure.

Materials and Methods

The NMR studies were carried out with recombinant C5a (75 residues) (rC5a[Met0]) obtained from a synthetic gene (Mandecki et al., 1985) expressed in a protease-deficient strain of *Escherichia coli* (Mandecki et al., 1986) which was grown under conditions giving high-level C5a expression (Mollison et al., 1987). The protein was purified by high-performance liquid chromatography (Carter et al., 1986). rC5a[Met0] is identical with human serum C5a except for the additions of an N-terminal methionine and glutathione moiety attached to Cys-27 and the deletion of the carbohydrate chain at Asn-64. No difference in the biological activity of human serum C5a and rC5a[Met0] could be detected for receptor binding, stimulation of lysosomal enzyme release, chemokinesis, or chemotaxis (Mollison et al., 1988b). Furthermore, the one-dimensional proton NMR spectrum of these species showed identical shifts for the aromatic and high field shifted methyl resonances (Zuiderweg, unpublished results), strongly indicating a similar folded structure.

Samples of approximately 7 mM rC5a[Met0] in ^2H_2O or 90% H_2O–10% ^2H_2O solutions at pH 2.3 without buffer were used for the NMR studies, which were acquired at 10 °C. Phase-sensitive NOESY data sets were collected with the standard pulse sequence (Jeener et al., 1980) and phase programs (States et al., 1982) and were acquired in the hypercomplex mode (Muller & Ernst, 1979; States et al., 1982) with the carrier placed in the center of the spectrum. NOESY spectra used for the determination of the secondary structure of C5a were recorded at 500 MHz on a General Electric GN500 spectrometer using mixing times of 25, 50, 75, and 100 ms. An incrementation of the NOE mixing time with 10% of the t_1 evolution time was utilized to shift out coherence transfers from the reference cross-peaks in an additional NOESY spectrum with a 25-ms mixing time (Macura et al., 1982). Solvent suppression was achieved by delivering a Dante pulse train from the low-power observe amplifier to the solvent resonance at all times except during t_2 and t_1 (Wider et al., 1984; Zuiderweg et al., 1986). A spectral width of ±2500 Hz was used, and data sets of 2K complex points in t_2 and 256 complex points in t_1 were acquired with 128 scans per t_1 value. A relaxation delay of 1 s was used. Amide proton exchange data were obtained from a series of COCONOSY experiments

[1] Abbreviations: NMR, nuclear magnetic resonance; NOE, nuclear Overhauser effect; 2D, two dimensional; COSY, 2D scalar correlated spectroscopy; NOESY, 2D nuclear Overhauser effect spectroscopy; rms, root mean square; rmsd, root-mean-square deviation.

(Haasnoot et al., 1984) in the phase-sensitive mode acquired on a GE GN500. Simultaneous scalar correlated (Aue et al., 1976) and NOESY data sets were accumulated with 8 h of acquisition time per set from a rC5a[Met0] solution freshly lyophilized from H_2O and dissolved in ^2H_2O at 10 °C, pH 2.3. With the carrier set at the low-field end of the spectrum, 400 real t_1 acquisitions of 48 scans each were acquired. The data used for the identification of NOEs in the aliphatic region of the spectrum were collected with a 7 mM C5a solution in ^2H_2O, 10 °C, pH 2.3, and were obtained at mixing times of 200 and 50 ms on a Bruker AM500 spectrometer. The spectra were acquired by oversampling in t_2 (spectral width ±5000 Hz) in order to increase the dynamic range (Delsuc & Lallemand, 1986) and to avoid base-line, phase, and intensity distortions caused by audio filters. The 200-ms data were acquired with 256 complex t_1 values of 64 scans each while the 50-ms data were collected with 256 complex t_1 values of 128 scans each. Low-power presaturation of the residual HO^2H signal was carried out during the relaxation delay of 1 s. The base line of the ω_2 slices was optimized by adjusting the receiver reference phase and acquisition delay to obtain a spectrum which needed no zero- or first-order phase correction.

Data processing was carried out in the format of the FTNMR program (Hare Research, Woodinville, WA) with in-house-written software utilizing a CSPI minimap array processor hosted by a VAX 8350 computer. For NOESY experiments, care was taken to optimize the base line of the spectra by manipulating the intensity of the first points of the free induction decays prior to Fourier transformations in both t_2 and t_1 (Otting et al., 1986). The data were digitally filtered as described in the figure legends. The occurrence of t_2 ridges was further minimized either by base-line correction of the ω_2 transformed slices with a fifth order polynomial, by automatic base-line correction (Pearson, 1977; Boelens et al., 1985; Zuiderweg, unpublished software) of the rows of the frequency domain matrix, or by both. The base plane of the spectrum was further improved by an automatic base-line correction of the ω_1 slices in the frequency domain matrix.

The calculations of C5a structures were carried out on a VAX 8600 computer using the DISMAN program (Braun & Go, 1985). Fourteen different starting structures were generated in which the backbone conformation was set in a random extended structure ($-165° < \phi < -75°$; $75° < \psi < 165°$) with fourteen independent seed numbers for the random choices in each starting conformation. The dihedral angles χ of the amino acid side chains in these starting structures were set to 180°. A protocol similar to that described by Braun and Go (1985) for the calculation of bovine pancreatic trypsin inhibitor structures was used. The DISMAN program used a variable target function for the calculation of structures from the distance data. First, the polypeptide chain was brought into a conformation compatible with constraints between atoms of residues which are adjacent in the sequence (short range), followed by convergence toward a target which also contained medium-range constraints while in the final stages all constraints were incorporated in the computation. The weight of the van der Waals distance constraints was concomitantly increased during the computations. The resulting structure was subsequently used as the starting structure for a new round of computations in which the same variable target function was used to move the conformation out of local minima. From test calculations it was decided that convergence of the structures toward the constraints did not improve after three minimization rounds. Therefore, all structures reported here were carried out with three rounds. The total calculation time for the structures reported in this paper was approximately 260 CPU hours.

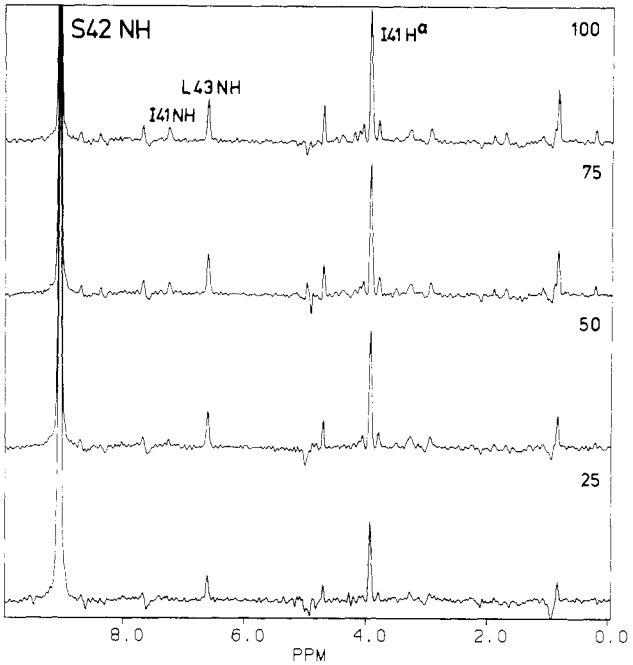

FIGURE 1: Cross sections parallel to ω_1 at the amide proton resonances of Ser-42 taken from NOESY spectra recorded with τ_m of 25, 50, 75, and 100 ms which were processed with sine bells shifted by 30 and 45° in ω_2 and ω_1, respectively. The sequential NOE peaks $d_{\alpha N}(i,i + 1)$ and $d_{NN}(i,i \pm 1)$ are indicated.

The DISMAN coordinate sets were modified to obtain the carbon positions corresponding to the methyl group pseudo atoms and were reformated to the Brookhaven Protein Databank format containing heavy atoms only. All graphics was carried out on an Evans and Sutherland MPS graphics system interfaced to a VAX 11/785 computer using the in-house-written Protein Display System software. Minimized rmsd superposition of the structures and addition of protons to the structures were carried out with the Protein Display System software.

RESULTS

Inspection of the NMR spectrum of C5a in H_2O [Figure 6 of Zuiderweg et al. (1988a)] shows that many NOEs occur between the amide protons of this protein. This implies that C5a contains a large amount of helical secondary structure (Billeter et al., 1982). Since helices form such tight structural entities in proteins, they determine the tertiary structure to a large extent. Therefore, it was of importance to obtain a good definition of the location and regularity of the helices from the NMR data prior to the tertiary structure determination. Accordingly, the 3D structure determination of C5a was divided into two steps: (a) the precise determination of the location of the helices and assessment of their regularity and (b) the folding of the tertiary structure of the protein on the basis of long-range distance constraints between the elements of secondary structure.

Secondary Structure of the Core of C5a. The secondary structure of C5a was delineated from quantitative measurements of the sequential NOEs between α and amide protons $[d_{\alpha N}(i,i + 1)]$ and between amide protons (d_{NN}), from identification of medium-range NOEs $[d_{\alpha N}(i,i + n)$, $n \geq 2$, and $d_{\alpha\beta}(i,i + 3)]$ (Wüthrich et al., 1984), and from identification of long, contiguous stretches of slowly exchanging amide

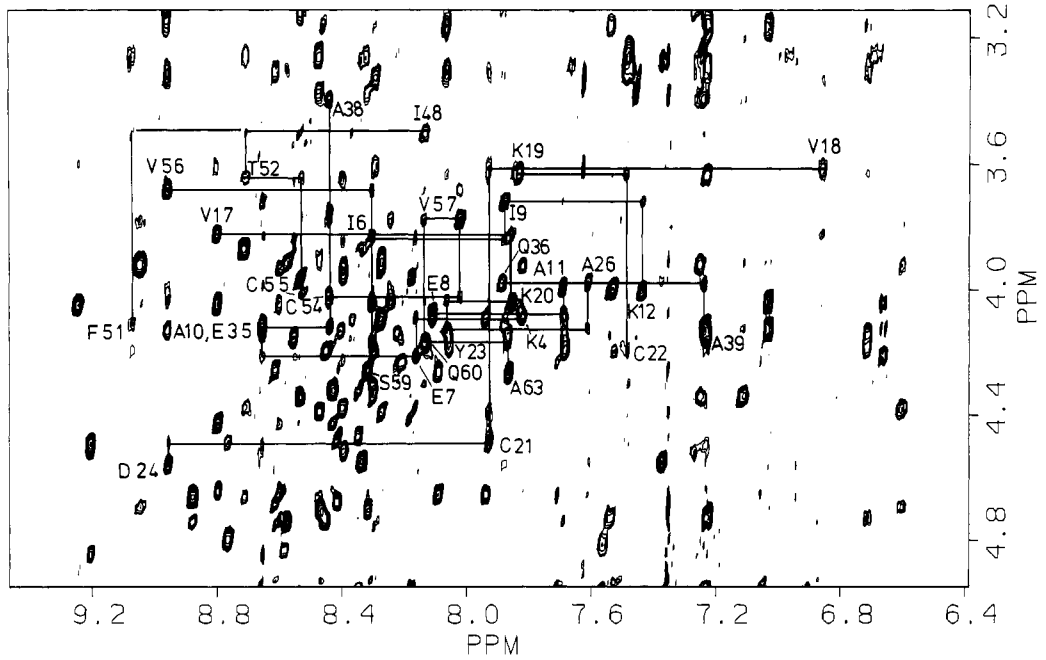

FIGURE 2: Fingerprint region of a 500-MHz ^1H NOESY spectrum or rC5a[Met0] in 90% H_2O–10% ^2H_2O, pH 2.3, 1.0 °C. The solvent resonance was suppressed by observe-field coherent time-shared radio-frequency saturation at all times except during t_1 and t_2. The NOE mixing time was 100 ms. The spectrum was processed with an unshifted sine bell in t_2 and a sinebell shifted by 30° in t_1. The digital resolution is 2.5 Hz/point in ω_2 and 4.9 Hz/point in ω_1. In this figure the observed $d_{\alpha N}(i,i+3)$ cross-peaks are documented; they are found at the intersections of the horizontal and vertical lines leading to the intraresidue NOEs identified by labels.

protons (Wagner, 1983; Zuiderweg et al., 1983a).

Figure 1 shows an example of cross sections through four NOESY spectra of C5a in H_2O recorded with mixing times of 25, 50, 75, and 100 ms at the amide proton position of Ser-42. It is seen that the NOE corresponding to $d_{\alpha N}(i,i+1)$ is intense and builds up rapidly, whereas the NOE corresponding to $d_{NN}(i,i+1)$ is much smaller and shows a slower build-up rate. This result shows that the Ile-41–Ser-42 link is of an extended type of structure. Build-up rates at $\tau_m = 0$ were extrapolated from the peak intensities from similar quality data as shown in Figure 1 for a large number of amide protons by a computer fitting routine. It was found that the build-up rates were directly proportional to the intensities at $\tau_m = 25$ ms. The rates were converted to distances with the build-up rate of a NOE cross-peak between Tyr 3,5-protons and Tyr 2,6-protons as a distance calibration and are presented in Table I. Due to limitations in signal/noise ratio for the smaller peaks, the accuracy of the distances in this table is not better than 10% (reflecting estimated errors of a factor of 2 in intensity measurements). The precision of the distances in the C-terminal region of C5a is undetermined because of increased mobility (see below).

Figure 2 shows an expansion of the NH–C_αH cross-peak region of a NOESY spectrum of rC5a[Met0] which documents the identification of the medium-range NOEs, $d_{\alpha N}(i,i+3)$. Observation of this type of NOE is essential to the delineation of helices (Zuiderweg et al., 1983a; Wüthrich et al., 1984). As can be seen, most of these NOEs are close to the noise level in the spectrum. Thus, absence of this type of connectivity in a particular region does not exclude helicity in that region. Furthermore, spectral overlap also prohibits a complete analysis of the possible presence and/or absence of these intermediate-range NOEs.

Quantitative amide proton exchange rates were obtained from a time series of 2D spectra of rC5a[Met0] freshly lyophilized from H_2O, recorded in ^2H_2O. The results of this study as derived from a series of five consecutive 2D spectra of 8-h individual acquisition time in which scalar and NOE information was simultaneously collected (Haasnoot et al., 1984) are listed in Table II for rC5a[Met0] at pH 2.3, 10 °C.

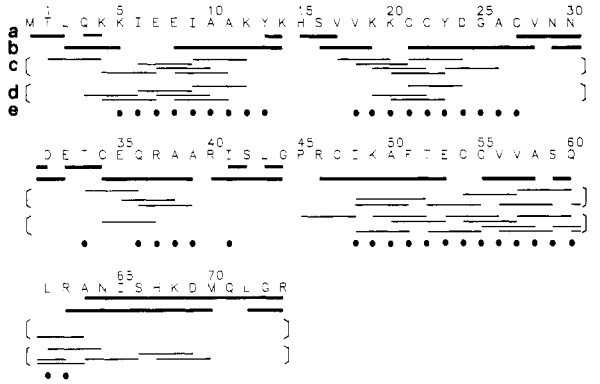

FIGURE 3: Overview of the NMR data obtained for the determination of the secondary structure of rC5a[Met0], pH 2.3, 10 °C. Lane a, $d_{\alpha N}(i,i+1)$ are indicated only when the corresponding cross-peaks were of large intensity; lane b, all d_{NN} are shown, irrespective of their intensity; lanes c and d, medium-range NOEs $d_{\alpha N}(i,i+3)$ and $d_{\alpha\beta}(i,i+3)$, respectively; lane e, slowly exchanging amide protons ($k_{ex} < 0.05$ h^{-1}) indicated by filled circles.

In Figure 3 the results of the above-mentioned studies are compiled in a schematic fashion. Positive evidence for helical structure (Wüthrich et al., 1984), i.e., absence of strong $d_{\alpha N}(i,i+1)$ NOEs (trace a) and presence of d_{NN} (b), $d_{\alpha N}(i,i+3)$ (c), and $d_{\alpha\beta}(i,i+3)$ (d) NOEs, is seen to accumulate in four separate contiguous regions of the protein: 4–13, 17–27, 34–39, and 45–63. In addition, long contiguous stretches of amide protons which exchange more slowly than 0.05 h^{-1}, an order of magnitude slower than that of amide protons in unstructured peptides (Englander et al., 1972), are also observed in these helical segments. Since it is very unlikely that these long contiguous regions of amide protons are buried in the interior of a protein as small as C5a, their slow exchange is taken to indicate hydrogen bonding (see Discussion). Thus, further evidence of helicity in the above-mentioned peptide regions of C5a is provided by these experiments. Taking into

Table I: Backbone Proton Distances in rC5a[Met0], pH 2.3, 10 °C[a]

residues	H_α-NH (Å)	NH-NH (Å)
Leu-2-Gln-3	3.1	2.9
Gln-3-Lys-4	2.9	3.0
Lys-4-Lys-5		2.8
Ile-6-Glu-7	3.8	
Ile-9-Ala-10	3.8	
Ala-10-Ala-11		2.9
Ala-11-Lys-12		3.0
Lys-12-Tyr-13		3.2
Tyr-13-Lys-14		4.0
Ser-16-Val-17		3.7
Val-17-Val-18	3.6	2.9
Val-18-Lys-19		2.6
Cys-21-Cys-22		3.0
Cys-22-Tyr-23		3.1
Tyr-23-Asp-24	3.5	3.0
Asp-24-Gly-25	4.3	2.9
Gly-25-Ala-26		3.6
Ala-26-Cys-27	3.6	3.6
Cys-27-Val-28	2.4	3.6
Val-28-Asn-29	2.2	
Asn-29-Asn-30	2.5	3.4
Asn-30-Asp-31	2.6	3.1
Asp-31-Glu-32		2.7
Glu-32-Thr-33	2.6	
Cys-34-Glu-35		2.8
Gln-36-Arg-37		3.2
Arg-37-Ala-38		2.6
Ala-39-Arg-40	3.5	
Arg-40-Ile-41	3.2	
Ile-41-Ser-42	2.2	
Ser-42-Leu-43	3.1	2.8
Leu-43-Gly-44		3.1
Arg-46-Cys-47		2.4
Ile-48-Lys-49	3.8	3.0
Lys-49-Ala-50		3.0
Ala-50-Phe-51		3.0
Phe-51-Thr-52	3.5	3.0
Thr-52-Glu-53	3.7	3.4
Cys-55-Val-56	3.6	3.1
Val-56-Val-57	3.8	2.9
Val-57-Ala-58	3.6	2.9
Arg-62-Ala-63		2.7
Ala-63-Asn-64	3.0	
Asn-64-Ile-65	2.9	3.0
Ile-65-Ser-66		2.8
Ser-66-His-67	2.7	
His-67-Lys-68	2.5	
Lys-68-Asp-69	2.6	3.0
Asp-69-Met-70	3.0	3.3
Met-70-Gln-71	3.0	
Leu-72-Gly-73	2.8	3.2

[a] Distances were obtained from the build-up rates of NOE cross-peaks observed in NOESY spectra recorded with mixing times of 25, 50, 75, and 100 ms. Estimated error is 10%.

Table II: Hydrogen/Deuterium Exchange of Amide Protons in rC5a[Met0], pH 2.3, 10 °C[a]

residue	rate (1/h)	residue	rate (1/h)	residue	rate (1/h)
T1	nd	A26	<0.007	F51	<0.007
L2	>0.20	C27	<0.007	T52	<0.007
Q3	>0.17	V28	0.14	E53	<0.007
K4	0.14	N29	>0.02	C54	<0.007
K5	0.033	N30	nd	C55	<0.007
I6	0.025	D31	nd	V56	<0.007
E7	<0.007	E32	nd	V57	<0.007
E8	0.042	T33	<0.007	A58	<0.007
I9	<0.007	C34	nd	S59	<0.007
A10	0.010	E35	>0.2	Q60	<0.007
A11	0.010	Q36	<0.007	L61	0.042
K12	0.025	R37	<0.007	R62	0.042
Y13	<0.007	A38	<0.007	A63	>0.2
K14	nd	A39	<0.007	N64	nd
H15[b]	nd	R40	>0.2	I65	0.14
S16	nd	I41	<0.007	S66	nd
V17	nd	S42	nd	H67	nd
V18	0.012	L43	0.066	K68 b)	nd
K19	<0.007	G44	>0.2	D69	nd
K20	0.016	P45		M70	nd
C21	<0.007	R46	nd	Q71	nd
C22	<0.007	C47	>0.2	L72	nd
Y23	<0.007	I48	<0.007	G73	nd
D24	0.010	K49	<0.007	R74	nd
G25	<0.007	A50	<0.007		

[a] Exchange rates were determined from a series of four COCONOSY experiments of 8 h each; nd indicates that the rate could not be determined since the corresponding cross-peaks were not observed in the first exchange spectrum. [b] Not observable due to spectral overlap or saturation.

account some uncertainty in the exact termination of the helices, a conservative interpretation of the data shows helices to be located in the peptide regions 4–12, 18–26, 34–39, and 46–63. Strictly speaking, the NOE results do not permit a distinction between α and other types of helices for these regions since only a very limited number of other intermediate-range NOEs [especially the diagnostic $d_{\alpha N}(i,i+4)$, see Wüthrich et al. (1984)] was observed. However, meaningful results were obtained for the 3D structure determination when an α-helical hydrogen-bonding pattern was imposed on these regions (see below). Therefore, it can be assumed with confidence that the NMR data obtained for rC5a[Met0] delineate α-helical stretches.

From Table I it is seen that the distances obtained for the $d_{\alpha N}(i,i+1)$ and d_{NN} connectivities in the helical regions of C5a are very close to their theoretical values (3.6 and 2.9 Å, respectively). Therefore, no indication of severe helical distortion is found in these regions with the exception of the occurrence of a long d_{NN} distance between Thr-52 and Glu-53 (3.4 Å). However, the $d_{\alpha N}(i,i+1)$ distance between these residues is also long (3.7 Å), which would be indicative of helicity. Moreover, amide proton exchange is slow in a contiguous region around these residues, and many medium-range NOEs bridging this area are observed (Figure 3). Although we currently do not have an explanation for the long d_{NN} distance between residues 52 and 53, all the other available data indicate that all helices in C5a are quite regular.

Conformation of the C-Terminal 11 Peptide. A simultaneous presence of strong $d_{\alpha N}(i,i+1)$ and d_{NN} NOEs was observed for residues beyond Ile-65 (Figure 3). Unfortunately, not many exact distances could be determined from the quantitative NOE studies for these residues due to spectral overlap. However, those for which such distances could be obtained showed that the apparent $d_{\alpha N}(i,i+1)$ and d_{NN} distances in this region are approximately equal and about 3 Å (Table I). This situation occurs only for a very narrow region of the backbone ϕ and ψ angles ($-120° > \phi > -140°$; $50° < \psi < 70°$), a conformation which has an energy of approximately 2 kcal above the minimum energy conformation (Brant & Flory, 1967). Therefore, it is very unlikely that such a specific and unfavorable conformation will pertain throughout a long peptide segment. An alternative explanation for these observations prevails: the apparent equal distances result from the dynamic averaging within an ensemble of the backbone conformations, analogous to the observations made for unrestricted peptides in aqueous solution (Gampe et al., 1987). This explanation is supported by the fact that no slowly exchanging amide protons could be detected in the region beyond residue 63 (Figure 3), by the observation that the chemical shifts of the C_α proton resonances in this region of C5a correspond more closely to that of random coil peptides (Bundi & Wüthrich, 1979) than that for the core of the protein (Figure 4), and by the fact that the line widths of the reso-

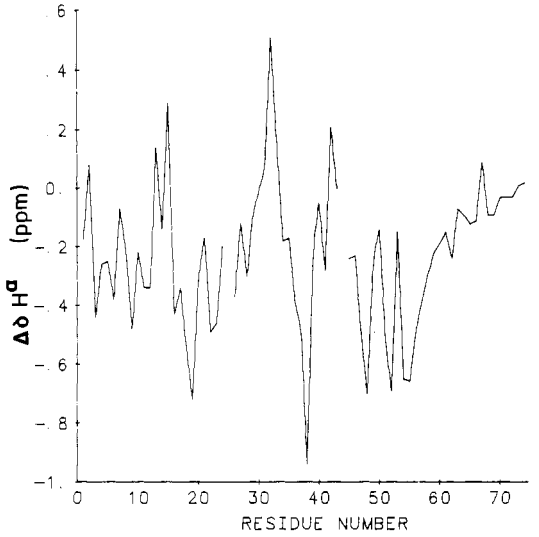

FIGURE 4: Difference between the chemical shifts for the $C_\alpha H$ resonances in rC5a[Met0] with those of corresponding amino acids in tripeptides.

nances become smaller in this area as compared to the rest of C5a. In the region 64–70 the NMR data do not support a complete random conformation but rather a dynamic ensemble weighed toward helicity as follows from the observation of $d_{\alpha\beta}(i,i+3)$ NOEs between residues 63 and 66, 66 and 69, and 67 and 70, while large $d_{\alpha N}(i,i+1)$ connectivities are simultaneously present (Figure 3). For the last few residues on the other hand, the NMR data are consistent with a complete disorder. Thus, it must be concluded that residues C terminal to Ala-63 increasingly acquire a dynamic random state as one proceeds toward the end of the peptide chain.

An interesting correlation follows from further inspection of the data in Figure 4; C_α proton chemical shifts in the helices of C5a are invariably high field from the corresponding random coil shifts. A similar correlation was found for the α-proton resonances in C3a (Nettesheim et al., 1988) and in *lac*-repressor headpiece (Zuiderweg et al., 1983b), both proteins which contain a large amount of helicity. α-Proton chemical shifts in the basic pancreatic trypsin inhibitor have been correlated with short $C_\alpha H$ to oxygen distances (Wagner et al., 1983). This correlation does not seem to hold for the shifts in α-helical proteins: short $C_\alpha H$ to oxygen distances in ideal α-helices are 2.54 and 2.71 Å for $\alpha(i)$–$O(i)$ and $\alpha(i)$–$O(i-1)$, respectively, which would predict downfield shifts on the basis of the correlation curve [Figure 1 of Wagner et al. (1983)]. Nevertheless, it appears that the observation of contiguous stretches of upfield-shifted α-protons may be used as another criterion for the delineation of helices.

In summary, four stable helices are found in rC5a[Met0], hereafter referred to as helices I (4–12), II (18–26), III (34–39), and IV (46–63). Beyond residue 63, helix IV loses stability and gradually becomes disordered. Essentially similar results for the secondary structure of C5a were obtained from data acquired at higher pH values and from C5a which lacks the N-terminal methionine residue and which is therefore more closely related to human serum C5a (Figure 5). These results allow the conclusion that the observed secondary structure is not an artifact of the low pH at which the data were obtained. Furthermore, the resonances of amide protons of the C-terminus of C5a in spectra obtained at pH 6.0 and 20 °C in 90% H_2O could not be observed due to solvent exchange, indicating that the C-terminus is also disordered in solution at more physiological conditions.

Tertiary Structure Constraints. Although assignments for

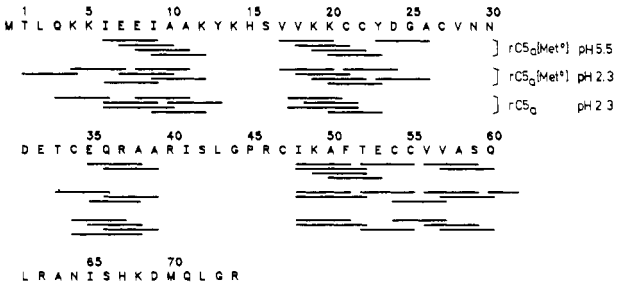

FIGURE 5: Comparisons of medium-range NOEs $[d_{\alpha N}(i,i+3)]$ for rC5a[Met0] at pH 5.5 (top), rC5a[Met0] at pH 2.3 (middle), and rC5a at pH 2.3 (bottom).

virtually all backbone protons and for the majority of the side-chain resonances were obtained (Zuiderweg et al., 1988a), the identification of NOE cross-peaks to individual proton pairs in the NOESY spectrum of C5a was not straightforward. This is due to the fact that many proton resonances overlap, making unique identification of the NOE pair impossible for many cross-peaks. The following combination of approaches has been used to resolve most of the ambiguities. NOE cross-peaks were analyzed in a NOESY spectrum recorded under conditions of spin diffusion (τ_m = 200 ms) (available as supplementary material; see paragraph at end of paper regarding supplementary material). In such a spectrum proton–proton selective NOEs are propagated to residue–residue NOEs, which makes identification easier since matrices of NOEs between characteristic residue patterns can be recognized. This approach may be viewed as similar to the use of NOESY–COSY (multiple) relay (Wagner, 1984) for the resolution of degeneracies, with the advantage of higher sensitivity but with the obvious disadvantage of less reliable specificity. Once it was determined which residues were spatially close from the spin-diffusive data, the cross-peaks still observable in a non-spin-diffusive spectrum (recorded with a NOE mixing time of 50 ms) were assigned to the specific proton pairs within those residues (spectrum available as supplementary material).

Secondly, ambiguities in cross-peak assignments were resolved by identifying the approximate spatial position of the different possible proton pairs in a preliminary model, obtained on the basis of the known secondary structure and several unambiguous key NOEs. In this way, many proton pairs could be excluded from the possible cross-peak assignment list since they were found to be located in spatially distant areas of the protein. Assignments compatible with the preliminary structure could be obtained for every ambiguous NOE cross-peak.

By use of this strategy, approximately 300 interresidue cross-peaks were analyzed in the aromatic, methylene, and methyl regions of the spin-diffusive NOESY spectrum of rC5a[Met0], pH 2.3, 10 °C. This data set was edited to obtain 73 NOEs between protons of residues further apart than three residues in the sequence (long-range NOEs) observable in the non-spin-diffusive spectrum. No long-range NOEs were observed between residues within the C-terminus itself, as was to be expected for this dynamically averaging peptide region, or between residues of the C-terminal peptide and other parts of the molecule, showing that this region does not preferentially dock on any region of the core of C5a. Consequently, no determination of the structure of this region was attempted, and the short- and medium-range NOEs for this area were not included in the final list of distance constraints.

A pictoral representation of the distribution of the long-range NOEs is given in Figure 6, where an α-carbon plot of

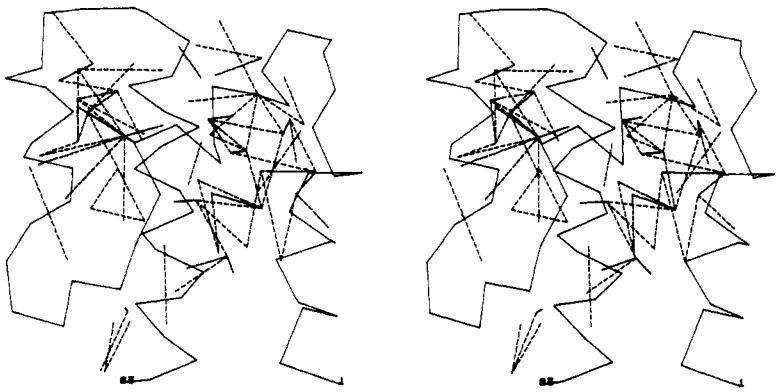

FIGURE 6: Representation of the location of long-range NOEs (NOEs between residues further apart than three units in the amino acid sequence) in the 3D structure of C5a 1–63. The α-carbon skeleton of the NMR structure coded A4 (see Table IV) was used. The dashed lines, representing the NOEs, are drawn to the positions of the involved protons or, if no stereospecific assignments were made for the protons, to the adjacent carbon atom.

a representative computed NMR structure is used to show the spread of the NOEs in the molecule. As can be seen, many long-range NOEs define the relative positioning of the helices while only a limited number of constraints are available for the loop regions.

Together with the short- and medium-range NOEs, a data set of 285 interresidue NOE distance constraints could be established for the region 0–63 of C5a from NOESY data sets recorded with τ_m = 50 ms. The constraints were modified into pseudo atom constraints (Wüthrich et al., 1983) for those protons which were not assigned stereospecifically. These included all methylene protons and all methyl groups except for those of the α-helical Val-18, Val-56, and Val-57 residues where stereospecific assignments were obtained for the methyl groups from the relative intensity of the intraresidue NH–methyl NOEs at 25-ms mixing time (Zuiderweg et al., 1985a).

Upper and lower limits for distances were obtained for the quantitative NOE data of Table I by using a 10% error range. Other NOEs were classified in several groups: as determined from the counting of contour levels in the base plane corrected 50-ms NOESY spectrum, weak NOEs were assigned a range of 3.0–4.0 Å, NOEs of medium intensity a range of 2.5–3.5 Å, and strong NOEs a range of 2.0–3.0 Å. NOEs which could not be quantified from the counting of contour lines (e.g., because of partial overlap or proximity to the diagonal or a noise band) were given a conservative range of 2.0–4.0 Å. An upper limit of 4 Å was chosen for the weak NOEs because some $d_{\alpha N}(i,i+4)$ cross-peaks could still be observed for helical regions of C5a in NOESY spectra recorded with a mixing time of 50 ms. The theoretical value of this distance in an ideal α-helix is 4.1 Å. The classification and ranges of the other categories of NOESY cross-peaks were derived from this standard.

Whenever no NOE could be detected between backbone protons of sequentially adjacent residues in uncrowded parts of the spectrum, a lower bound of 3.5 Å was imposed on the corresponding distance in the constraints list to create a repulsive force between these protons. The upper limits of the NOE constraints were modified by adding the appropriate factors for pseudo atom corrections (Wüthrich et al., 1983). For those structure calculations which included hydrogen-bond constraints, two distances per O(i)–NH($i+4$) bond were given for the following residues within the regions O(4)–NH(8) upto O(8)–NH(12) (helix I), O(18)–NH(22) upto O(22)–NH(26) (helix II), O(34)–NH(38) and O(35)–NH(39) (helix III), and O(46)–NH(50) upto O(59)–NH(63) (helix IV). The O–N distance was assigned the range 2.7–3.0 Å and the O–NH distance the range 1.8–2.0 Å (Williamson et al., 1985).

Table III: Statistics of C5a Structure Calculations with 285 NOE Constraints[a]

structure	NV_{low}	$\sum V_{low}$	NV_{upp}	$\sum V_{upp}$	rmsv
C5aN1	32	5.52	70	21.83	0.21
C5aN2	23	3.05	74	20.32	0.19
C5aN3	32	5.30	74	18.61	0.19
C5aN4	28	5.21	71	20.62	0.20
C5aN5	20	3.27	77	20.19	0.20
C5aN6	32	6.14	74	17.87	0.20
C5aN7	35	4.88	81	21.38	0.21
C5aN8	28	5.99	77	21.00	0.21
C5aN9	25	4.46	72	19.14	0.20

[a] NV_{low} and NV_{upp} indicate the number of violations with respect to the lower and upper bounds; $\sum V_{low}$ and $\sum V_{upp}$ give the total violation in angstrom units for lower and upper bounds, respectively; rmsv represents the root mean square violation in angstrom units calculated as $[\sum (V_{low}^2 + V_{upp}^2)/N]^{1/2}$, where N is the total number of constraints (285).

Biochemical evidence identifying the pairing of the cysteine residues as 21–47, 22–54, and 34–55 (Zimmerman & Vogt, 1984) was utilized as three constraints per disulfide and incorporated in the constraint list in some of the structure calculations. The ranges were as follows: C_β–S_γ, 3.0–3.1 Å; S_γ–S_γ, 2.0–2.1 Å. In this way, the total of 285 interresidue NOE distance constraints for the region 1–63 of rC5a[Met0] was augmented for some structure calculations by 61 hydrogen and disulfide constraints to yield a data set of 346 entries. The complete list of all distance constraints used in the computations is available as supplementary material.

Structure calculations were carried out with the DISMAN dihedral angle folding distance geometry program (Braun & Go, 1985). Several different structure calculations were performed. With only the NOE data, 14 structure calculations were carried out from 14 starting structures which were independently generated. The calculations resulted in nine equally well converged structures (Table III). With the same 14 random starting structures, 14 calculations were made in which the full list of constraints (NOEs, hydrogen bond, and disulfide bond) was used as the target function. The results of these calculations for 11 equally well converged structures are listed in Table IV.

Over 80% of the remaining violations in all structures is less than 0.4 Å while the largest violations never exceeded 1.1 Å. Residual constraint violations did not map in a particular region of the protein; their distribution appeared random. This was not the case for violations in trial runs of the program in which consistent violations were observed for two constraints in all resulting structures. Upon reinspection of the data, alternative assignments could be made for the cross-peaks

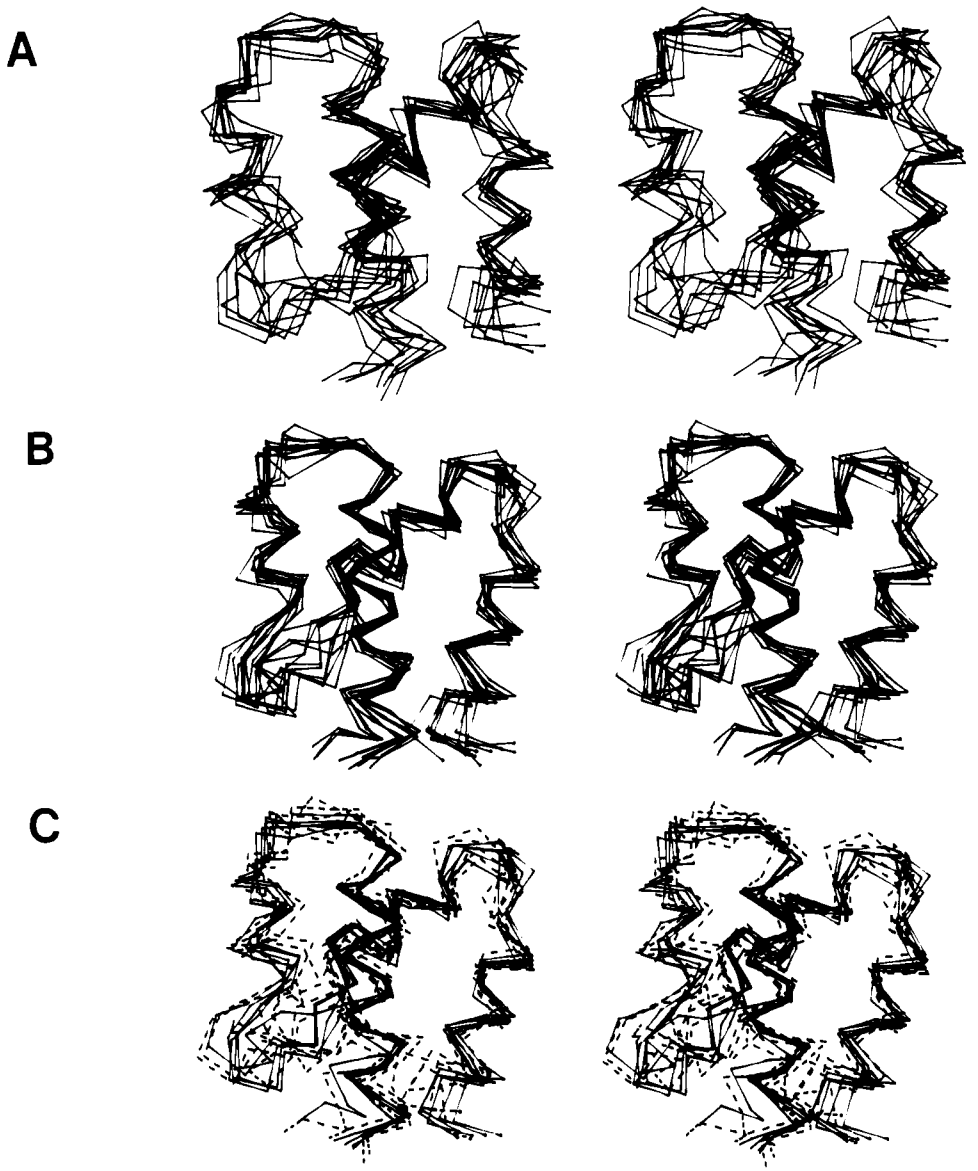

FIGURE 7: (A) Superposition of the nine equally well converged structures (listed as N1–N9 in Table III) obtained from calculations in which the 285 NOE constraints were taken as a target function. The rmsd of the C_α positions 1–63 was minimized for the ensemble by rotation and translation of the individual structures. (B) Superposition of the 11 equally well converged structures (listed as A1-A11 in Table IV) obtained from calculations in which all 338 constraints (NOEs, H-bond, and disulfide) were taken as a target function. The global rmsd of the C_α positions 1–63 was minimized. (C) Superposition of five representative structures of the calculations involving only NOE constraints (dashed) and five representative structures in which all constraints (NOEs, H-bond, and disulfide) were used as a target function (drawn). The structures were superimposed after rmsd minimization of the α-carbon positions for the peptide regions 3–12, 17–26, 34–39, and 46–60.

involved. Thus, it appeared that inconsistencies in the earlier data were signaled by the program; consequently, we take the random distribution of violations with respect to the current target function to indicate that such inconsistencies are absent.

Figure 7A shows a superposition of the nine equally well converged structures obtained with DISMAN structure calculations involving NOE distance constraints only. As can be seen, the relative positions of the helices are very well defined. In contrast, considerable variation is found for the conformation of the peptide loops connecting the helices. This reflects the relatively small amount of experimental data for these regions available to date. The average RMSD in this ensemble of structures is 2.43 ± 1.2 Å and 2.88 ± 1.1 Å for all C_α atoms and all heavy atoms, respectively (table available as supplementary material). The rms variation for the individual C_α positions for this ensemble of structures is plotted in Figure 8A. Figure 7B shows a superposition of the 11 equally well converged structures obtained with DISMAN structure calculations involving all distance constraints. The average rmsd in this ensemble of structures is 1.67 ± 1.0 Å and 2.13 ± 0.8 Å for all C_α atoms and all heavy atoms, respectively (table available as supplementary material). The rms variation for the individual C_α positions for this ensemble of structures is plotted in Figure 8B. A superposition of the C_α skeletons of five representative structures from each ensemble is shown in Figure 7C.

Discussion

Precision of the Structures. Central questions in the discussion of the precision of NMR structures are whether structures completely different from the obtained ensemble of structures would also be compatible with the NMR data and if the local spread of structures found reflects indeterminacy and imprecision in the data rather than poor sampling of local conformational space by the comuter algorithm. Since relatively imprecise distance constraints by themselves cannot determine the chirality of the structures, it is in principle possible that an ensemble of mirror-image structures exists

Table IV: Statistics of C5a Structure Calculations with 346 NOE, H-Bond, and Disulfide Constraints[a]

structure	NV_{low}	$\sum V_{low}$	NV_{upp}	$\sum V_{upp}$	rmsv
C5aA1	39	5.97	82	16.95	0.16
C5aA2	40	6.49	90	18.42	0.16
C5aA3	39	5.21	87	17.15	0.15
C5aA4	33	4.22	80	14.37	0.13
C5aA5	50	5.94	86	18.70	0.17
C5aA6	34	4.84	96	19.52	0.16
C5aA7	43	5.34	87	17.20	0.14
C5aA8	41	6.14	92	18.47	0.17
C5aA9	46	6.76	88	18.83	0.17
C5aA10	41	5.35	88	18.46	0.17
C5aA11	45	4.97	78	18.27	0.17

[a] NV_{low} and NV_{upp} indicate the number of violations with respect to the lower and upper bounds; $\sum V_{low}$ and $\sum V_{upp}$ give the total violation in angstrom units for lower and upper bounds respectively; rmsv represents the root-mean-square violation in angstrom units calculated as $[\sum(V_{low}^2 + V_{upp}^2)/N]^{1/2}$, where N is the total number of constraints (346).

which is equally compatible with the NMR data. For C5a, however, right-handed helicity is unambiguously determined from the precise short- and medium-range NOEs: a left-handed helix of L-amino acids would yield extremely short distances for the $d_{\alpha N}(i,i+1)$ NOEs (2.2 Å), and all medium-range distances are beyond the NOE detection limit (>4.5 Å). This clearly conflicts with our experimental observations. Thus, the presence of four right-handed helices in C5a serves as an internal reference for the total chirality of the molecule, and a mirror-image structure is exceedingly unlikely to be compatible with the data.

We have carried out additional computations to determine whether the local spread of structures obtained is representative of the conformational space accessible within the data constraints. We found that an increase of the upper bounds of the constraints by 0.5 Å (except for the quantitative NOEs) in computations involving NOEs only resulted in an increase of the rms variation by 0.3 Å for the α-carbons and 0.5 Å for all heavy atoms as compared to the calculation shown in Table III. This substantial increase in rms variation upon loosening of the bounds is indeed indicative of a spread in structures that is dictated by the data and not by a biased undersampling of conformational space by the program. Furthermore, if the spread of structures would have been determined by the random residual violations of the target function, a limited loosening of the constraints would not result in an increase in spread. That the residual violations do not determine the spread is further demonstrated by the fact that structure calculations carried out with the additional tight hydrogen-bonding constraints did result in a decrease in rmsd (see Results), but not in an increase of violations (cf. Tables III and IV).

The overall precision of the structural ensemble obtained from calculations involving NOE constraints only is an average 2.43-Å rmsd between the C_α atoms and a 2.88-Å average for all heavy atoms. From Figure 8A it is seen that the structures are much better defined for local areas: the α-helical regions I, II, and IV and their relative positioning are determined with a precision of around 1 Å. Thus, it is found that the relatively large numbers for the overall rmsd are mainly determined by the loop and terminal regions of the molecule, for which only a limited amount of NOE data could be identified. Whether the lack of NOE data for the loop regions is caused by a physical absence of structure due to mobility or by a static structure which does not display long-range NOEs because it is generally spatially remote from the core of the molecule remains to be determined.

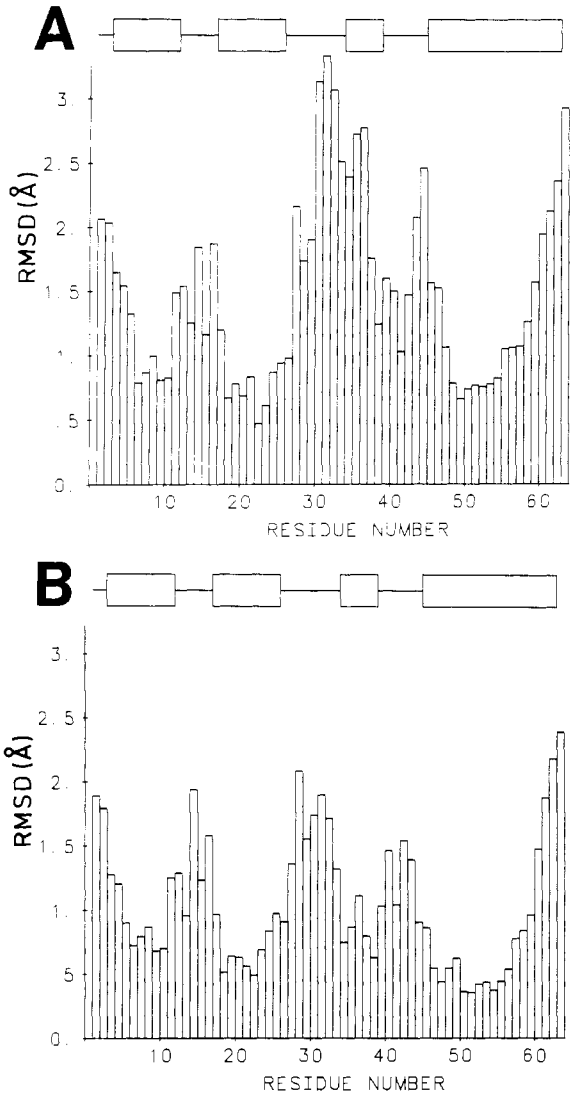

FIGURE 8: (A) rmsd of the individual α-carbons of the ensemble of nine structures shown in Figure 7A (NOEs only) with respect to a calculated average structure. The global rmsd over the α-carbons 1–63 was minimized. (B) rmsd of the individual α-carbons of the ensemble of 11 structures shown in Figure 7B (all constraints) with respect to a calculated average structure. The global rmsd over the α-carbons 1–63 was minimized.

More precise structure calculations using a list of distance constraints that included hydrogen-bonding and disulfide linkage constraints were carried out (Figure 7B). The disulfide linkages were directly obtained from experimental data (Zimmerman & Vogt, 1984), while the hydrogen-bonding constraints were obtained indirectly from the NMR data delineating the secondary structure. As noted under Results, the long contiguous regions of slowly exchanging amide protons observed for the helical regions of C5a are most likely to reflect hydrogen bonding, rather than solvent inaccessibility of the amide moieties in a small protein like C5a. This was rigorously tested in the structure calculations in which an α-helical hydrogen-bonding pattern was assumed for the slowly exchanging amides within the helices (see Results). As is shown in Table IV, these constraints could be incorporated in the calculations, without increasing the residual constraints violations of the resulting structures. Furthermore, as is shown in Figure 7C where the ensembles of structures for both calculations are superimposed, the ensemble of hydrogen-bond-constrained structures falls within the ensemble of structures which were not subject to these constraints. It is found that the helical

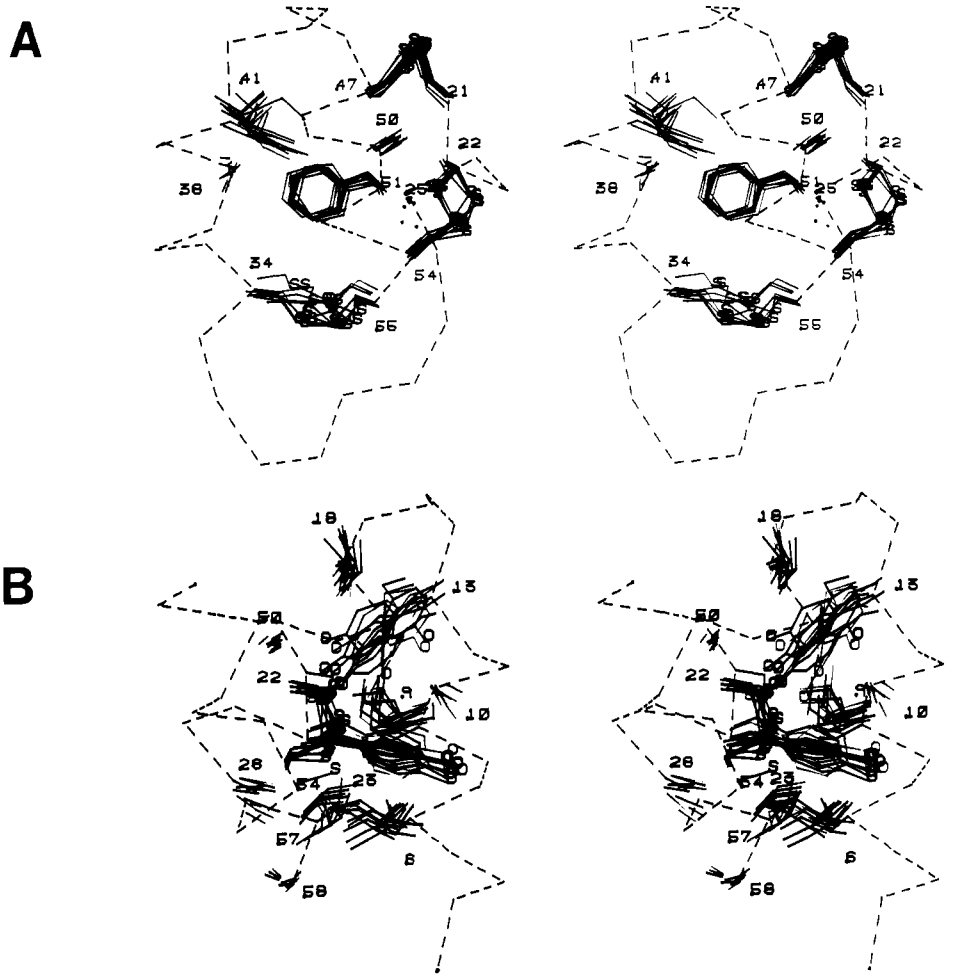

FIGURE 9: Superposition of internal helical residues in the ensemble of 11 structures obtained with all constraints. The global rmsd between the heavy atoms of residues, 6, 9, 10, 13, 18, 21, 22, 23, 25, 26, 34, 38, 41, 47, 50, 51, 54, 55, 57, and 58 was minimized. (A) Region of the central core. Indicated are C21, C22, G25, C34, A38, I41, A50, F51, C54, and C55. (B) Region of contact between helix I and the core. Indicated are I6, I9, A10, Y13, V18, C22, Y23, A26, A50, C54, V57, and A58.

pitch is the same in both ensembles. It can therefore be concluded that the NOE data are compatible with other independently derived experimental information (disulfide linkages) and with a regular α-helical structure for its main structural components. Conversely, these results clearly indicate that the slow exchange of long, contiguous regions of amide protons in this molecule is primarily caused by hydrogen bonding rather than by solvent inaccessibility. Indeed, in C5a many amide protons in the helices are located at the surface of the molecule, and no clear correlation of faster exchange of these groups with solvent accessibility is observed (e.g., Glu-7 is external but exchanges extremely slowly, see Table I). Exchange rates differ however between the helices in C5a; exchange rates increase in the order helix IV, helix II, helix III, and helix I as follows from Table I and from studies carried out at pH 5.8 and 10 °C (unpublished results). These data suggest that the exchange rates are correlated to the solvent accessibility of entire helices or, equivalently, to the number of hydrophobic contacts that have to be disrupted in order to allow opening of the helices. That amide proton exchange does not allow the exact delineation of helical termination follows for example from the observation that Lys-19 exchanges slowly, while it cannot be involved in a helical hydrogen-bonding pattern since the helix terminates at Val-18 as determined from the NOE data.

The precision of the structures generated with the additional constraints for hydrogen and disulfide bonds is 1.67 Å for the α-carbons and 2.13 Å for all heavy atoms. From Figure 8B it is found that the structure is determined better than 0.7 Å at the α-carbon positions for the bulk of the helical regions I, II, and IV. It follows from the comparison of panels A and B of Figure 8 that additional restraining of the helices causes an expected decrease of variation not only in those regions but also in the poorly defined loop regions. This is due to the better definition of the terminal positions of the helices which restrict the accessible conformational spaces for the connecting loops.

The average rms deviation between the heavy atom positions of the helical residues in the fully restrained ensemble of structures is 1.25 Å after minimization of the global rms deviation between these residues. A 1.01-Å rmsd between the heavy atoms of the internal helical residues (residues I6, I9, A10, Y13, V18, C21, C22, Y23, G25, A26, C34, A38, I41, C47, A50, F51, C54, C55, V57, and A58) is obtained after a global rmsd minimization over these residues. The resulting superposition of these latter residues is presented in Figure 9, where only parts of the C5a backbone ensemble are shown for reasons of clarity. It is found that in particular the positions of residues F51 and Y23 are determined with high precision. This results from the large number of constraints for these residues even though most of these constraints are very imprecise due to the large pseudoatom corrections for these residues.

In summary, it is shown that the conformational spread of the structures is determined by the NOE data and not by the sampling of the algorithm or by the residual violations in the calculations. Furthermore, the NMR data are compatible with

the regular α-helicity for the main structural units and with the independently determined disulfide linkages.

The obtained ensemble reflects those structures which are compatible with the data and with the standard atomic radii and bond-length, -angle, and -torsion parameters, which are not treated as variables in the DISMAN program. Also, the peptide bond dihedral angles are strictly kept at 180° in the structure calculations. However, although the obtained structures are physically plausible for those reasons, no consideration is given to molecular forces such as van der Waals attraction, 1–4 interactions, and electrostatics. Minimization of these potential energy terms needs to be carried out, and the obtained ensemble of structures may therefore be viewed as starting structures for a restrained molecular dynamics optimization procedure, which simultaneously minimizes the potential energy and residual violations of the conformations (Zuiderweg et al., 1985b; Kaptein et al., 1985; Clore et al., 1985, 1986). However, the incorporation of hydrogen-bond constraints in the present DISMAN structure calculations already represents a partial refinement: after it was determined that the NOE data by themselves dictate the amide protons to be close to hydrogen-bonding acceptor groups, it would be extremely unfavorable if the bonds could not be formed due to nonideal distances between the donor and acceptor atoms. Therefore, the ensemble of hydrogen-bond-constrained structures in which those distances are regularized should be a better representation of physical reality than the ensemble in which those bonds were not regularized. In order to optimize other energetic terms, we are currently carrying out extensive restrained molecular dynamics calculations.

Description of Structural Features. The structure of C5a 1–63 can be described as an approximately antiparallel bundle of four helices. This coiled-coil arrangement of helices is expected for a structure of this type and has been observed for several other four-helix proteins (Stenkamp et al., 1983). The molecular shape of C5a 1–63 is close to that of an oblate ellipsoid with long and short axes of approximately 33 and 22 Å, respectively. As discussed with the secondary structure of C5a, helices I, II, and IV are quite regular and are compatible with an α-helical pitch. The N-terminal helix is amphiphilic, and its hydrophobic residues are tightly interdigitated with hydrophobic residues on helices II and IV. The internal residues of helices II, III, and IV are all in close contact with the side chain of the central Phe-51 residue, and the relative positioning of these helices is stabilized by disulfide linkages between residues 21 and 47, 22 and 54, and 34 and 55. These disulfide linkages follow from the NMR structure determination involving only NOEs as well as from biochemical studies on porcine C5a (Zimmerman & Vogt, 1984).

In contrast to the helical folding, the interconnecting loop regions are less well defined. A relatively tight turn of five residues connects helices I and II. The side chain of one of the residues in this turn, Tyr-13, is buried and appears to stabilize the turn; the other side chains of the turn are external. A long, poorly defined loop of eight residues connects helix II with helix III. It contains mainly hydrophilic residues except for Val-28, which is in close contact with residues at the C-terminus of helix IV. The odd cysteine in C5a, Cys-27, is located in this loop. It has been protected by a glutathione moiety in the studied protein and is pointing into solution in the calculated structures. The backbone fold of the turn between helices II and III is not completely undefined; differences in sequential NOE intensities (see Table I) in this region indicate the presence of stable structure. Evidence is obtained for the presence of a tight turn between residues 41 and 43 in the loop between helix III and helix IV from a strong $d_{\alpha N}(i,i+2)$ NOE between those residues. Residue 41 (Ile) is internal as evidenced by many NOEs from this residue with core residues (Figure 9A). Of the other hydrophobic residues in this loop, Leu-43 is at least partially exposed to solvent. The side chains of Arg-40 and Ser-42 are exposed to solvent while the aliphatic part of the side chain of Arg-46 contacts Val-17 and is hence partly buried. At this stage of refinement, no specific structural role for the Gly–Pro combination at positions 44 and 45 in the loop between helices III and IV can be recognized. These amino acids are conserved in human, bovine, porcine, rat, and murine C5a [anaphylatoxin sequences have been aligned and compared by Greer (1986)]. However, an amino acid side chain at position 44 would point toward the disulfide linkage between residues 47 and 21 in the core of the molecule and may therefore not be tolerated.

Exposed hydrophobic residues in the structure of C5a are Ala-11, Ala-39, and Val-56, while Val-17, Leu-43, and Ile-48 are partially buried. As is the case with Tyr-13 and Phe-51, also the partially buried Tyr-23 plays a structural role; it stabilizes the interaction between helix I and the core of the molecule (see Figure 9). The positions of the tyrosyl residues in the structure is in full agreement with data on nitration of these residues: Tyr-23 is preferentially nitrated while Tyr-13 is nitrated very slowly (Johnson & Chenoweth, 1985; J. Henkin, personal communication). Both histidine residues (at sequence positions 15 and 67) are exposed to solvent. The function of a conserved Gly residue at position 25 is very clear; a side chain at this position would collide with the central Phe residue (see Figure 9). All hydrophilic and charged amino acids are facing the solvent in C5a with the possible exception of Arg-37, from which the aliphatic chain has NOE contacts with the core Phe residue. Interestingly, this residue is conserved in all known C5a (and C3a) sequences and therefore likely to be involved in a structural role, perhaps by forming a salt bridge with an anionic side chain. A candidate residue for this interaction is Glu-32, which is conserved in all known C5a sequences (but not in the other anaphylatoxins).

Comparison with the C3a Structure. The conformation of C5a is closely related to that of C3a, for which a crystal structure has been determined (Huber et al., 1980). As in C5a, helices II, III, and IV of C3a are folded around a central phenylalanine residue. In Figure 10 we have superimposed the backbone of C3a, as obtained from further refined coordinates which were kindly supplied by Dr. J. Deisenhofer (Max-Planck-Institut, München), with the ensemble of C5a structures. It is found that the sequence and spatial locations of helices II, III, and IV in C5a correspond very closely to those of the helices in C3a. Deviations are found for the loop regions especially between helices II and III, where the C3a backbone follows a path that lies outside of the ensemble of C5a structures. This is likely to be due to an amino acid insertion in C3a in this region. Apparently large differences between the structures exist for the terminal regions. In C5a in solution, the C-terminus is only poorly defined, where for C3a in the crystal a regular helix is found followed by a peptide in a stable β-turn-type conformation. In C5a in solution a docked N-terminal helix is observed which corresponds to a disordered region in the crystal for C3a. However, a NMR investigation of the secondary structure of C3a (Nettesheim et al., 1988) has shown that in solution disorder occurs for the C-terminus and that helicity prevails for the N-terminus, just as observed for C5a. Thus, the differences in biological activities between C3a and C5a (see the introduction) are very unlikely to be caused by major conformational differences between the two

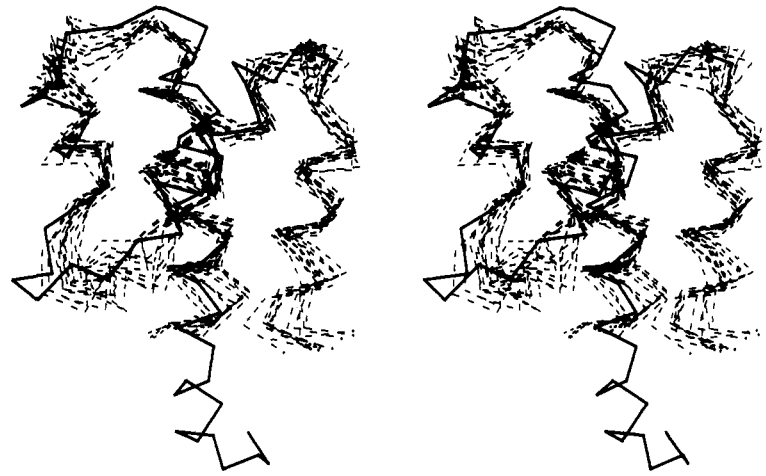

FIGURE 10: Superposition of the NMR ensemble obtained from all 346 constraints with the C3a crystal structure (solid line) after minimization of the rms deviation of the C_α positions between the helices of this structure and structure A4.

molecules. Rather, the differentiation in biological activities is due to the differences in the chemical characteristics of their surface residues, which are very different indeed (Greer, 1985, 1986).

A superposition of a predicted C5a structure (Greer, 1985) with an NMR ensemble obtained from DISMAN structure calculations with a slightly smaller distance constraints set was recently presented and extensively discussed elsewhere (Zuiderweg et al., 1988b). The conclusions drawn here on the basis of the present structures are identical; the predicted N-terminal and core conformation follow closely the experimental ensemble while major differences occur between the structures in the C-terminal part, where for the model structure a stable helix was predicted. Since part of the C5a receptor binding site is located in the C-terminal part of this region, this difference is very important for the understanding of C5a function.

Inferences for C5a Structure–Function Relationships. Typically, biological assays of C5a are carried out under physiological conditions. We have therefore assessed the structural features of C5a in solution at more neutral pH conditions. As described under Results, the secondary structure of C5a is independent of pH in the range pH 2.3 to pH 6.5. The fast exchange of amide protons in the C-terminal region at neutral pH values in H_2O solutions shows disorder for this part of the molecule at these conditions as well. Moreover, key NOEs defining the crude folding of the protein were unchanged at higher pH values. Thus, although no precise structure determination was carried out at higher pH values, no indication has been found for major conformational changes in the molecule at more neutral pH values. Therefore, it seems justified that inferences for C5a biological activity can be made from the structure obtained at low pH.

Regions of C5a that interact with the receptor are well dispersed in the protein. The C-terminus of the molecule is of importance for its biological activity as follows from the fact that removal of the C-terminal arginyl residue results in a large decrease of receptor binding affinity (Fernandez et al., 1978; Gerard et al., 1979; Chenoweth & Hugli, 1978). However, upon complete removal of 11 C-terminal residues, C5a (1–63) retains a significant fraction of its receptor-binding affinity (10^{-6} M versus 10^{-11} M for native C5a; Edalji et al., 1987). Also, removal of N-terminal residues reduces receptor affinity (Damerau et al., 1983; Gerard et al., 1985; Edalji et al., 1987), and the disulfide linked core has been implicated as a binding site as well (Johnson et al., 1987; Johnson & Chenoweth, 1985).

We have addressed the question whether the core of the C5a molecule tends to stabilize a specific conformation of the C-terminus of the molecule by intramolecular interactions. Although no long-range NOEs were observed between these parts of the molecule (see above), it would still be feasible that weak interactions take place which do not give rise to NOEs, but which could affect the chemical shifts of core residues involved in a putative weak binding interaction. Therefore, we carried out an NMR investigation of C5a (1–63). Although the removal of the 11 C-terminal residues did negatively affect the long-term stability of the protein, no significant chemical shift changes were detected for the remaining residues in the 2D spectra of fresh solutions of this derivative, except for those close to the clipping site (Edalji et al., 1987; Nettesheim et al., unpublished data). These data indicate that the C-terminus does not dock preferentially on any area of the C5a molecule and is therefore highly unlikely to be biased toward any other structure than the observed residual helicity by the core presence. Conversely, the 10^5 M loss in receptor affinity of C5a as caused by the enzymatic removal of the C-terminus is not due to a conformational change of the core. It appears therefore that receptor binding sites located in the C-terminus and in the remainder of the molecule are independent of each other.

Enzymatic removal of eight N-terminal residues also resulted in a marked decrease of receptor binding affinity. However, changes in NOEs for the central phenylalanine residue were detected for this derivative by one-dimensional difference NOE techniques (Edalji et al., 1987). This apparent change in core conformation is easily understood from the structure of C5a: disruption of the N-terminal helix would affect the interaction between this helix and the core of the molecule and is therefore likely to cause changes in atomic positions. Thus, the decrease of receptor affinity as caused by modifications at the N-terminus does not necessarily infer that this part of the molecule is a receptor binding site. This notion explains the apparent discrepancy between the resulting decrease in affinity upon enzymatic removal of the N-terminus (Damerau et al., 1983; Gerard et al., 1985; Edalji et al., 1987) and the results of site-specific mutagenesis experiments (Mollison et al., 1988a) where it was found that changes in the amino terminus did not affect affinity.

Although no conclusive evidence has been obtained as yet for the location(s) of receptor binding residues in the core of C5a, the present structural data can identify residues which are likely candidates for the interaction. As described above, no major differences are found between the solution confor-

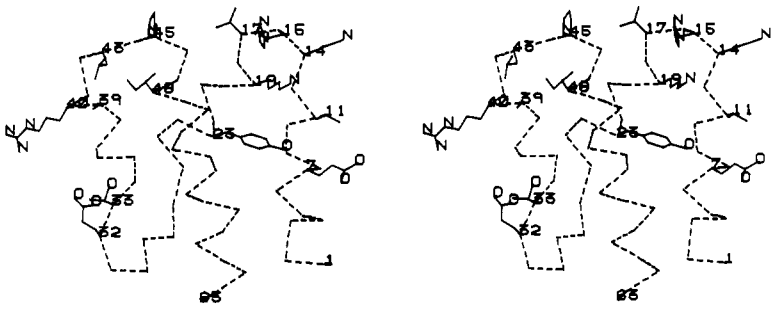

FIGURE 11: Location of core residues which may interact with the C5a receptor. The selected residues Q7, A11, K14, H15, V17, K19, Y23, E32, T33, A39, R40, L43, P45, and I48 fulfill three criteria: (i) location on the solvent interface; (ii) chemically conserved in five C5a species; (iii) chemically different from the corresponding residues in four C3a species.

mations of C5a and C3a. Nevertheless, C5a and C3a bind to different receptors. As a further difference, all receptor interaction energy is located in the C-terminus of C3a (Lu et al., 1984), in contrast to C5a. Therefore, with use of a strategy previously proposed (Greer, 1986), surface residues in the core of C5a which are chemically conserved within the family of C5a species, but which are chemically different from the corresponding residues in the family of C3a species, should be viewed as candidates for the additional interaction of C5a with its receptor. Comparisons were made for this purpose between the sequences of human (Fernandez & Hugli, 1978), porcine (Gerard & Hugli, 1978), rat (Cui et al., 1985), bovine (Gennaro et al., 1986), and murine (Wetsel et al., 1987) C5a and human (Hugli, 1975), porcine (Corbin & Hugli, 1976), rat (Jacobs et al., 1978), and murine (Domdey et al., 1982) C3a, and the resulting candidate residues were edited against the C5a structure. N-terminal residues which fulfill all three criteria are Glu-7 and Ala-11, but they were shown not to be involved in receptor interaction (Mollison et al., 1988a). Core residues of interest according to the criteria are Lys-14, His-15, Val-17, Lys-19, Glu-32, Thr-33, Ala-39, Arg-40, Leu-43, and Ile-48. Tyr-23, which is partially exposed but which also has a function in stabilizing the helix I–helix II interaction, and Pro-45, which is fully exposed but which is likely to have an as yet undetermined structural function (see above), may be added to this list. These residues are shown in Figure 11, from which it is seen that no clear contiguous region for these residues emerges from this analysis. However, the majority of external, C5a-conserved residues which differ from C3a map in the "top" of the molecule. Some of these residues were recently targeted by site-specific mutagenesis techniques combined with NMR structural studies (Mollison et al., 1988a). Interestingly, a small but significant decrease in receptor affinity was found for a species in which Arg-40 was replaced with Gly, in agreement with the above analysis that the "top" of the molecule may be involved in the additional C5a receptor interaction. Other mutants in this region are currently being prepared to further test this hypothesis.

In conclusion, the tertiary structure of rCa[Met0] has been determined with high precision for the backbone of the helical regions in the protein (rmsd < 0.8 Å) while the side chains of the internal helical residues are obtained with a precision that is slightly worse (1 Å). Considerable variation in structure is found for the loop regions in the members of the NMR ensemble; nevertheless, the results clearly indicate which residues are internal and which are external in these areas. The overall 3D structure for C5a was found to be very closely related to that of C3a. The 3D structural information is currently being utilized to direct and interpret site-specific mutagenesis studies on the protein.

Acknowledgments

We thank Dr. W. Mandecki for the synthesis of the C5a DNA, Dr. J. Henkin, M. Seavy, L. Fayer, and R. P. Edalji for their involvement in the expression and purification of the recombinant C5a protein, and Dr. J. Greer and Dr. S. W. Fesik for the critical reading of the manuscript.

Supplementary Material Available

One table containing the complete list of distance constraints used in the structure calculations, two tables listing the mutual rms deviations in the structural ensemble calculated with NOEs only and in the structural ensemble calculated with all constraints, and four figures showing NOESY spectra of C5a at two mixing times and showing the fingerprint region of scalar correlated spectra of C5a in H_2O and in ^2H_2O, freshly lyophilized from H_2O (12 pages). Ordering information is given on any current masthead page.

Registry No. Complement component C5a, 80295-54-1.

References

Aue, W. P., Bartholdi, E., & Ernst, R. R. (1976) *J. Chem. Phys. 64*, 2229.

Billeter, M. Braun, W., & Wüthrich, K. (1982) *J. Mol. Biol. 155*, 321.

Boelens, R., Scheek, R. M., Dijkstra, K., & Kaptein, R. (1985) *J. Magn. Reson. 62*, 378.

Brant, D. A., & Flory, P. J. (1967) *J. Mol. Biol. 23*, 47.

Braun, W., & Go, N. (1985) *J. Mol. Biol. 186*, 611.

Bundi, A., & Wüthrich, K. (1979) *Biopolymers 18*, 285.

Carter, G. W., Mollison, K. W., Fayer, L., Fey, T., Krause, R., Henkin, J., & Edalji, R. P. (1985) *Complement 2*, 15.

Chenoweth, D. E., & Hugli, T. E. (1978) *Proc. Natl. Acad. Sci. U.S.A. 75*, 3943.

Chenoweth, D. E., Erickson, B. W., & Hugli, T. E. (1979) *Biochem. Biophys. Res. Commun. 86*, 227.

Clore, G. M., Gronenborn, A. M., Brünger, A. T., & Karplus, M. (1985) *J. Mol. Biol. 186*, 435.

Clore, G. M., Nilges, M., Sukumaran, D. K., Brünger, A. T., Karplus, M., & Gronenborn, A. M. (1986) *EMBO J. 5*, 2729.

Corbin, N. C., & Hugli, T. E. (1976) *J. Immunol. 117*, 990.

Cui, L.-X., Ferreri, K., & Hugli, T. E. (1985) *Complement 2*, 18.

Damerau, B., Zimmerman, H., Wustefeld, H., Czarniak, K., & Vogt, W., (1983) *Immunobiology (Stuttgart) 164*, 229.

Delsuc, M. A., & Lallemand, J. Y. (1986) *J. Magn. Reson. 56*, 343.

Domdey, H., Wiebauer, K., Kazmaier, M., Muller, V., Odink, K., & Fey, G. (1982) *Proc. Natl. Acad. Sci. U.S.A. 79*, 7619.

Edalji, R. P., Mollison, K. W., Zuiderweg, E. R. P., Fey, T. A., Krause, R. A., Conway, R. G., Miller, L., Lane, B., Henkin, J., Greer, J., & Carter, G. W. (1987) *Fed. Proc., Fed. Am. Soc. Exp. Biol. 46*, 980.
Englander, S. W., Downer, S., & Teitelbaum, H. (1972) *Annu. Rev. Biochem. 41*, 903.
Fernandez, H. N., & Hugli, T. E. (1978) *J. Biol. Chem. 253*, 6955.
Fernandez, H. N., Henson, P. M., Otani, A., & Hugli, T. E. (1978) *J. Immunol. 120*, 109.
Gampe, R. T., Jr., Connolly, P. J., Rockway, T., & Fesik, S. W. (1988) *Biopolymers 27*, 313.
Gennaro, R., Simonic, T., Negri, A., Mottola, C., Secchi, C., Ronchi, S., & Romeo, D. (1986) *Eur. J. Biochem. 155*, 77.
Gerard, C., & Hugli, T. E. (1978) *J. Biol. Chem. 255*, 4710.
Gerard, C., Chenoweth, D. E., & Hugli, T. E. (1979) *J. Reticuloendothel. Soc. 26*, 711.
Gerard, C., Showell, H. J., Hoeprich, P. D., Jr., Hugli, T. E., & Stimler, N. P. (1985) *J. Biol. Chem. 260*, 2613.
Greer, J. (1985) *Science (Washington, D.C.) 228*, 1055.
Greer, J. (1986) *Enzyme 36*, 150.
Haasnoot, C. A. G., Van de Ven, F. J. M., & Hilbers, C. W. (1984) *J. Magn. Reson. 56*, 343.
Huber, R., Scholze, H., Paques, E. P., & Deisenhofer, J. (1980) *Hoppe-Seyler's Z. Physiol. Chem. 361*, 1389.
Hugli, T. E. (1975) *J. Biol. Chem. 250*, 8293.
Hugli, T. E. (1981) *CRC Crit. Rev. Immunol. 1*, 321.
Jacobs, J. W., Rubin, J. S., Hugli, T. E., Bogardt, R. A., Mariz, I. K., Daniels, J. S., Daughaday, W. H., & Bradshaw, R. A. (1978) *Biochemistry 17*, 5031.
Jeener, J., Meier, B. H., Bachmann, P., & Ernst, R. R. (1979) *J. Chem. Phys. 71*, 4546.
Johnson, R. J., & Chenoweth, D. E. (1985) *J. Biol. Chem. 260*, 10339.
Johnson, R. J., Tamerius, J. D., & Chenoweth, D. E. (1987) *J. Immunol. 138*, 3856.
Kaptein, R., Zuiderweg, E. R. P., Scheek, R., Boelens, R., & Van Gunsteren, W. F. (1985) *J. Mol. Biol. 182*, 179.
Khan, S. A., Erickson, B. W., Kawahara, M. S., & Hugli, T. E. (1985) *Complement 2*, 42.
Lu, Z.-X., Fok, K.-F., Erickson, B. W., & Hugli, T. E. (1984) *J. Biol. Chem. 259*, 7367.
Macura, S., Wüthrich, K., & Ernst, R. R. (1982) *J. Magn. Reson. 46*, 269.
Mandecki, W., Mollison, K. W., Bolling, T. J., Powell, B. S., Carter, G. W., & Fox, J. L. (1985) *Proc. Natl. Acad. Sci. U.S.A. 82*, 3543.
Mandecki, W., Powell, B. S., Mollison, K. W., Carter, G. W., & Fox, J. L. (1986) *Gene 43*, 131.
Mayer, M. M. (1979) *Rev. Infect. Dis. 1*, 483.
Mollison, K. W., Fey, T. A., Krause, R. A., Mandecki, W., Fox, J. L., & Carter, G. W. (1987) *Agents Actions 21*, 366.
Mollison, K. W., Mandecki, W., Zuiderweg, E. R. P., Fayer, L., Fey, T. A., Krause, R. A., Conway, R. G., Miller, L., Edalji, R. P., Shallcross, M. A., Lane, B., Fox, J. L., Greer, J., & Carter, G. W. (1988a) *Proc. Natl. Acad. Sci. U.S.A.* (in press).
Mollison, K. W., Edalji, R. P., Fey, T. A., Krause, R. A., Conway, R. G., & Carter, G. W. (1988b) *Fed. Proc., Fed. Am. Soc. Exp. Biol. 2*, 1604.
Müller, L., & Ernst, R. R. (1979) *Mol. Phys. 38*, 963.
Nettesheim, D. G., Edalji, R. P., Mollison, K. W., Greer, J., & Zuiderweg, E. R. P. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 5036.
Otting, G., Widmer, H., Wagner, G., & Wüthrich, K. (1986) *J. Magn. Reson. 66*, 187.
Pearson, G. A. (1977) *J. Magn. Reson. 27*, 265.
Stenkamp, R. E., Sieker, L. C., & Jensen, L. H. (1983) *Acta Crystallogr., Sect. B: Struct. Sci. B39*, 697.
Wagner, G. (1983) *Q. Rev. Biophys. 16*, 1.
Wagner, G. (1984) *J. Magn. Reson. 57*, 497.
Wagner, G., Pardi, A., & Wüthrich, K. (1983) *J. Am. Chem. Soc. 105*, 5948.
Ward, P. A. (1970) *Arthritis Rheum. 13*, 181.
Weigle, W. D., Goodman, M. G., Morgan, E. L., & Hugli, T. E. (1983) *Springer Semin. Immunopathol. 6*, 173.
Wetsel, R. A., Ogata, R. T., & Tack, B. F. (1987) *Biochemistry 26*, 737.
Wider, G., Macura, S., Anil Kumar, Ernst, R. R., & Wüthrich, K. (1984) *J. Magn. Reson. 56*, 207.
Williamson, M. P., Havel, T. F., & Wüthrich, K. (1985) *J. Mol. Biol. 182*, 295.
Wüthrich, K., Billeter, M., & Braun, W. (1983) *J. Mol. Biol. 169*, 949.
Wüthrich, K., Billeter, M., & Braun, W. (1984) *J. Mol. Biol. 180*, 715.
Zimmerman, B., & Vogt, W. (1984) *Hoppe-Seyler's Z. Physiol. Chem. 365*, 151.
Zuiderweg, E. R. P., Kaptein, R., & Wüthrich, K. (1983a) *Proc. Natl. Acad. Sci. U.S.A. 80*, 5837.
Zuiderweg, E. R. P., Kaptein, R., & Wüthrich, K. (1983b) *Eur. J. Biochem. 137*, 279.
Zuiderweg, E. R. P., Boelens, R., & Kaptein, R. (1985a) *Biopolymers 24*, 601.
Zuiderweg, E. R. P., Scheek, R. M., Boelens, R., Van Gunsteren, W. F., & Kaptein, R. (1985b) *Biochimie 67*, 707.
Zuiderweg, E. R. P., Hallenga, K., & Olejniczak, E. T. (1986) *J. Magn. Reson. 70*, 336.
Zuiderweg, E. R. P., Mollison, K. W., Henkin, J., & Carter, G. W. (1988a) *Biochemistry 27*, 3568.
Zuiderweg, E. R. P., Henkin, J., Mollison, K. W., Carter, G. W., & Greer, J. (1988b) *Proteins: Struct. Funct., Genet. 3*, 139.