Articles

Three-Dimensional Structure of Porcine $C5a_{desArg}$ from ^1H Nuclear Magnetic Resonance Data

Michael P. Williamson*,‡ and Vincent S. Madison§

Physical Methods Department, Roche Products Ltd., P.O. Box 8, Welwyn Garden City, Herts, U.K. AL7 3AY, and Physical Chemistry Department, Hoffmann-La Roche, Inc., Nutley, New Jersey 07110

Received June 14, 1989; Revised Manuscript Received November 9, 1989

ABSTRACT: Two-dimensional nuclear magnetic resonance spectra of porcine $C5a_{desArg}$ (73 residues) have been used to construct a list of 34 hydrogen bonds, 27 dihedral angle constraints, and 151 distance constraints, derived from nuclear Overhauser effect data. These constraints were used in restrained molecular dynamics calculations on residues 1–65 of C5a, starting from a folded structure modeled on the crystal structure of a homologous protein, C3a. Forty-one structures have been calculated, which fall into three similar families with few violations of the imposed constraints. Structures in the most populated family have a root-mean-square deviation from the average structure of 1.02 Å for the C^α atoms, with good definition of the internal residues. There is good agreement between the calculated structures and other nuclear magnetic resonance data. The structure is very similar to that recently reported for human C5a [Zuiderweg et al. (1989) *Biochemistry 28*, 172–185]. Some biological implications of these structures are discussed.

The complement system is an essential component of host immune defense and is made up of about 20 serum proteins that interact sequentially to produce a variety of proinflammatory and cytolytic effects (Jose, 1987). One of these is the anaphylotoxin C5a, which is cleaved from the much larger protein C5 during activation of the complement system. The C-terminal arginine residue of C5a is rapidly hydrolyzed by a specific carboxypeptidase to yield $C5a_{desArg}$, which appears to be the physiologically active form of the molecule. It has many potent proinflammatory activities, including chemotaxis and cellular adhesion of neutrophils, stimulation of neutrophils and macrophages, and stimulation of the immune system (Goldstein, 1988; Hugli, 1981) and has been implicated in several inflammatory conditions, including rheumatoid arthritis, psoriasis, and adult respiratory distress syndrome. An antagonist of C5a would therefore be potentially a very useful therapeutic agent.

The sequences of C5a proteins from several species are known, including human (Fernandez & Hugli, 1978) and porcine (Gerard & Hugli, 1980). All the C5a sequences contain roughly 75 residues, and some (including human) are glycosylated near the C-terminus. All known C5a sequences contain six or seven half-cystines, for which a linkage has been deduced (Zimmerman & Vogt, 1984) and confirmed by direct synthesis (Chino et al., 1988). There is no crystal structure of C5a, although there is a crystal structure of the homologous protein C3a (Huber et al., 1980), which has been used to construct a model for human C5a (Greer, 1985). Solution structures of human (Zuiderweg et al., 1989) and bovine (Zarbock et al., 1988) C5a have been deduced, using NMR[1]-based constraints.

We have recently described the assignment of the ^1H NMR spectrum of porcine C5a and its secondary structure in solution (Williamson, 1989). On the basis of this assignment, we here present a large number of structural constraints and use them to derive a solution structure of porcine C5a, described as an ensemble of 41 closely similar structures. The structures were calculated by restrained molecular dynamics and energy minimization, starting from a model that was derived from the crystal structure of human C3a. The structures are compared, and their biological implications are discussed.

Materials and Methods

$C5a_{desArg}$ was isolated from fresh porcine blood as previously described (Williamson, 1989). The NMR experiments were carried out with approximately 3 mM $C5a_{desArg}$ in water at low pH, at 500 and 400 MHz. Further details are given in Williamson (1989) and in the caption to Figure 1.

C5a structures were displayed and manipulated by using the Roche Interactive Molecular Graphics (RIMG) software package (Müller et al., 1986), and energy minimization and molecular dynamics trajectories were calculated by using the program CHARMM (version 19) (Brooks et al., 1983), in a manner similar to that described (Brünger et al., 1987). The CHARMM version 18 parameters and energy functions were used with slight modifications (Fry et al., 1989). For pairwise electrostatic energies, the dielectric constant was numerically equal to r (in angstroms), the distance between the charges. Reduced formal charges of ±0.25 were used for the nonneutral chain termini and side chains of Asp and Lys residues. Minimization was conducted by using 200 steepest descent steps followed by assumed-basis Newton–Raphson steps until

* Address correspondence to this author at the Department of Molecular Biology and Biotechnology, University of Sheffield, Sheffield, U.K. S10 2TN.
‡ Roche Products Ltd.
§ Hoffmann-La Roche, Inc.

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

0006-2960/90/0429-2895$02.50/0

the energy converged to 0.01 kcal/mol (up to 5000 steps). Calculations were carried out on a VAX 11/750 for the original model and on a VAX 8800 for restrained dynamics and minimization.

The porcine C5a model was constructed starting from an energy-minimized human C3a crystal structure. The C5a sequence contains one deletion and one insertion relative to C3a, both of which occur in loops. These were modeled by keeping the new loops as close as possible to those in C3a. Residue Phe34 of C3a was deleted and the chain joined. Residues Cys27 and Cys34 were then constrained, and the loop was minimized. Similarly, an extra lysine residue was added between Tyr15 and Pro16 of C3a and the loop subjected to restrained minimization. Both new loops were geometrically correct and had low energies. The amino acid sequence of C3a was then replaced by the porcine C5a sequence, keeping the maximum number of heavy atoms in the same positions, and adding extra atoms in classical trans/gauche conformations. The N-terminus of C5a was made as a regular type III α-helix and docked onto the rest of the protein with residues Leu2 and Ile6 interdigitating hydrophobic residues in helices B and D. This necessitated changes to the dihedral angles ψ of residues 15 and 16 and ϕ of residue 16. The resulting structure was relaxed, to give a low-energy strain-free structure. The model appears very similar to that constructed by Greer (1987).

NOEs were measured from NOESY spectra acquired by using a mixing time of 200 ms. It was ascertained that the degree of spin diffusion in these spectra was low, from the observation that the intensities of NOEs to methylene pairs were often very different. If extensive spin diffusion were occurring, the NOEs would have similar intensity (Williamson, 1987; Neuhaus & Williamson, 1989). The intensity of NOESY cross peaks was measured crudely by counting contour levels, and cross peaks were classified as strong, medium, or weak. For cross peaks involving methyl groups, one contour level was subtracted before classification, because of the inherently higher intensity of these cross peaks. The distance calibrations were made by consideration of standard distances in protein structures (Wüthrich, 1986). Most sequential NOEs between $C^{\alpha}H$ of residue i and NH of residue $i + 1$ were present, many as weak cross peaks. The maximum distance between these protons is 3.6 Å, implying a cutoff distance for weak cross peaks of <3.6 Å (1 Å = 0.1 nm). To allow for other effects that might influence cross-peak intensities (for example, motional averaging and peak width variation) the cutoff was set at 4.0 +0.8/−1.5 Å. Similarly, the maximum intraresidue distance between $C^{\alpha}H$ and NH is 2.9 Å, and about $^2/_3$ of these cross peaks were of at least medium intensity. The cutoff for medium was therefore set at 3.0 +0.5/−1.0 Å. The cutoff for strong NOEs was set at 2.5 ± 0.5 Å, based on an approximate r^{-6} dependence. Amide protons appeared to need different calibration, judging by the large number of weak cross peaks found in regular helical stretches, and a cutoff of 3.5 +0.8/−1.5 Å was set for weak NH–NH cross peaks. At a late stage in structure refinement, distances corresponding to NOEs observed in regular helical stretches were set to those expected for regular α-helical structure, i.e., 2.8 ± 0.5 Å for NH–NH NOEs, 3.4 +0.5/−0.8 Å for $d_{\alpha N}(i,i+3)$ NOEs, and 2.8 +0.5/−0.8 Å for $d_{\alpha\beta}(i,i+3)$ NOEs. A single average distance for diastereotopic protons in nonstereospecifically assigned methylene and methyl groups was obtained as $\langle r^{-6}\rangle^{-1/6}$. For prochiral methyl groups that gave separate resonances but could not be assigned stereospecifically from the NMR data, an algorithm was applied at each iteration of the optimization that chose the stereospecific assignment that gave the best fit to the distance constraints.

The distance constraints were converted to energy constraints by adding the additional term

$$E = (\mathrm{WNOE})(kT/2)[(r - r_0)/\delta r]^{\mathrm{ENOE}}$$

where WNOE = a weighting factor with values from 0.01 to 3.0, k = the Boltzmann constant, T = absolute temperature (300 K), r = interproton distance or $\langle r^{-6}\rangle^{-1/6}$ average for topologically equivalent groups, r_0 = NOE-defined distance, δr = experimental uncertainty in distance, and ENOE = an exponent with values of 2 or 4. Note that in general (as described above in detail) δr takes on different values for $r < r_0$ and $r > r_0$, because of motional averaging effects.

All amide bonds (including side chains) were constrained to be trans, using a force constant of 20 kcal/(mol·rad^2). The amide bond to proline 45 was assumed to be trans, on the basis of the lack of an NOE between Pro45 $C^{\alpha}H$ and Gly44 $C^{\alpha}H$, and between Pro45 $C^{\alpha}H$ and Gly44 NH. The three disulfide bridges were linked by covalent bonds. The three-bond coupling constant between NH and $C^{\alpha}H$ was used to restrict allowed values of the dihedral angle ϕ (Williamson et al., 1985). It was assumed to be very small where no NH–α COSY cross peak was found and small where the cross peak was of low intensity; the ϕ angles in these cases were constrained to −60°, with force constants of 10 and 5 kcal/(mol·rad^2), respectively. For those residues with only one β proton, the coupling constant between $C^{\alpha}H$ and $C^{\beta}H$ was measured in a high-resolution phase-sensitive double-quantum-filtered COSY spectrum, using cross sections along the ω_2 axis (Marion & Wüthrich, 1983), and using inverse Fourier transformation and zero filling to 16 384 data points, giving a final digital resolutionof 0.4 Hz/pt. Values of greater than 10 Hz were taken to indicate a trans orientation and constrained to 180° with a force constant of 10 kcal/(mol·rad^2).

Hydrogen bonds were defined by distances O–N of 2.9 ± 0.5 Å and O–H of 1.9 ± 0.3 Å. Initially, 21 hydrogen bonds were defined, corresponding to slowly exchanging amide protons in well-defined α-helical structure (residues 16–27, 34–39, and 45–59). At a later stage in the structure determination, other hydrogen bonds were added. These hydrogen bonds are not necessary for maintaining the chain fold, which is determined by NOEs (Williamson et al., 1985; Kline et al., 1988), but are useful in that they provide tighter constraints than the NOEs and keep the local secondary structure better formed. They corresponded to hydrogen bonds that were found frequently in structures and also involved slowly exchanging amide protons, with the addition of several α-helix-type hydrogen bonds in the N-terminal helix.

Starting from the C5a model, the angle constraints were applied and NOE constraints were added gradually, increasing WNOE from 0.01 to 0.1, reminimizing between steps. Molecular dynamics was initiated from the minimized structure in two steps: heating to 1000 K over 10 ps, and equilibration at 1000 K for a total of 50 ps, altering WNOE and ENOE in a stepwise manner, as described (Fry et al., 1989). Subsequently, each optimized structure was obtained from 20 ps of dynamics at 1000 K (continuing from the end of the trajectory from the previous step at 1000 K), 15 ps of linear quenching to near 0 K, and minimization. The dynamics at 1000 K consisted of four segments of 5 ps each with WNOE, ENOE = 0.3, 2; 3.0, 2; 0.6, 4; and 3.0, 4. The gradual increase in WNOE and ENOE has been shown to produce low NOE violations, by allowing considerable flexibility at the start of the trajectory while ending up with heavy penalties for large violations (Fry et al., 1989). The quenching step used the

values WNOE = 3.0, ENOE = 4, and the minimization used WNOE = 3.0, ENOE = 4, followed by WNOE = 1.0, ENOE = 4. This last weighting approximates a square well potential with a width given by the uncertainties in the NOE-defined distances.

The conformations generated were analyzed for similarities on the basis of RMS fits of α-carbons and classified into families (Fry et al., 1989). Conformers were first sorted in order of increasing energy. The lowest energy conformer was the first member of the first family. The next conformer was added to the first if it fitted with an RMS cutoff of less than 2 Å; otherwise it became the first member of the second family. Each subsequent conformer was compared to the defined families in order and became a member of the first family that it fitted within the cutoff; otherwise it was placed in a new family. For comparisons within and between families, an average structure was calculated (of no physical meaning), which was compared to average structures from other families and to individual conformers within the family. Variation in local conformation was probed by taking four-residue segments, calculating the average C^α positions for the segment, and then calculating RMS deviations from the average for each conformer and averaging. RMSD values of the dihedral angle ϕ were calculated as

$$\mathrm{RMSD} = [(1/N)\sum_{i=1}^{N}(\phi_i - \phi_i^0)^2]^{1/2}$$

RMS violations of distance constraints were calculated with respect to the upper and lower limits of the distance constraints, by using the complete set of 151 NOE constraints and 68 hydrogen bond constraints (34 hydrogen bonds). Several low-energy members of each family were used to calculate ring-current shifts [using the parameters of Johnson and Bovey (1958)] and NOE contacts, by using the programs VNMR and DISCALC (Hoch, 1983), respectively.

Results

Input Constraints. Figure 1 shows part of the NOESY spectrum of porcine $C5a_{desArg}$ in H_2O, with an indication of some of the sequential NOEs present, which were used during the sequential assignment. By use of these assignments, NOESY spectra were systematically searched and used to generate a list of all assignable NOEs, plus a number of other NOESY cross peaks whose assignment was ambiguous. Many of the assigned NOEs were intraresidue or sequential and were not used as input to the structure calculations because they were not informative with the rather crude intensity quantitation used. This left 131 assigned NOEs in addition to the ambiguous ones. These NOEs were then compared to those expected from the model, to determine whether the model would be a suitable starting point for restrained dynamics calculations. Figure 2 compares the predicted NOEs with those used for the final calculations (which are not very different from the initial 131 NOEs) and demonstrates that the model fits the observed NOEs quite well, although the turn between helices A and B and the placement of helix A are not correct in detail.

The model was therefore optimized by using restrained molecular dynamics and energy minimization. After several structures had been optimized, they were studied for short interproton distances and for violations of NOE constraints, to see if any of the ambiguous NOEs could be identified or if any of the existing NOE constraints seemed inconsistent with the structures. The constraint list was then altered appropriately, and structures were recalculated. This recursive procedure was carried out several times, and toward the end of this procedure, some sequential NOE constraints were tightened and additional hydrogen bond constraints were added (vide supra). It should be pointed out that these additional constraints were only added when it was clear that the correct global folding was already established. The complete list of 153 NOEs was made up of 19 intraresidue, 48 sequential NH–NH and 13 other sequential, 4 (i,i+2), 16 $C^\alpha H(i)$ to $NH(i+3)$, 3 $C^\alpha H(i)$ to $C^\beta H(i+3)$, 12 other (i,i+3), and 38 long-range NOEs, summarized in Figure 2. Two incorrect long-range NOEs were included in the list at this stage (vide infra) and are omitted from Figure 2. This final set of constraints was used in deriving all of the structures discussed below.

The two methyl groups of Val48 were distinguishable and had very different NOE contacts. Measurement of the coupling constant between $C^\alpha H$ and $C^\beta H$ (10.2 Hz) showed that the two protons are approximately trans. With this information, the NOEs observed could be used to assign the two methyl groups stereospecifically, as shown in Zuiderweg et al. (1985). The two methyl groups of Leu18 were also distinguishable but could not be assigned stereospecifically from a simple analysis of intraresidue NOEs and coupling. In the course of optimizing each structure, a particular chiral assignment was made for the NOEs from each of the Leu18 methyl groups, as described under Materials and Methods.

The torsion angle and hydrogen bond constraints applied are summarized in Figure 3. The complete list of distance, angle, and hydrogen bond constraints is available as supplementary material (see paragraph at end of paper regarding supplementary material).

During the course of the assignment (Williamson, 1989), it had become clear that the C-terminal helix becomes progressively less well-defined as the terminus is approached and that the C-terminus of the protein is disordered in solution. Accordingly, the eight C-terminal residues have been omitted from the calculations, and all further results relate to the first 65 residues, with the assumption that the remaining eight residues have little preferred structure in solution.

Restrained Dynamics Calculations and Structures. Restrained molecular dynamics calculations using 153 NOE constraints resulted in a set of 33 structures, all of low energy. However, inspection of the NOE violations revealed that two of the constraints were not met in any of the structures. On checking against NOESY spectra, it was clear that both NOESY cross peaks could be assigned differently, and so the two constraints were removed, to give the final set of 151 constraints. The 33 structures were reminimized by using the revised constraints, which caused an average RMS movement of 0.75 Å in the C^α positions, the changes being localized to the areas of the two deleted constraints. As a result of the reminimization, the energies of the structures were reduced by about 150 kcal/mol each, and the numbers and sizes of NOE violations were reduced substantially. A further 8 structures were calculated by restrained molecular dynamics using the revised constraints. These 8 structures were indistinguishable from the 33 reminimized structures, and all 41 structures will be considered together.

The structures all fit the NOE constraints reasonably well (Table I). The largest violation in any structure was 1.14 Å, but violations of greater than 0.8 Å were rare. All the violations larger than 0.8 Å came from three constraints on the N-terminal helix (given in Table II), which is probably indicative of difficulties in docking the N-terminal helix against the rest of the protein rather than incorrect constraints, as each

Table I: Violations of NOE Distance Constraints, by Structure[a]

structure	family	RMSD violation of distance constraints (Å)	number of violations in range								
			−0.4	−0.2	0	0.2	0.4	0.6	0.8	1.0	1.2
1	1	0.152	1	1	176	19	12	6	3	1	0
2	1	0.148	1	0	181	15	12	6	4	0	0
3	1	0.146	1	0	181	16	9	8	4	0	0
4	1	0.148	1	1	178	20	10	5	4	0	0
5	1	0.147	1	2	172	23	12	5	3	1	0
6	2	0.152	0	2	169	27	13	5	1	2	0
7	1	0.145	1	1	170	25	12	6	4	0	0
8	1	0.141	0	2	175	22	9	8	3	0	0
9	1	0.150	1	1	178	16	13	5	5	0	0
10	1	0.150	1	1	182	13	11	7	3	1	0
11	2	0.155	1	1	176	15	17	4	4	1	0
12	1	0.153	0	1	181	17	11	5	2	2	0
13	1	0.148	1	1	177	18	12	6	2	2	0
14	1	0.144	1	1	178	20	9	6	3	1	0
15	2	0.148	1	0	176	23	11	5	1	2	0
16	1	0.151	1	2	179	16	11	6	3	1	0
17	1	0.159	1	2	171	22	15	4	2	1	1
18	1	0.150	1	2	177	20	10	5	3	0	1
19	1	0.150	0	2	178	15	15	5	3	0	1
20	1	0.150	1	1	177	22	10	4	3	0	1
21	1	0.145	1	1	178	19	12	4	3	1	0
22	2	0.146	1	1	176	19	16	3	1	2	0
23	1	0.147	1	0	180	19	10	4	4	1	0
24	1	0.152	1	0	180	18	13	2	3	2	0
25	1	0.150	1	1	177	20	10	6	3	1	0
26	1	0.150	1	1	176	20	11	5	5	0	0
27	1	0.146	1	2	177	19	9	7	3	1	0
28	2	0.146	1	1	176	22	12	4	1	2	0
29	1	0.145	1	0	179	20	11	3	3	2	0
30	1	0.144	1	0	176	20	13	4	5	0	0
31	1	0.152	1	1	178	19	10	6	2	2	0
32	1	0.154	1	1	178	18	9	7	5	0	0
33	1	0.153	1	1	180	17	9	7	2	2	0
34	1	0.144	0	2	178	18	12	3	6	0	0
35	3	0.151	1	1	175	20	12	8	1	0	1
36	3	0.153	2	1	175	19	11	7	3	0	1
37	3	0.151	1	2	173	19	14	8	1	0	1
38	1	0.142	1	2	175	20	13	5	2	1	0
39	2	0.140	1	0	176	22	12	4	4	0	0
40	1	0.145	1	1	177	17	12	7	4	0	0
41	2	0.148	1	1	172	24	12	5	4	0	0

[a] Structures 1–8 were calculated directly; structures 9–41 were calculated by using the two incorrect NOEs and then reminimized without them. The violations are measured as differences from (constraint + upper bound) and (constraint − lower bound). The number of distances in each 0.2-Å interval is listed under the column giving the maximum deviation for that interval.

Table II: Correlation of Large NOE Violations with Family

NOE constraint[a]	no. of structures with RMSD >0.8 Å		
	family 1	family 2	family 3
Tyr23 $C^{\delta}H$ to Ile6 $C^{\gamma}Me$	16	0	0
Ala10 NH to Ala50 $C^{\alpha}H$	11	5	0
Lys19 NH to Glu9 $C^{\alpha}H$	0	7	3
total no. of structures in family	31	7	3

[a] These are the only NOE violations of >0.8 Å.

Table III: Comparison of Families of C5a Structures, Obtained by Using a 2.0-Å Cutoff

family	no. of members in family	lowest energy structure in family[a]	RMS displacement ± SD within family[b] (Å)	RMS of average structures (Å) vs	
				family 2	family 3
1	31	−328	1.02 ± 0.27	1.5	2.6
2	7	−315	0.99 ± 0.20		1.7
3	3	−294	0.80 ■ 0.13		

[a] Protein potential energy in kcal/mol (not including the constraint energy), for the final minimization. WNOE = 1, ENOE = 4, and peptide bond planarity was not constrained. [b] Calculated by using the 65 C^{α}s, compared to the averaged coordinates.

violation occurs in less than half the structures. It is also possible that the N-terminal helix has sufficient conformational freedom that the NOE contacts are conformationally averaged and that the NOE constraints used are genuinely inconsistent with any single static structure. Support for this suggestion comes from the fact that the occurrence of the largest violations correlates with the family classification (vide infra), implying that the different families may be different solutions to the problem of producing a structure using mutually incompatible NOE data.

Angle constraints were well met. For those residues with dihedral angles ϕ set to −60°, the constraint was met with an RMSD of 9.7°. The standard deviation of the torsion angle was used as a measure of the spread of dihedral angles between structures. For residues with ϕ constrained, the standard deviation from the mean for this angle had a mean value of 5.1°, compared to 21.7° for the unconstrained residues, indicating significantly less scatter for the constrained residues. As might be expected, standard deviations for ϕ were highly correlated with those for ψ, showing that the regions of well-defined backbone conformation include the dihedral angle ψ as well as ϕ.

The structures were grouped into families, as described (Fry et al., 1989), using RMS fits of C^{α} carbons with a 2.0-Å cutoff. There are only three, rather similar, families, as shown in Table III. The three members of family 3 are from consecutive runs

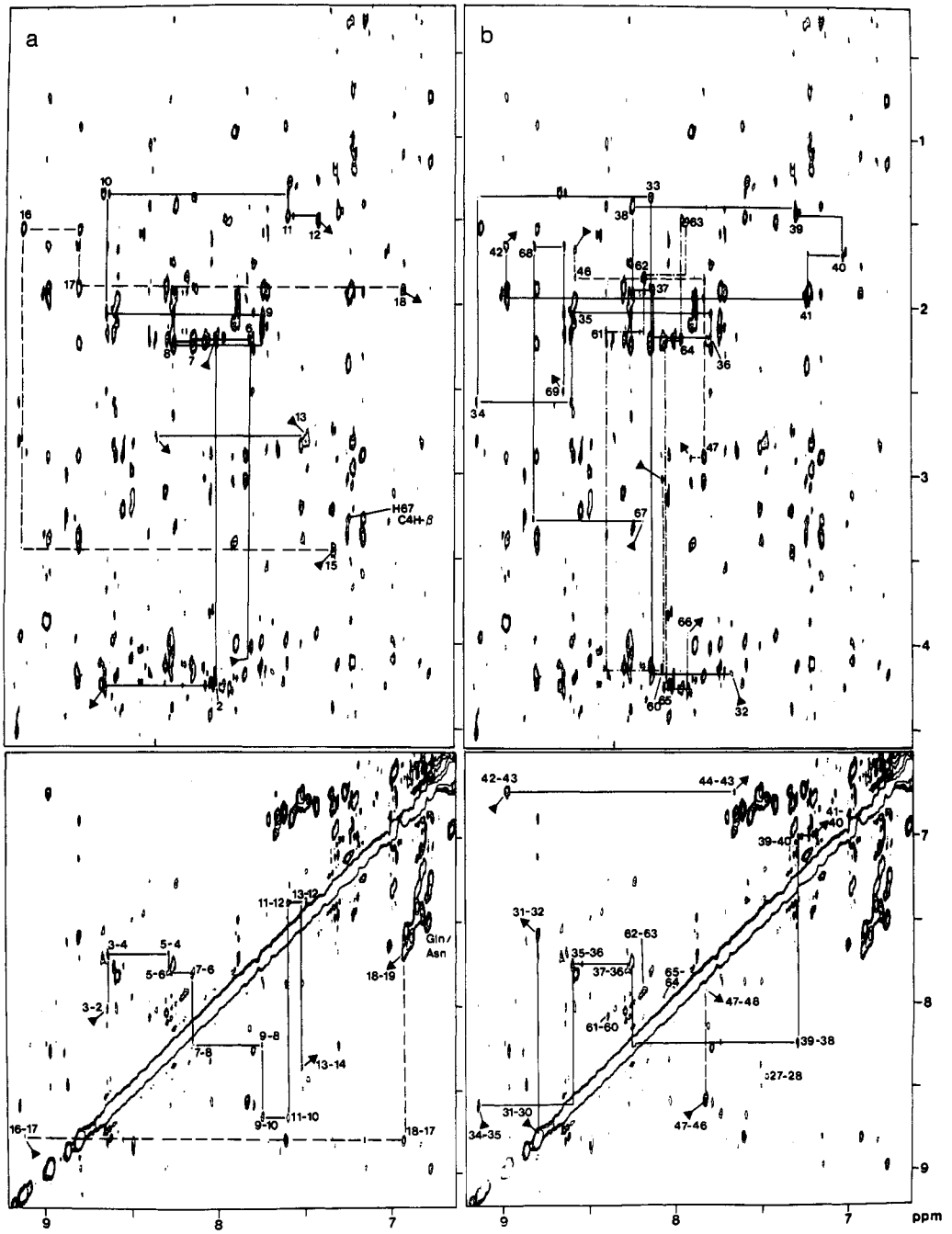

FIGURE 1: Part of a 500-MHz NOESY spectrum of porcine $C5a_{desArg}$ in 90% H_2O/10% D_2O at 35 °C, pH 3.1, 400 scans per increment, spectral width in F_2 4673 Hz, and τ_m 200 ms (randomly varied by 10% each scan), showing the sequential connectivities used to assign the spectrum. The time-domain data were truncated to 1200 × 266 points and transformed into 4096 × 1024 using sine bells shifted $\pi/24$ in t_2 and $\pi/3$ in t_1. The F_2 region shown contains the signals from all aromatic and amide protons. The same spectrum is plotted in both parts, but with different sequential connectivities marked: residues 1–14 (solid line) and 15–19 (dashed line) in (a), and residues 30–44 (solid line), 45–48 (dashed line), 59–66 (dotted and dashed line), and 67–69 (solid line) in (b). The beginning and end of sequential connectivities are marked by solid arrowheads. The connectivities in the top part all involve $C^{\beta}H$, except for residues 2, 5, 32, 60, and 65, which involve $C^{\alpha}H$, and Thr33, which involves C^{γ} methyl. In the top part the numbers refer to intraresidue NOEs, which are coincident with COSY or relayed COSY connectivities. The His67 C4H–$C^{\beta}H$ connectivity is shown in (a). The connectivities for the slowly exchanging amide protons 19–27, 38, 39, 41, and 48–59 are not marked here, as they are included in Williamson (1989).

and so may be an unlikely arrangement which was dependent on starting conditions. Members of families 1 and 2 are scattered throughout the trajectories (see Table I). The families differ in the loop connecting helices A and B and in the way in which helix A lies on the rest of the structure. The backbone conformations of all the structures overlapped in groups by family are shown in Figure 4. The average backbone structures for each family are compared in Figure 5.

The structures overlay very well in general chain fold and orientation, particularly along the helices, but less well in the loops, particularly the loop between helices B and C. This is a consequence of local disorder, as shown by the mean RMS deviation of four-residue chain segments from the average position (Figure 6a). The poor definition of the loops is probably related to the paucity of constraints in these regions, rather than any inherent flexibility, as there are several small values of the three-bond coupling constant $^3J_{HN\alpha}$ in the loops as well as a number of nonrandom sequential NOEs. Values for the standard deviations of the ϕ and ψ dihedral angles residue by residue have a very similar pattern to that shown in Figure 6a, except where ϕ is constrained, in which case its

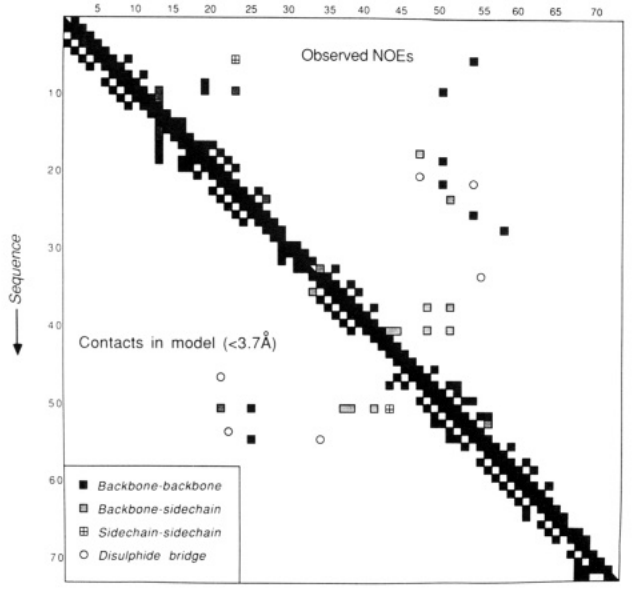

FIGURE 2: Diagonal plot of the NOEs used for the calculation of the structure of C5a, and comparison with the NOEs expected from the original model deduced from the C3a structure. Both axes are calibrated with the sequence of the protein. A filled square at the position (x, y) in the plane indicates an NOE between backbone protons (NH, $C^{\alpha}H$ or $C^{\beta}H$, and $C^{\delta}H$ of proline) of the two residues x and y. A gray square indicates an NOE between the backbone protons of one residue and the side-chain protons of the other, while a square containing a cross indicates an NOE between two side-chain protons. Where two residues are connected by more than one NOE, the NOE involving the greater number of backbone protons is marked. Disulfide bridges are marked by circles. The top half shows the NOEs observed in NOESY spectra using a 200-ms mixing time, and the bottom half shows the NOEs predicted from the original model, using a 3.7-Å cutoff. This cutoff was chosen because most, but not all, sequential NOE connectivities between NH and $C^{\alpha}H$ (maximum distance 3.6 Å) were observed.

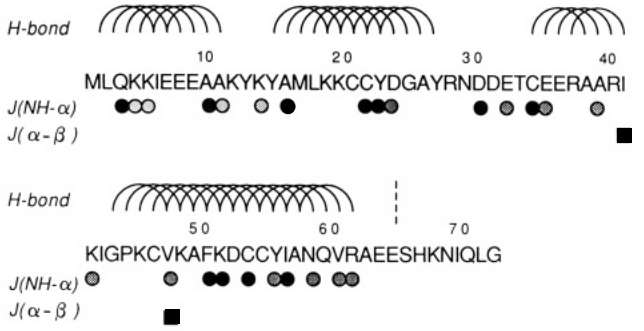

FIGURE 3: Diagram showing the hydrogen bond and angle constraints used for the final calculations. The sequence of C5a is given by the one-letter code. All hydrogen bond constraints link O of residue i with NH of residue $i + 4$. Circles show where the ϕ dihedral angle of a residue was constrained to −60° in consequence of a small value of the three-bond coupling constant between NH and $C^{\alpha}H$. Filled circles denote angles constrained with a force constant of 10 kcal/(mol·rad^2), and gray circles denote those constrained with a force constant of 5 kcal/(mol·rad^2). Squares show residues for which the angle between $C^{\alpha}H$ and $C^{\beta}H$ was constrained to 180° with a force constant of 10 kcal/(mol·rad^2), because of measured large coupling constants between these protons. The dashed line after Glu65 indicates that only the first 65 residues were used in the calculations.

standard deviation is often smaller.

The internal side chains are well-defined, but, as expected, those on the surface are not. This trend is seen strikingly in Figure 6b, where there is a periodic variation in the side-chain RMS deviation along the helices, with a periodicity of about 4 residues, and also in the plots of superimposed structures in Figure 7. For the 31 members of family 1, the average

Table IV: Standard Deviations from Average Position, by Residue Type[a]

residue	positions	standard deviation (Å)
Cys	21, 22, 34, 47, 54, 55	0.42, 0.06, 0.51, 0.22, 0.07, 0.29
Val	48, 61	0.08, 0.41
Ile	6, 41, 43, 57	0.43, 0.15, 0.42, 0.63
Leu	2, 18	0.69, 0.39
Phe	51	0.20
Tyr	13, 15, 23, 27, 56	1.49, 0.92, 0.12, 1.75, 1.85
Met	1, 17	1.17, 1.02
Thr	33	0.72
Asp	24, 30, 31, 53	0.72, 1.12, 1.21, 0.63
Asn	29, 59	1.27, 0.87
Glu	7, 8, 9, 32, 35, 36, 64	1.25, 1.33, 0.91, 1.46, 1.35, 1.46, 1.19
Gln	3, 60, 65	1.55, 1.17, 1.27
Lys	4, 5, 12, 14, 19, 20, 42, 46, 49, 52	1.45, 1.18, 1.08, 1.55, 1.06, 0.90, 1.44, 0.93, 1.40, 1.29
Arg	28, 37, 40, 62	1.73, 1.68, 1.58, 1.96

[a] RMS deviations from the averaged structure were calculated over the heavy atoms after aligning the backbone atoms for that residue. Values for Gly, Ala, and Pro are not included as they are not meaningful. For Phe and Tyr residues the RMS deviations have been corrected for ring flipping of 180°.

RMS displacement from the averaged structure has been calculated for various sets of backbone and side-chain atoms. As shown in Table III, the RMS displacement within the family is 1.02 ± 0.27 Å, when calculated by using only the 65 C^{α}s. When all 259 backbone atoms (N, C^{α}, C′, O) are used, the RMS displacement is slightly reduced, to 0.99 ± 0.25 Å. On adding the side chains of the internal residues Ile6, Glu9, Tyr13, Leu18, Cys21, Cys22, Tyr23, Cys34, Ile41, Ile43, Cys47, Val48, Phe51, Cys54, Cys55, and Ile57 (a total of 322 atoms), the RMS displacement only increases to 1.06 ± 0.31 Å, showing that these side chains have essentially no more variability than the overall backbone, but addition of all heavy atoms (a total of 525 atoms) increases the RMS displacement to 1.84 ± 0.31 Å.

The RMS deviations in side-chain position shown in Figure 6b form a simple and useful measure of variability but are not easy to compare because of the different sizes of the various residues. Therefore, residues are compared by residue type in Table IV, from which the relatively low variability of, for example, Glu9 and Lys20 are apparent.

Discussion

Structural Features. The solution structure of C5a derived here is an approximately antiparallel bundle of four helices and is very similar to the structures previously described for human and bovine C5a (Zuiderweg et al., 1989; Zarbock et al., 1988). Comparison of Figure 8 with Figures 9 and 11 of Zuiderweg et al. (1989) shows that the internal side chains are in similar positions in human and porcine C5a. The core of the protein is held together not only by a substantial hydrophobic cluster but also by three disulfide bridges. As one would expect, hydrophobic residues are on the inside, and hydrophilic on the outside. There is a highly charged cluster on the surface of the protein, made up of residues Asp30, Asp31, Glu32, Glu35, Glu36, Arg40, and Arg62.

Many of the highly conserved residues have obvious structural roles. Of the two glycine residues, Gly44 forms a turn with Pro45, while Gly25 could not be replaced by any other residue, as the side chain would clash with the ring of Phe51. The aromatic rings of Phe51, Tyr13, and Tyr23 all appear to play a structural role; Phe51 is a key residue in the hydrophobic cluster that forms the core of the protein, and Tyr13 and Tyr23 both help to dock the N-terminal helix onto

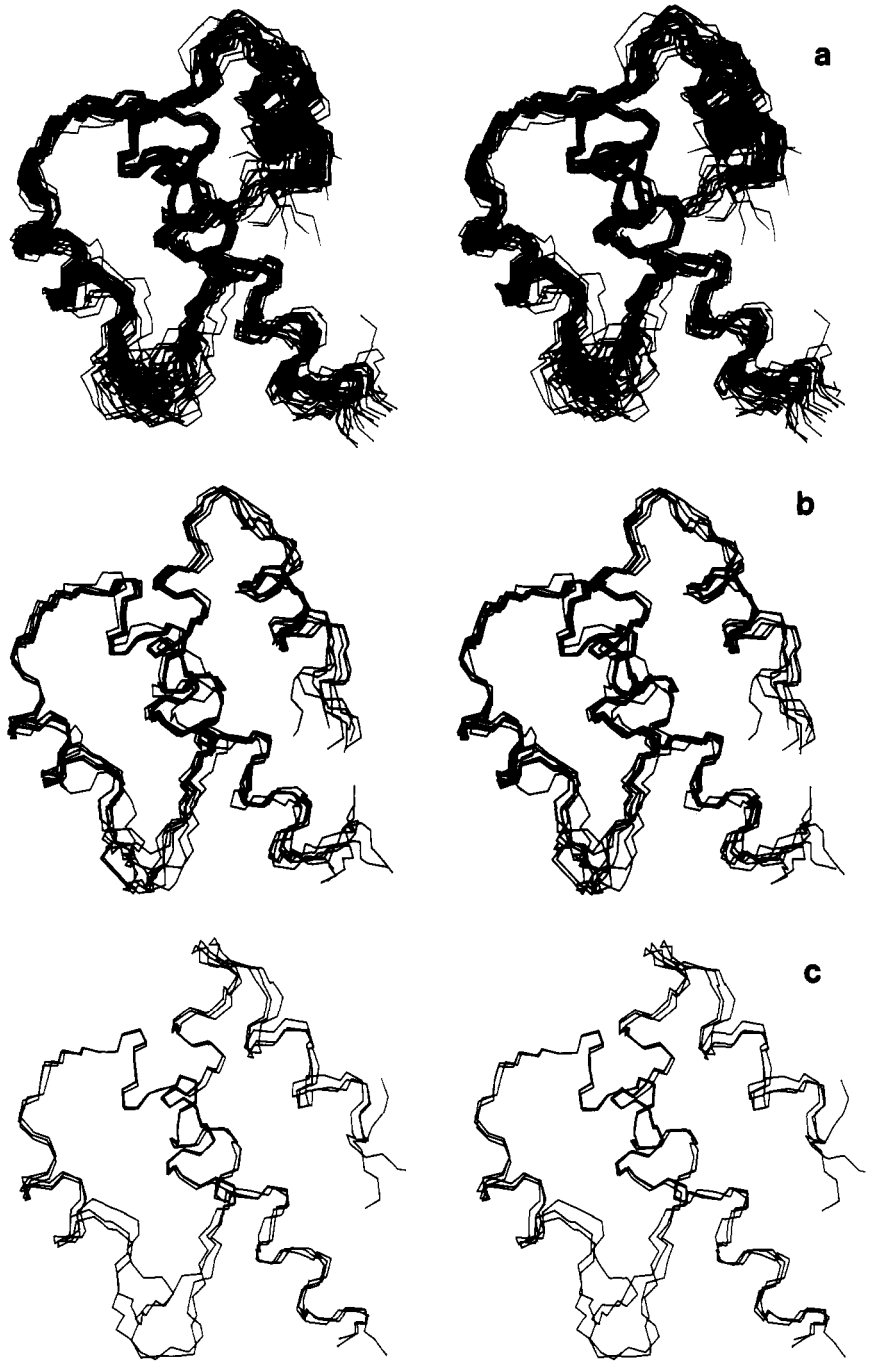

FIGURE 4: Stereoviews of the 41 structures of porcine C5a grouped into three families; backbone atoms (N, C^{α}, C′) only are shown. The structures are superimposed for an RMS best fit against the C^{α} positions in the average structure for each family, by using residues 5–8, 18–25, 37–39, and 46–60. (a) The 31 members of family 1. (b) The seven members of family 2. (c) The three members of family 3.

the body of the protein. Tyr27 (which is a cysteine in human C5a) and Tyr56 are both on the outsides of helices where they would be exposed to solvent and show considerable conformational mobility.

Arg37 is conserved in all C3a, C4a, and C5a sequences known (Greer, 1986). It appears to play a structural role, as the methylene groups in the side chain lie over the top of Phe51, and the guanidinium group hydrogen bonds to backbone carbonyl groups in helix B for two thirds of the members of family 1. In the remaining third of the structures, the guanidinium group of Arg37 is on the surface and often bonds to Glu32. Arg28 also has a structural role, as its side chain is buried between helices B and C in 90% of the members of family 1. The methylene groups of Arg28 make contact with a number of hydrophobic residues, particularly in the loop 28–33. The guanidinium group of Arg28 hydrogen bonds to the backbone carbonyl of Glu32 and to the side chain of Asn59. Interestingly, this residue is valine in human and mouse C5a but is a polar residue (Asp, Asn, Gln) in C3a. This could imply a different conformation for the loop 28–33 in C3a.

Tyr13 is conserved in C3a, C4a, and C5a, and Lys19 has the conservative replacement Arg in C3a. These two residues are very close together in the solution structures derived here, with the methylene groups of the side chain of Lys19 being held in a cavity formed by the rings of Tyr13 and Tyr23. For Lys19, only the terminal amino group (which is exposed to solvent) varies its position in family 1. For the neighboring Lys20, the first two atoms of the side chain have nearly fixed positions while the terminal three atoms have variable positions.

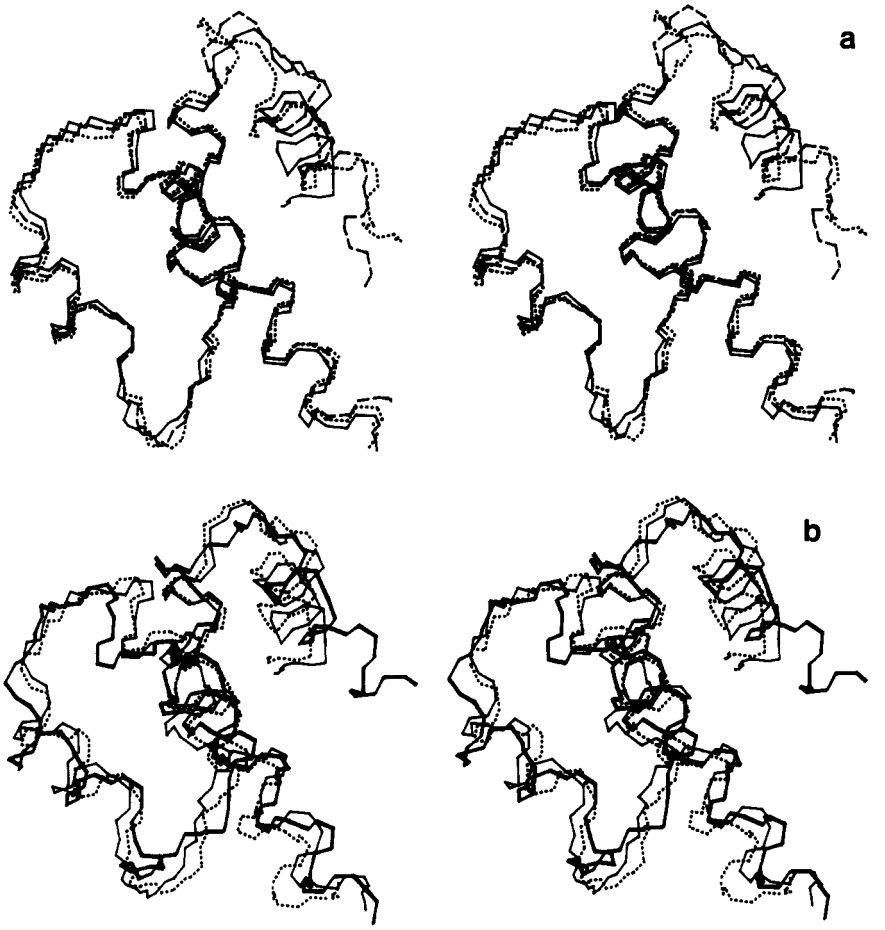

FIGURE 5: (a) Comparison of the three families found with a 2-Å cutoff, superimposed for an RMS best fit against the C^{α} positions in the overall average structure. Family 1—solid line. Family 2—dashed line. Family 3—dotted line. The only significant difference is to be found in the N-terminal helix (at the upper right). (b) Comparison of the starting model with structures generated starting from the model and from a fully extended conformation. Starting model—bold solid line. Average of family 1—light solid line. Average of six runs refolded from fully extended—dotted line. Both panels are shown as stereoviews, with backbone atoms only, and were best fitted by using C^{α} positions from residues 5–8, 18–25, 37–39, and 46–60.

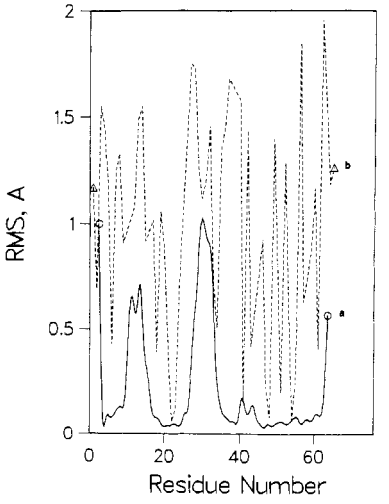

FIGURE 6: (a) (Solid line) Plot of the RMS deviation of C^{α} coordinates from the averaged position, calculated for 4-residue segments from all 41 structures. The values are plotted against the midpoint of the segment, i.e., 2.5 for residues 1–4. (b) (Dashed line) Plot of the RMS deviation from the average position for all heavy atoms in a residue, against residue number (using all 41 structures). For each comparison, the atoms N, C^{α}, and C′ were overlapped, so that the RMS deviation reflects only the variation in side-chain conformation. Gly, Ala, and Pro have been omitted, because their RMS deviations are meaningless.

The conservatively replaced residues Tyr13, Lys19, Lys20, and Ile41 form part of a hydrophobic band (with protruding polar groups) across the surface of the protein, which may have structural significance.

The conformers resulting from the dynamics calculations reported here are closely similar, with the most striking difference being the variable location of the N-terminus with respect to the rest of the protein. It seems likely that this apparent conformational freedom is a real feature of the molecule and not an artifact of the calculations, because amide exchange data, and (for residues 1–3) a reduction in line width and NOE intensity, indicate a greater mobility for the N-terminal helix than for the rest of the protein (except for the eight C-terminal residues, which have no defined conformation in solution). A similar suggestion has been made for the N-terminal helix of C3a in solution (Chazin et al., 1988).

Precision and Accuracy of the Structure. It is a general observation that protein structure determinations using NMR data are more precise (i.e., show a smaller spread between structures) than the individual input constraints (Wüthrich, 1989). This is explained as reflecting the additional steric constraints imposed by the heavy atoms and the high degree of correlation between distances (Driscoll et al., 1989). Improved precision can be obtained by increasing the number of constraints used, and by the use of stereospecific assignments. For example, nine structures of the 74-residue protein Tendamistat were calculated by using 842 NOE-derived distance constraints, 86 angle constraints, 17 hydrogen bonds, and 2 disulfide bridges, with 46% of the prochiral groups stereospecifically assigned, to give an RMS deviation from the

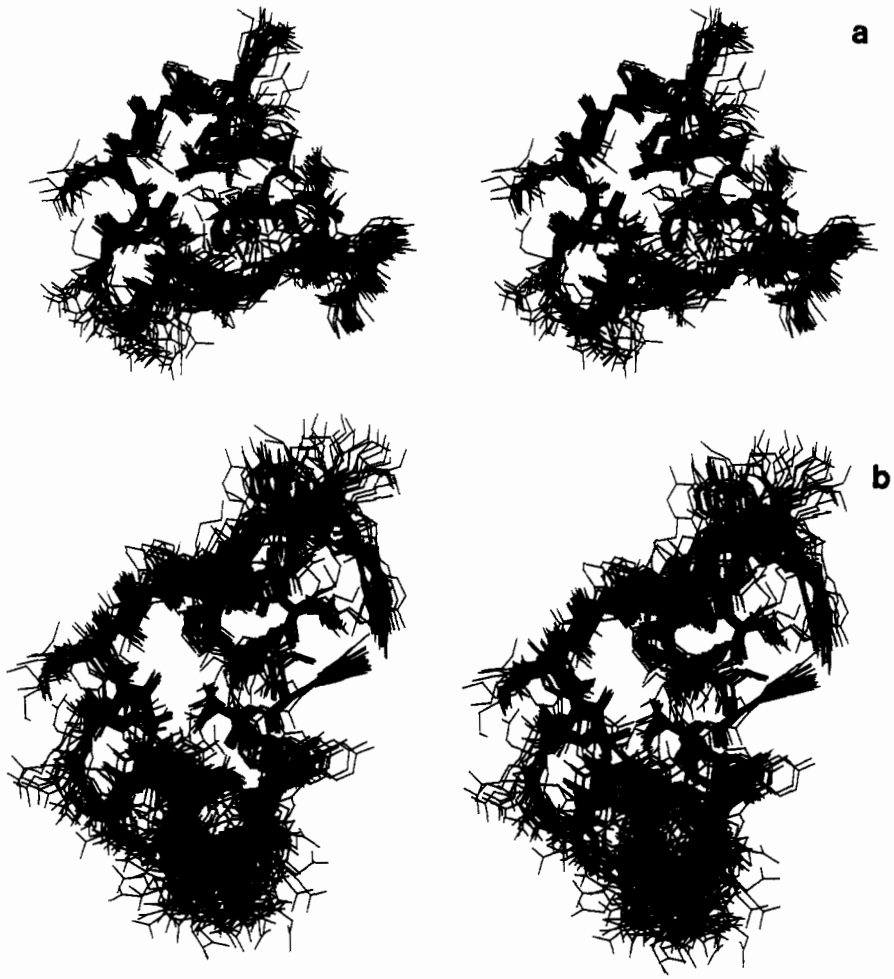

FIGURE 7: Stereoview of the 31 structures of family 1, superimposed for best fit to the indicated C^{α} positions of the average structure for family 1. The orientation is the same as that in Figure 4. All heavy atoms are shown: (a) residues 34–58, using C^{α}s 34–58 for alignment; and (b) residues 13–47, using C^{α}s 18–25 and 37–39 for alignment.

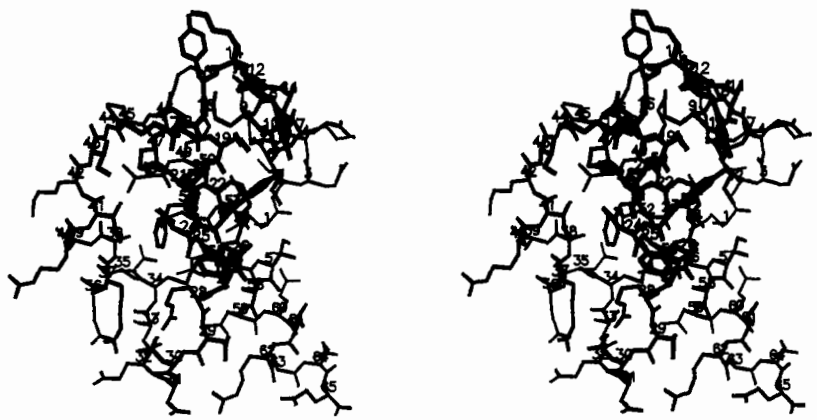

FIGURE 8: Stereoview of the lowest energy member of family 1. Residue numbers are indicated near the C^{α}s. The orientation is the same as that in Figure 4.

average of 0.85 Å for the backbone atoms (Kline et al., 1988). By contrast, only 151 NOE-derived distance constraints and 27 angle constraints were used here, using only two stereospecific assignments, but the RMS deviation was still a very respectable 0.99 Å for backbone atoms (vide supra). From the several factors contributing to this precision, two are worth stressing: first, the extensive periods of equilibration and quenching during the molecular dynamics calculations, and the use of incremental NOE weighting; and second, the use of a large number (34) of hydrogen bond constraints. Hydrogen bonds are particularly powerful constraints, because of their small error and their inherently medium- or long-range nature, and are of great importance in increasing structural precision.[2]

We have shown that the structures obtained here are precise but have not addressed the question of whether they are accurate. A possible source of bias is the fact that the calcu-

[2] In response to a reviewer's comment, we should point out that our weighted-averaging treatment of diastereotopic protons does not artificially increase the precision by locking the conformation around diastereotopic protons. None of the diastereotopic pairs used in NOE constraints was locked into the conformation of the original model. This applies also to diastereotopic aromatic pairs, where ring flips of 180° were common in successive runs.

lations started with a model structure, and it could therefore be that the resulting structures may be inaccurate because of their similarity with an inaccurate model. Indeed, helices B–D of the resulting structures are closely similar to those of the model (with an RMSD of 1.2 Å for the C^{α} within these helices), as expected from the comparison of NOEs given in Figure 2. However, the loops and the orientation of the N-terminal helix are substantially different (RMSD of 3.4 Å for residues not in helices B, C, or D, and RMSD of 2.8 Å for all 65 C^{α}s), indicating that the dynamics calculations have succeeded in moving the structures away from the conformation in the model, where appropriate (see Figure 5b). In particular, the interdigitation of helix A into helices B and D in all the resultant structures is substantially different from that in the starting model.

As a further check, six structures were optimized by using the final set of constraints but no disulfide bonds, starting from a fully extended conformation. After the normal equilibration and the first optimization run, the protein had folded into the four-helix bundle, and the cysteines were positioned so that the disulfide bonds could be added for the final minimization. Furthermore, all six structures belonged to family 1 within the maximum deviation of 2 Å RMS, and the peptide energies and violations of NOE constraints were comparable to those obtained for the original members of family 1. The average of the six structures is compared both to the starting model and to family 1 in Figure 5b, from which the good agreement of the two sets of calculated structures and their substantial differences from the starting model are apparent.

The question of accuracy is best answered by considering whether the structures satisfy all the available information, both that used to determine the structures and that which has not been used (Williamson et al., 1985; Arseniev et al., 1988; Kline et al., 1988). As discussed above, deviations from the dihedral angle ϕ constraints are small, and the list of NOE violations given in Table I shows that the structures agree with the NOE input reasonably well. One should also consider whether the reverse is true, that is, whether the NOESY spectra agree with what would be expected from the structures.

Several low-energy structures were used to calculate expected short interproton distances, and NOESY spectra were examined to see if all the short distances corresponded to locations of cross-peak intensity. Since there were a number of sequential NOEs between $C^{\alpha}H$ and NH that were not observed, and the maximum distance for this contact is 3.6 Å (Wüthrich, 1986), the cutoff distance for potential NOEs was set at 3.7 Å. There were only six short distances that were found in most or all of the structures and that did not correspond to NOESY cross peaks. Three of these involved interactions between aromatic protons and side-chain protons, and one was between Lys46 $C^{\beta}H$ and one of the Leu18 methyls, all of which would be strongly affected by small changes in side-chain orientations. The other two, both predicted to be strong, were Lys12 $C^{\beta}H$–Tyr15 HN and Gly44 HN–Cys47 $C^{\beta}H$. These may indicate some problem with the local conformation in these turn regions, but overall the fit is excellent. Some authors [e.g., Marion et al. (1987) and Arseniev et al. (1988)] have reconstructed NOESY spectra from the structures to demonstrate how well the structures fit the NOESY data. This approach was not considered useful here, because of the lack of assignments for many of the side-chain protons in C5a (Williamson, 1989).

Chemical shifts are another set of parameters that were not used in the structure determination and that can be used to verify the structures (Kline et al., 1988). The precision with which chemical shifts can be calculated is low (Pardi et al., 1983), except for perturbations arising from ring-current shifts, which can be large and reasonably accurately calculated, particularly for side-chain protons (Perkins & Wüthrich, 1979). By use of several low-energy structures, consistent large differences from the random coil shifts (Bundi & Wüthrich, 1979) were predicted for Gly25 $C^{\alpha}H$ (roughly 1 ppm each), Ala38 $C^{\beta}H$ (roughly 0.8 ppm), and Ala38 $C^{\alpha}H$ (roughly 0.5 ppm). The largest chemical shift differences from random coil values found were for Gly25 $C^{\alpha}H$ (1.1 ppm) and Ala38 $C^{\alpha}H$ (1.0 ppm), in good agreement with predictions. All the other large differences from random coil values found for side-chain protons could be accounted for qualitatively by the proximity of aromatic rings in the structures.

The location of slowly exchanging amide protons was used as a criterion for hydrogen bond constraints, and so it is hardly surprising that all the slowly exchanging amides were hydrogen bonded in the structures. A number of other hydrogen bonds were also formed, of which the bonds (O to N) 59–63 and 37–40 occurred in all structures examined. The first of these is an additional α-helical hydrogen bond at the C-terminus of helix D, while the second is a 3_{10}-type turn at the C-terminus of helix C. It is noteworthy that hydrogen bonds 36–40 and 37–41 were also present, so the hydrogen bond 37–40 made up a bifurcated hydrogen bond, which is quite commonly found in crystal structures at helix termini (Richardson, 1981).

Biological Implications. The last few years have seen extensive efforts aimed at defining the residues in C5a responsible for its receptor binding, most notably a series of site-directed mutagenesis experiments from the group at Abbott (Mollison et al., 1989). There is good evidence for involvement of the C-terminus in receptor binding, but other residues are also involved. These have been variously suggested to be located in the "core", consisting of residues 20–37, 47–49, and 50–62 (Johnson et al., 1987), to include residues 23 and 27 (Johnson & Chenoweth, 1985a,b), residues 24–26 (Hahn, 1987), or residues 3–5, 7–13, and 19–25 (Greer, 1985). Site-directed mutagenesis and NMR experiments (Mollison et al., 1989) imply that residues 1–5, 7–9, 11, 12, 23, 24, 32, 33, and 35 are *not* important in binding to the receptor, although they may be involved in maintaining structural integrity, that residues 21 and 26 are important in maintaining correct three-dimensional structure, and that residue 40 (or a residue close to it) and residues in the segment 13–20 may be involved in receptor binding.

The structures described here confirm the features described for human C5a (Zuiderweg et al., 1989), particularly the large spatial separation between residues 13–20, 40, and the C-terminus. Residue 40 is separated from residues 13–20 by Ile43 and Met17, both of which residues tend to be conservatively replaced in C5a from different species, but different in C3a and C4a. These two residues would therefore be interesting targets for site-directed mutagenesis. As mentioned above, residues Tyr13 and Tyr23 form a hydrophobic cavity that restricts the conformational freedom of Lys19 and appears to keep the side chains of Lys19 and Lys20 close together. These two residues are Lys or Arg in all known C3a, C4a, or C5a structures and are presumably important residues for the function of the proteins. It is not yet clear whether they are necessary for the structural integrity of the proteins or whether they form a common recognition site for the receptors.

ACKNOWLEDGMENTS

We thank Dr. A. Kröhn and A. B. Meade (Roche Welwyn) for constructing the model of porcine C5a and making the coordinates available and Drs. G. Englert and A. Labhardt

(Roche Basel) for generous allocations of time on a Bruker AM-500.

Supplementary Material Available

Two tables giving the complete list of NOE and hydrogen bond distance constraints and the complete list of angle constraints (5 pages). Ordering information is given on any current masthead page.

References

Arseniev, A., Schultze, P., Wörgötter, E., Braun, W., Wagner, G., Vašák, M., Kägi, J. H. R., & Wüthrich, K. (1988) *J. Mol. Biol. 201*, 637.

Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S., & Karplus, M. (1983) *J. Comput. Chem. 4*, 187.

Brünger, A. T., Clore, G. M., Gronenborn, A. M., & Karplus, M. (1987) *Protein Eng. 1*, 399.

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

Chazin, W. J., Hugli, T. E., & Wright, P. E. (1988) *Biochemistry 27*, 9139.

Chino, N., Kubo, S., Nishiuchi, Y., Kumagaye, S.-I., Kumagaye, K. Y., Takai, M., Kimura, T., & Sakakibara, S. (1988) *Biochem. Biophys. Res. Commun. 151*, 1285.

Driscoll, P. C., Gronenborn, A. M., Beress, L., & Clore, G. M. (1989) *Biochemistry 28*, 2188.

Fernandez, H. N., & Hugli, T. E. (1978) *J. Biol. Chem. 253*, 6955.

Fry, D. C., Madison, V. S., Bolin, D. R., Greeley, D. N., Toome, V., & Wegrzynski, B. B. (1989) *Biochemistry 28*, 2399.

Gerard, C., & Hugli, T. E. (1980) *J. Biol. Chem. 255*, 4710.

Goldstein, I. M., (1988) in *Inflammation: Basic Principles and Clinical Correlates* (Gallin, J. I., Goldstein, I. M., & Snyderman, R., Eds.) Raven Press, New York.

Greer, J. (1985) *Science (Washington, D.C.) 228*, 1055.

Greer, J. (1986) *Enzyme 36*, 150.

Hahn, G. S. (1987) U.S. Patent no. 4692511.

Hoch, J. C. (1983) Ph.D. Thesis, Harvard University.

Huber, R., Scholze, H., Paques, E. P., & Deisenhofer, J. (1980) *Hoppe-Seyler's Z. Physiol. Chem. 361*, 1389.

Hugli, T. E. (1981) *CRC Crit. Rev. Immunol. 1*, 321.

Johnson, C. E., & Bovey, F. A. (1958) *J. Chem. Phys. 29*, 1012.

Johnson, R. J., & Chenoweth, D. E. (1985a) *J. Biol. Chem. 260*, 7161.

Johnson, R. J., & Chenoweth, D. E. (1985b) *J. Biol. Chem. 260*, 10339.

Johnson, R. J., Tamerius, J. D., & Chenoweth, D. E. (1987) *J. Immunol. 138*, 3856.

Jose, P. J. (1987) *Br. Med. Bull. 43*, 336.

Kline, A. D., Braun, W., & Wüthrich, K. (1988) *J. Mol. Biol. 204*, 675.

Marion, D., & Wüthrich, K. (1983) *Biochem. Biophys. Res. Commun. 113*, 967.

Marion, D., Genest, M., & Ptak, M. (1987) *Biophys. Chem. 28*, 235.

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. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 292.

Müller, K., Amman, H. J., Doran, D. M., Gerber, P., & Schrepfer, G. (1986) in *Innovative Approaches in Drug Research* (Harms, A. F., Ed.) Elsevier, Amsterdam.

Neuhaus, D., & Williamson, M. P. (1989) *The Nuclear Overhauser Effect in Structural and Conformational Analysis*, VCH Publishers, New York.

Pardi, A., Wagner, G., & Wüthrich, K. (1983) *Eur. J. Biochem. 137*, 445.

Pardi, A., Hare, D. R., Selsted, M. E., Morrison, R. D., Bassolino, D. A., & Bach, A. C., II (1988) *J. Mol. Biol. 201*, 625.

Perkins, S. J., & Wüthrich, K. (1979) *Biochim. Biophys. Acta 576*, 409.

Richardson, J. S. (1981) *Adv. Protein Chem. 34*, 167.

Williamson, M. P. (1987) *Magn. Reson. Chem. 25*, 356.

Williamson, M. P. (1989) *J. Mol. Biol. 206*, 407.

Williamson, M. P., Havel, T. F., & Wüthrich, K. (1985) *J. Mol. Biol. 182*, 295.

Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, John Wiley & Sons, New York.

Wüthrich, K. (1989) *Science (Washington, D.C.) 243*, 45.

Zarbock, J., Gennaro, R., Romeo, D., Clore, G. M., & Gronenborn, A. M. (1988) *FEBS Lett. 238*, 289.

Zimmerman, B., & Vogt, W. (1984) *Hoppe-Seyler's Z. Physiol. Chem. 365*, 151.

Zuiderweg, E. R. P., Boelens, R., & Kaptein, R. (1985) *Biopolymers 24*, 601.

Zuiderweg, E. R. P., Nettesheim, D. G., Mollison, K. W., & Carter, G. W. (1989) *Biochemistry 28*, 172.