37. X-x. Xu, J. Zhu, D.-z. Huang, W.-s. Zhou, 1984 International Congress of Pacific Basin Societies, PAC CHEM '84, Honolulu, Hawaii, 16 to 21 December 1984, Abstr. 10E102.
38. W.-b. Guan, W.-j. Huang, Y.-c. Zhou, J.-z. Gong, *Acta Pharmacol. Sinica* **3**, 139 (1982).
39. Z. Ye *et al.*, *J. Trad. Chin. Med.* **3**, 95 (1983).
40. W. K. Milhous, D. L. Klayman, C. Lambros, paper presented at the XI International Congress for Tropical Medicine and Malaria, Calgary, Alberta, Canada, 20 September 1984.
41. H.-m. Gu, B.-f. Lü, Z.-x. Qu, *Acta Pharmacol. Sinica* **1**, 48 (1980).
42. Z. L. Li, H. M. Gu, D. C. Warhurst, W. Peters, *Trans. R. Soc. Trop. Med. Hyg.* **77**, 522 (1983).
43. China Cooperative Research Group on Qinghaosu and its Derivatives as Antimalarials, *J. Trad. Chin. Med.* **2**, 31 (1982).
44. D. Wang and X. Liu, *Zhongguo Yaoli Xuebao* **4**, 191 (1983); *Chem. Abstr.* **99**, 187154n (1983).
45. China Cooperative Research Group on Qinghaosu and its Derivatives as Antimalarials, *J. Trad. Chin. Med.* **2**, 45 (1982).
46. J.-b. Jiang, G.-q. Li, X.-b. Guo, Y. C. Kong, K. Arnold, *Lancet* **1982-II**, 285 (1982).
47. Perhaps in (*6*) this is a typographic error since there were 2099 patients originally treated.
48. China Cooperative Research Group on Qinghaosu and its Derivatives as Antimalarials, *J. Trad. Chin. Med.* **2**, 25 (1982).
49. D. Zhu, B. Huang, Z. Chen, M. Yin, *Acta Pharm. Sinica* **15**, 509 (1980); *Chem. Abstr.* **94**, 167365a (1981).
50. D. Zhu *et al.*, *Zhongguo Yaoli Xuebao* **4**, 194 (1983); *Chem. Abstr.* **99**, 187025w (1983).
51. X.-d. Luo, H. J. C. Yeh, A. Brossi, *Heterocycles*, in press.
52. H. M. Gu, D. C. Warhurst, W. Peters, *Trans. R. Soc. Trop. Med. Hyg.* **78**, 265 (1984).
53. J. Whaun *et al.*, in *Conference Proceedings of Polyamines: Basic and Clinical Aspects* (VNU Science Press, Utrecht, Netherlands, in press).
54. R. Docampo and S. N. J. Moreno, in *Free Radicals in Biology*, W. A. Pryor, Ed. (Academic Press, New York, 1984), vol. 6, pp. 243–288.
55. China Cooperative Research Group on Qinghaosu and its Derivatives as Antimalarials, *J. Trad. Chin. Med.* **2**, 9 (1982).
56. Y. Li, P. Yu, Y. Chen, L. Li, Y. Gai, D. Wang, Y. Zheng, *Acta Pharmaceut. Sinicia* **16**, 430 (1981).
57. M. Cao, S. Hu, M. Li, S. Zhang, *Nanjing Yaoxueyuan Xuebao* **1**, 53 (1982); *Chem. Abstr.* **100**, 34720h (1984).
58. G. Q. Li, *Chung Hua I Hsueh Tsa Chih* **62**, 293 (1982).
59. H.-m. Gu *et al.*, *Chung-kuo Yao Li Hsueh Pao* **2**, 138 (1981); *Chem. Abstr.* **95**, 161913b (1981).
60. J. Jiang *et al.*, *Zhongguo Yaoli Xuebao* **4**, 197 (1983); *Chem. Abstr.* **99**, 187026x (1983).
61. Y. Li *et al.*, *K'o Hsueh T'ung Pao* **24**, 667 (1979); *Chem. Abstr.* **91**, 21136u (1979).
62. Y. Li, *Huaxue Shiji* **2**, 88 (1982); *Chem. Abstr.* **97**, 91182c (1982).
63. Y. Li, P. Yu, Y. Chen, R. Ji, *Huaxue Xuebao* **40**, 557 (1982); *Chem. Abstr.* **98**, 4420h (1983).
64. Y. Li, P.-l. Yu, I.-h. Chen, J.-y. Chi, *Yao Hsueh T'ung Pao* **15**, 38 (1980); *Chem. Abstr.* **96**, 6883u (1982).
65. Q. Yang, W. Shi, R. Li, J. Gan, *J. Trad. Chin. Med.* **2**, 99 (1982).
66. R. Li, T. Y. Liao, K. Y. Huang, L. L. Chou, *Chung Ts'ao Yao* **12**, 20 (1981); *Chem. Abstr.* **96**, 45792h (1982).
67. L. W. Scheibel, E. Bueding, W. R. Fish, J. T. Hawkins, in *Malaria and the Red Cell*, J. W. Eaton and G. J. Brewer, Eds. (Liss, New York, 1984), pp. 131–142.
68. J.-a. Wu and R.-y. Ji, *Acta Pharmacol. Sinica* **3**, 55 (1982).
69. H. M. Gu, D. C. Warhurst, W. Peters, *Biochem. Pharmacol.* **32**, 2463 (1983).
70. S. Ding *et al.*, *Bull. Chin. Mater. Med.* **6**, 25 (1981).
71. W. H. Li, H. L. Shu, G. Y. Xu, Y. L. Zeng, *Yao Hseuh Hsueh Pao* **17**, 783 (1982).
72. Z. Wang, Y. Zhu, S. Zhang, X. Lu, *Yaoxue Xuebao* **16**, 466 (1981); *Chem. Abstr.* **97**, 61086g (1982).
73. R. Zang, J. P. Chang, S. Xu, Y. Li, *Yaoxue Xuebao* **16**, 460 (1981); *Chem. Abstr.* **97**, 115368w (1982).
74. R. Zhang, S. Xu, Y. Li, *Acta Pharmaceut. Sinica* **16**, 460 (1981).
75. S. Liu and Q. Gao, *Zhongcaoyao* **15**, 109 (1984); *Chem. Abstr.* **100**, 1978509q (1984).
76. X. Shen, K. Yan, Z. Luo, Y. Tian, M. Zeng, *Yaowu Fenxi Zazhi* **3**, 24 (1983); *Chem. Abstr.* **99**, 10924n (1983).
77. Y. Li, G. Song, Y. Gao, *Fenxi Huaxue* **11**, 545 (1983); *Chem. Abstr.* **99**, 1284204y (1983).
78. B. Wang and F. Yin, *Zhongcaoyao* **15**, 97 (1984); *Chem. Abstr.* **101**, 55364f (1984).
79. J. F. Mi and C. Y. Shen, *Yao Hseuh Hsueh Pao* **17**, 421 (1982).
80. X. Wang and Y. Zhang, *Yaowu Fenxi Zazhi* **3**, 353 (1983); *Chem. Abstr.* **100**, 74036w (1984).
81. Y. Ding and C. Wang, *Fenxi Huaxue* **11**, 158 (1983); *Chem. Abstr.* **99**, 115281d (1983).
82. Y. D. Zhang, C. G. Wang, C. Y. Xu, Y. J. Zhuang, Y. L. Zeng, *Yao Hsueh Hsueh Pao* **17**, 212 (1982).
83. H. Shu, G. Xu, W. Li, Y. Zeng, *Fenxi Huaxue* **10**, 678 (1982); *Chem. Abstr.* **99**, 43606p (1983).
84. X.-d. Luo, H. J. C. Yeh, A. Brossi, J. L. Flippen-Anderson, R. Gilardi, *Helv. Chim. Acta* **67**, 1515 (1984).
85. G. Li *et al.*, *J. Trad. Chin. Med.* **2**, 125 (1982).
86. W. Le *et al.*, *Yaoxue Xuebao* **17**, 187 (1982); *Chem. Abstr.* **96**, 210480q (1982).
87. W. Yue, J. You, J. Mei, *Zhongguo Yaoli Xuebao* **5**, 60 (1984); *Chem. Abstr.* **100**, 167771x (1984).
88. L. Wu, H. Yang, Y. Yang, *Yaoxue Xuebao* **18**, 7 (1983); *Chem. Abstr.* **99**, 16157m (1983).
89. W. Le, J. You, J. Mei, *Yaoxue Xuebao* **18**, 619 (1983); *Chem. Abstr.* **100**, 44939a (1984); *Yaoxue Xuebao* **17**, 187 (1982).
90. R. Chen, Z. Qu, M. Zeng, J. Li, *Yaoxu Tongbao* **18**, 410 (1983); *Chem. Abstr.* **100**, 17251p (1984).
91. K. K. Zhuang, *J. New Medicine* (No. 6), 39 (1979).
92. R. Qian, Z. Li, J. Yu, D. Ma, *J. Trad. Chin. Med.* **2**, 271 (1982).
93. G. Huang and Y. Zhao, *J. Trad. Chin. Med.* **3**, 171 (1983).
94. National Institutes of Health, NCI Report on NSC 369397 (tested on 25 Oct. 1983).

RESEARCH ARTICLE

Model Structure for the Inflammatory Protein C5a

Jonathan Greer

C5a is a biologically active molecule produced during the course of serum complement activation (*1*). As an anaphylatoxin, it shares with C3a and C4a the ability to induce contraction of smooth muscle, increase vascular permeability and release histamine from mast cells and basophilic leukocytes. However, C5a also has the ability to recruit monocytic and polymorphonuclear leukocytes to inflammatory sites by stimulating their locomotion (chemokinesis) and directing their migration (chemotaxis). C5a also fosters the release from these inflammatory cells of free radicals and tissue-digesting enzymes that are believed to play a major role in the tissue injury accompanying inflammation. Consequently, inhibition of C5a activity is of considerable therapeutic interest for the treatment of inflammatory diseases such as rheumatoid arthritis.

Human C5a is a relatively small protein with a sequence of 74 amino acid residues (*2*). It contains one oligosaccharide (molecular size 2800 to 3000 daltons), which is attached to an asparagine at position 66[64] [see (*3*)]. The total molecular size of human C5a is around 11,000 daltons. Porcine C5a has also been sequenced (*4*); it consists of 74 residues but has a Glu rather than Asn at position 66[64] and therefore contains no carbohydrate. The sequences of C3a, C4a, and C5a are all clearly homologous, suggesting a common evolutionary origin for these molecules (*1*).

Studies have been performed on C5a in order to identify the parts of the molecule that are essential for receptor binding and biological activity. C5a desArg, from which the COOH-terminal Arg residue has been removed, shows a significant decrease in chemotactic activity (*5*). When the pentapeptide that corresponds to the COOH-terminal five residues of C5a is examined, no chemotactic or chemokinetic activity is observed (*6*). On the other hand, the balance of the molecule containing the remaining 69 NH_2-terminal residues, while chemotactically inactive, does compete with C5a for binding to the C5a receptor (*7*). Thus, it appears that the NH_2-terminal portion and the COOH-terminus of the C5a molecule interact with the receptor and are involved in producing biological activity. Similar modification and peptide studies on C3a (*1*) show that only the COOH-terminal residues of C3a are required or necessary for full biological activity.

A considerable increase in our ability

Jonathan Greer is senior project leader in the Physical Biochemistry Laboratory of the Computer Assisted Molecular Design Group at Abbott Laboratories, Abbott Park, Illinois 60064.

to probe and understand the function of C5a would come from knowledge of its three-dimensional structure. Although the structure of C5a is not known from experimental sources, the structure of the homologous C3a molecule has been determined by x-ray diffraction studies (*8, 9*). Consequently, comparative modeling methods may be used to produce a tentative model structure of the C5a molecule.

Abstract. *The complement cleavage product C5a is a potent stimulant of inflammatory processes; thus, inhibition of C5a activity is of therapeutic interest. The three-dimensional structure of the major portion of C5a was modeled from the homologous C3a crystal structure by comparative modeling techniques. The model shows that core residues of C5a are completely conserved, while external residues differ from C3a. Even though the amino-terminal 12 residues of C3a are disordered in the crystal, this sequence in C5a may form an amphipathic helix. The distribution of species sequence differences in the complete C5a structure suggests a possible receptor binding site.*

Comparative modeling methods have been applied to a large number of homologous proteins including α-lactalbumin (*10*), bacterial (*11*), and mammalian (*12–15*) serine proteases and aspartyl proteases (*16*). In reporting such studies, it is important to give precise details of how the modeling was performed and some sense of the degree of reliability or confidence in the various parts of the model structure. Such assessments are based on the general and local degree of sequence homology and on its relation to the extent of structural conservation.

The properties of the reported crystal structure of C3a suggest that the modeling of C5a be considered in three segments. The central portion of the molecule, residues 13 through 73, is well defined in the C3a crystal structure and provides a reasonable basis for comparative molecular modeling. The COOH-terminus of the C3a molecule has a defined conformation in the crystal; however, this appears to be due to an intermolecular contact that is peculiar to the crystal. Hence, the conformation of this portion of the C3a molecule is different and perhaps disordered in solution (*9*). Finally, no electron density appears for the NH_2-terminal 12 residues of C3a in the crystal and hence their structure is completely unknown.

Modeling the central portion of the C5a molecule. The first step in modeling this part of C5a requires the accurate alignment of the C5a sequence to that of C3a (Fig. 1). While the two molecules are clearly homologous, the degree of homology is sufficiently low that it is difficult in some cases to distinguish true sequence homology from chance correspondence of the sequences. To aid in positioning additions and deletions in the sequence, the three-dimensional structure of the C3a molecule was considered. In particular, the sequences were aligned so that no additions or deletions would occur in the middle of α-helices (*12, 13*) (Fig. 1). Thus, for example, the alignment shown for residues 63[61] to 77[74] in Fig. 1 was selected over that of Fernandez and Hugli (*2*) shown below because residues 47[45] to 73[71] form one

```
                  65        70        75
C3a  ... L R R Q H A R A S H - - L G L A R

C5a  ... L R - - - A N I S H K D M Q L G R
                     65        70
```

long α-helix in C3a (*9*) and the deletion of three residues would cause a gap in the helix, while the concomitant addition of two residues would give a bulge out of the helix, both of which are very unlikely. The alignment for this region shown in Fig. 1 is further supported by the fact that other sequences of C3a from pig, rat, and mouse (*17–19*) fail to preserve the A X X S H sequence at residues 68[66] to 72[70].

From the sequence alignment of Fig. 1, the structure for the central portion of C5a was generated by "mutation" of the C3a sequence. Only one residue addition was necessary, around position 15[14], and one deletion, around position 34[32] (Fig. 2). The model structure begins at residue 13[11] and forms a turn through residue 16[15]. The single residue addition occurs in this loop. Residues 17[16] through 27[26] form an α-helix. Residues 28[27] to 34[32] are another turn in which the residue deletion occurs. Positions 35[33] through 41[39] and 47[45] through 73[71] form two α-helices with a turn lying between them composed of residues 42[40] through 46[44]. Thus, as has been found previously for the serine proteases (*12–14*), all the additions and deletions occur in external loops that lie between the secondary structure elements that are conserved and that contain the significant sequence homology.

Residue Asn 66[64], which lies near the COOH-terminus of the long α-helix, is the site of the carbohydrate attachment to C5a (see arrow in Fig. 2). The model shows this site to be well exposed with all the space necessary for binding a large oligosaccharide. It lies far enough from the COOH-terminus not to interfere directly with receptor binding of Arg 76[74], yet close enough to exert some influence on biological activity (*20*).

When examined in detail, the model structure for C5a shows a remarkable property. The residues that change in sequence when C3a is "mutated" to form C5a are shown in yellow in Fig. 2, whereas the residues that remain conserved in the two proteins are shown in red. Figure 2 shows that all side chains that form the core of these two anaphylatoxin molecules, especially those around Phe 53[51] and Arg 39[37], are complete-

```
C3a#        5         10        15        20        25        30        35        40
            - - - - - - - -
C3aH  S V Q L T E K R M N K V G K Y - P K E L R K C C E D G M R Q N P M R F S C E R R T R
C5aH      T L Q K K I E E I A A K Y K H S V V K K C C Y D G A C V N N D E - T C E Q R A A
C5aP      M L Q K K I E E E A A K Y K Y A M L K K C C Y D G A Y R N D D E - T C E E R A A
          - - - - - - - -
C5a#              5         10        15        20        25        30        35

C3a#   45        50        55        60        65        70        75
C3aH  F I S L G E A C K K V F L D C C N Y I T E L R R Q H A R A S H L G L A R
C5aH  R I S L G P R C I K A F T E C C V V A S Q L R A N I S H K D M Q L G R
C5aP  R I K I G P K C V K A F K D C C Y I A N Q V R A E E S H K N I Q L G R
C5a#  40        45        50        55        60        65        70
```

Fig. 1. Amino acid sequences of C3a and C5a. The numbering shown on the top line is that of the C3a molecule (line labeled "C3a#"); on the bottom line that of C5a (labeled "C5a#"). The human C3a sequence (*9*) is labeled "C3aH," human C5a is "C5aH" (*2*), and porcine C5a is "C5aP" (*4*). The solid lines above and below the sequences mark the positions of the α-helices in the crystal structure of C3a (*9*). The single-letter code for amino acids (or residues) is A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr.

ly conserved. The core of these two molecules is identical in sequence, and no adjustments need to be made to the main chain of the C5a model to accommodate changes in buried side chains. Thus, while there is only 36 percent identity in the sequence between these two proteins, virtually all these identities form the core of the structure. In contrast, most of the external residues on the surface of the molecule are changed from C3a to C5a, and these changes are often dramatic in terms of residue character (see, for example, positions 24[23], 29[28], 32[31], 42[40], 50[48], and 65[63] in Figs. 1 and 2).

Model for the NH_2-terminus of C5a. Since no crystal structure exists for the

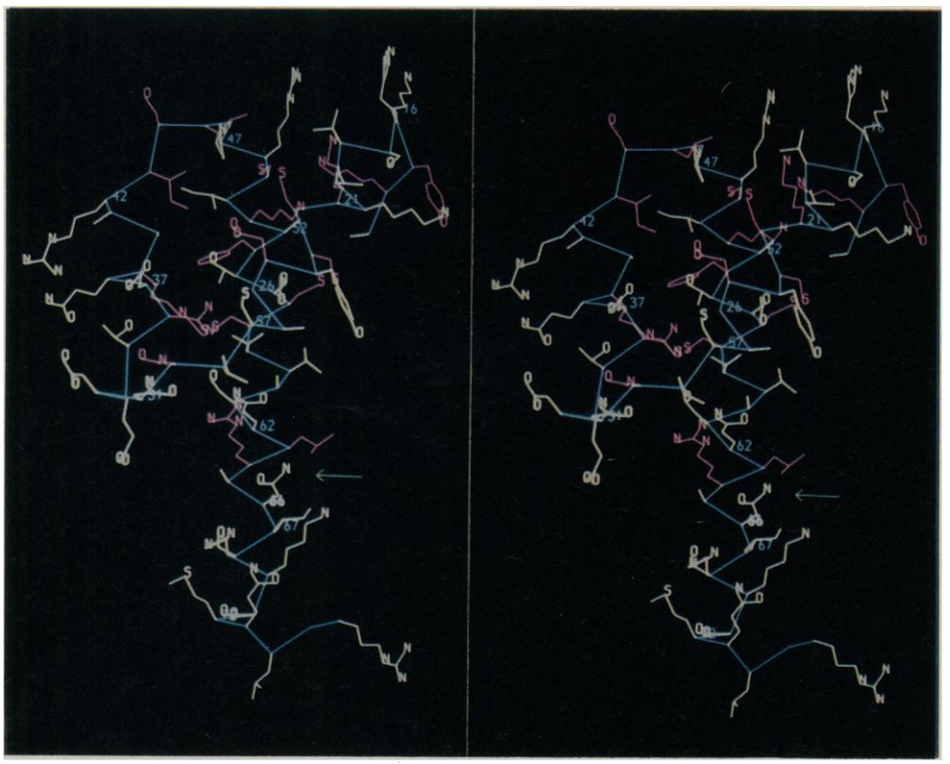

Fig. 2. The model structure of C5a in stereo. (Blue) The α-carbon plot of the molecule; (red) the side chains conserved in human C3a and C5a; (yellow) the side chain residues in C5a that differ from C3a. Residue numbering corresponds to that for C3a. All the internal side chains in the structure are conserved between C3a and C5a. When the sequences are compared (Fig. 1), it is clear that the external residues are highly variable and can be dramatically different in the two proteins. Asn 66[64], labeled in white and marked by the green arrow, is the site of the bound carbohydrate in C5a. It is adjacent to the COOH-terminus of the molecule.

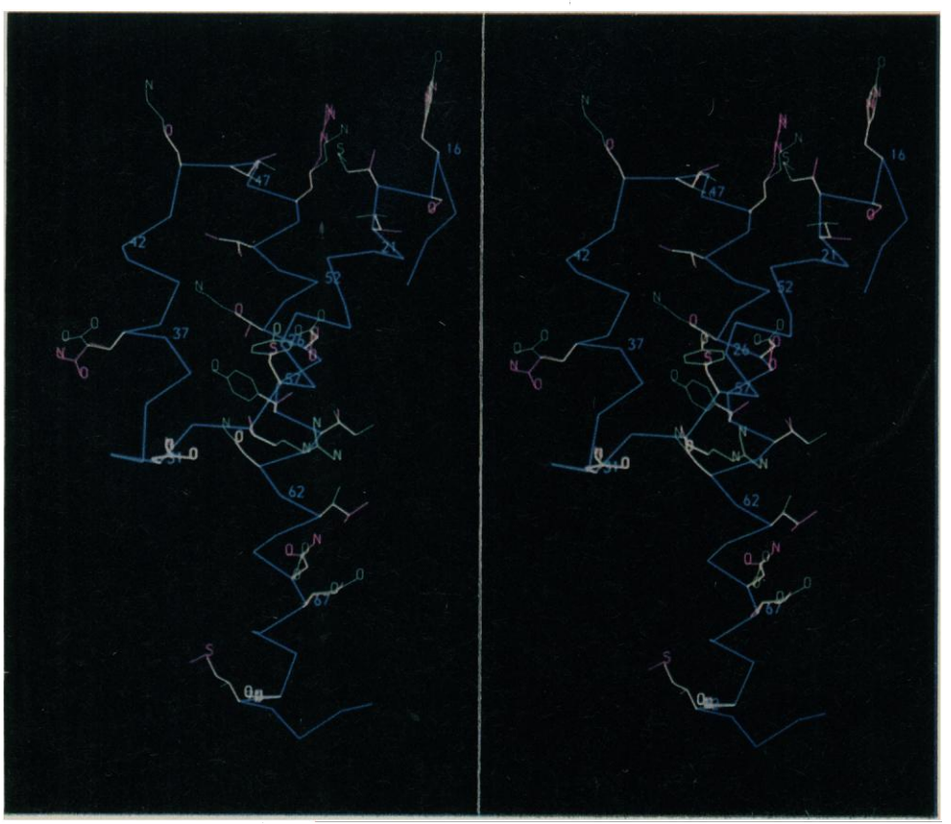

Fig. 3. The sequence differences between human C5a (red) and porcine C5a (green) are shown superimposed on the α-carbon plot (blue) of C5a. Atoms common to both the human and porcine side chains appear in yellow. Side chains with the identical sequence in both species are not shown. The substitutions are distributed throughout the surface of the molecule while the core of the structure continues to be completely conserved.

NH_2-terminal dodecapeptide of C3a, the NH_2-terminus of C5a cannot be modeled by comparative techniques. Nevertheless, because it may be involved in receptor binding, a model for this portion of C5a would be valuable in probing functional properties.

Close examination of the sequence of residues 3[1] to 11[9] in C5a (Fig. 1) reveals an interesting and suggestive property. Residues Leu 4[2], Ile 8[6], and Ile 11[9] are hydrophobic, while residues Thr 3[1], Gln 5[3], Lys 6[4], Lys 7[5], Glu 9[7], and Glu 10[8] are all very hydrophilic. Such an arrangement of residues is highly suggestive of an amphipathic α-helix that lies on the surface of a protein where it is mostly exposed to solvent and thus mainly hydrophilic but has one face that is hydrophobic which would contact the rest of the protein molecule.

As a means of ascertaining whether this region of C5a might be a reasonable candidate for an α-helix, secondary structure predictions were made, by the method of Chou and Fasman (*21*), on the NH_2-terminal sequences of human C3a and C5a (Table 1). While the prediction for the NH_2-terminus of C3a is ambiguous as to whether it is α- or β-like, for the NH_2-terminus of C5a, it is strongly and distinctly helical. As a result of the above considerations, the nonapeptide consisting of residues 3[1] to 11[9] of C5a was folded into an α-helix.

How does this NH_2-terminal helix interact with the central portion of the C5a molecule? As a means of discovering possible binding positions for this helix, the available sequences for human and porcine C5a were compared by placing them in the model of the C5a structure. It was evident (Fig. 3) that all the changes between the two sequences were on the surface of the molecule and that the core residues were completely conserved in these two sequences, as expected. Furthermore, the changed residues were spread throughout the whole surface of the molecule and not concentrated in any particular area. However, one region of the surface had no residue changes and was conserved: around residues 23[22] to 27[26] and 52[50], 53[51], 56[54], 57[55], 60[58], 63[61], and 64[62]. Examination of this conserved region revealed

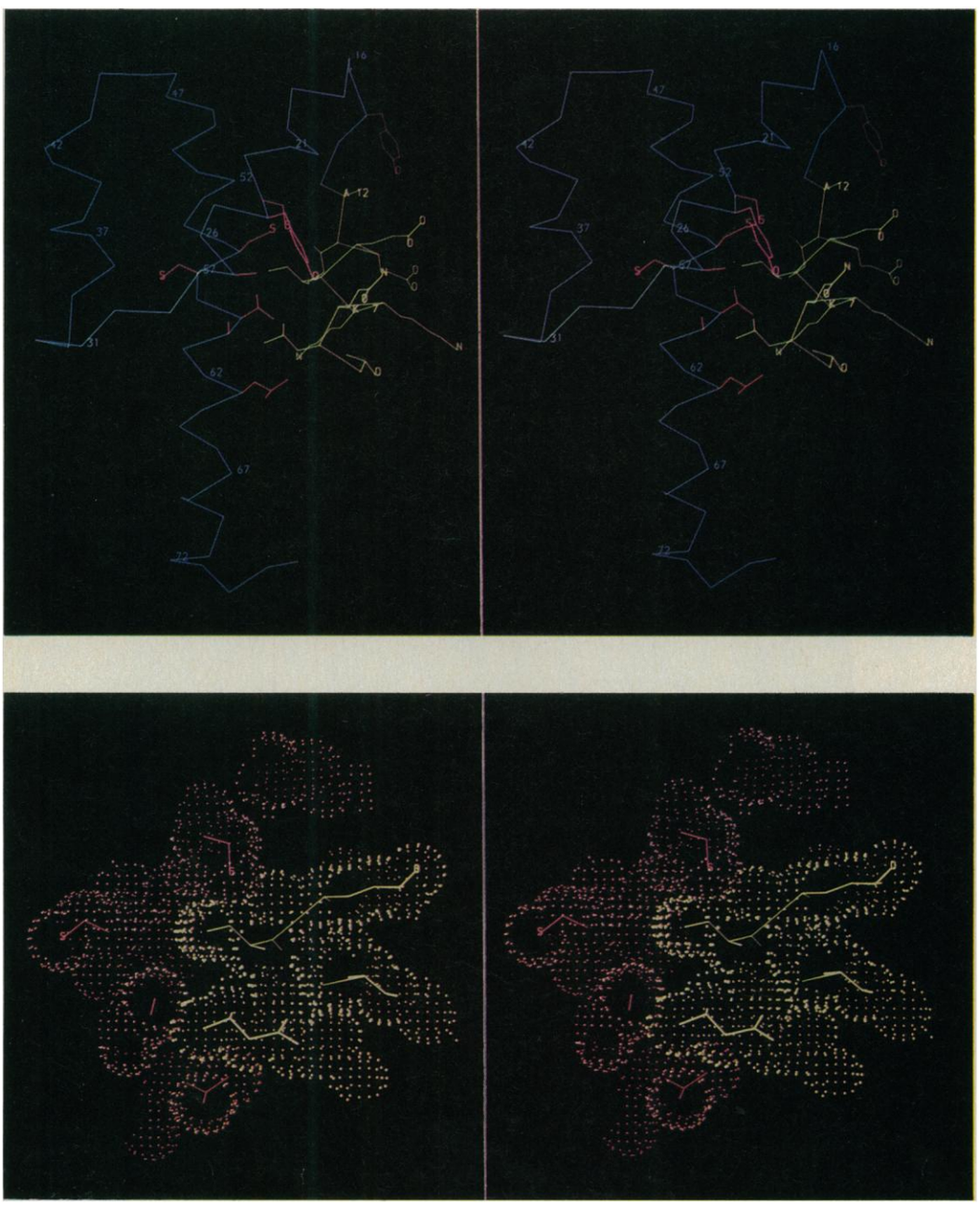

Fig. 4. The docked position of the putative NH_2-terminal α-helix of C5a on the central portion of the molecule. (Top) The resulting structure is shown in skeletal form. The NH_2-terminal helix consisting of α-carbon plot and side chains for residues 3[1] to 12[10] is shown in green. The residues that are in contact with this helix on the central portion of the molecule are shown in red on a blue α-carbon plot. These residues are part of the helix between 17[16] and 27[26] and of the long helix from 47[45] to 73[71]. (Bottom) A surface representation of the docked NH_2-terminal helix (green) onto the remainder of the molecule (red) is shown. Only a slice of the depth of the full structure found in the top panel is included here for clarity. This slice includes the bottom of Leu 4[2] and the top of Ile 8[6]. Both skeletal and surface descriptions were used to manually dock this helix onto the molecule.

a hydrophobic patch that could form the binding site for the hydrophobic face of the putative NH_2-terminal helix.

The nonapeptide helix consisting of residues 3[1] to 11[9] was docked manually onto the above hydrophobic patch. Both skeletal (Fig. 4, top) and surface (Fig. 4, bottom) representations were used in order to maximize interacting surface yet minimize overlap of atoms. The resulting structure (Fig. 4) has the NH_2-terminal helix interacting with two helices on the central portion of the molecule: with a part of the helix from 17[16] to 27[26] and also with a section of the long COOH-terminal helix. The contact is completely hydrophobic and is formed of residues Leu 4[2] and Ile 8[6] on the NH_2-terminal helix interdigitated with residues Cys 23[22], Tyr 24[23], Ala 27[26], Cys 56[54], Val 59[57], Ala 60[58], and Leu 63[61] on the central portion of the molecule. Fitting the helix onto this patch was greatly simplified in that most of the side chains in the contact are either small or relatively fixed in conformation; thus no initial adjustments in the side chain positions were required to obtain a reasonable docking conformation (Table 2).

The residues on both sides of the contact are conserved in porcine C5a except that Val 59[57] changes to Ile, and Ile 11[9] goes to Glu. Both these differences can easily be accommodated since they lie at the edge of the contact (Figs. 3 and 4).

When the NH_2-terminal nonapeptide helix of human C5a was docked onto the rest of the molecule as described above, residues Ile 11[9] and Ala 12[10] were close enough to join Ala 13[11] (whose position had been derived from the C3a crystallographic structure) with only minor adjustments. In this conformation, residues 11[9] through 16[15]—including the added residue in C5a over C3a (Fig. 1)—quite naturally form the turn between the NH_2-terminal helix and the

Table 1. Chou and Fasman predictions of secondary structure for the NH_2-terminus of C3a and C5a.

Protein	Residues		<Pα>*	<Pβ>*	Type predicted
	Begin	End			
C3a	1[−1]	9[7]	1.120	1.014	α
	1[−1]	5[3]	0.996	1.208	β
C5a	4[2]	14[12]	1.256	0.929	α

*α-Helix- and β-sheet-forming propensities. See (*21*) for a definition of these values.

Table 2. Contact distances between the NH_2-terminal helix and the rest of C5a. Indicated contact distances are for the closest distance of any atom-atom interaction (both main chain and side chain) between the indicated two residues. "++++" indicates all distances larger than 5.0 Å. Residues Thr 3[1], Lys 6[4], Glu 9[7], and Glu 10[8] make no contacts shorter than 5.0 Å with the rest of the molecule.

Position in rest of C5a	Positions in NH_2-terminal helix				
	4[2] Leu	5[3] Gln	7[5] Lys	8[6] Ile	11[9] Ile
20[19], Lys	++++	++++	++++	++++	++++
23[22], Cys	++++	++++	++++	2.87	++++
24[23], Tyr	++++	3.45	++++	4.83	++++
26[25], Gly	++++	++++	++++	4.95	++++
27[26], Ala	4.05	++++	++++	2.88	++++
56[54], Cys	++++	++++	++++	3.06	4.55
57[55], Cys	++++	++++	++++	4.88	++++
59[57], Val	4.49	++++	3.48	3.62	4.33
60[58], Ala	3.01	++++	++++	4.71	++++
63[61], Leu	3.32	++++	3.90	++++	++++

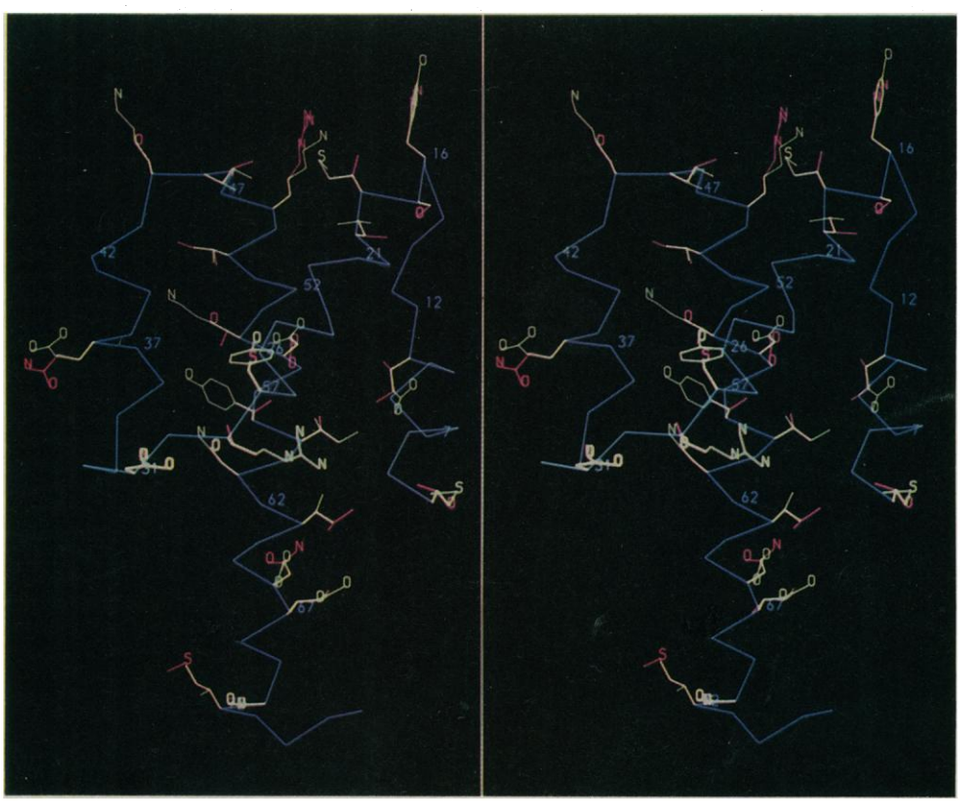

Fig. 5. The sequence changes are compared for human C5a (red) and porcine C5a (green) on the full proposed model for C5a (blue) including the putative NH_2-terminal α-helix. Atoms common to both human and porcine side chains appear in yellow. The contact between the NH_2-terminal helix and the rest of the molecule is completely preserved with the possible exception of Val 59[57] → Ile which can easily be accommodated in this structure. It is interesting that, despite the widespread nature of the changes between these two species, a conserved region does appear in the upper right of the molecule as shown, which includes residues 5[3], 6[4], 7[5], 9[7], 10[8], and 20[19] through 26[25]. This region is highly hydrophilic and may represent a tentative location for the receptor binding site on C5a.

helix from residues 17[16] through 27[26].

The C5a model structure can be further refined by energy minimization techniques with the program VFFPRG of Hagler and co-workers (*22, 23*). This method permits the accommodation of the addition and deletion in the loops at 11[9] to 16[15] and 28[27] to 33[32], respectively (Fig. 1), as well as optimization of the structure and contact interactions of the NH_2-terminal helix.

The proposal of an α-helix at the NH_2-terminus and its docking of C5a leads to the questions whether this helix and interaction occur in C3a and if so why it does not appear in the crystal structure (*9*). The contact residues (Table 2) were examined in the sequence of C3a. It was apparent that, although the NH_2-terminus of C3a might form an α-helix (Table 1), this helix could not form the same hydrophobic contact as in C5a. The Ile 8[6] of C5a becomes an Arg in C3a; which is both charged and too large to fit into the same contact as the Ile. Furthermore, a number of residues on the other side of the contact are changed including Ala 27[26] → Met, Val 59[57] → Tyr, Ala 60[58] → Ile, and Tyr 24[23] → Glu; these changes modify the contour of the molecular surface quite dramatically in this region of the structure. Thus the identical interaction, as in C5a, cannot occur. An alternative docking might be found for an NH_2-terminal helix in C3a onto the rest of the molecule, but it appears unlikely that this helix could be accommodated in the crystal structure of C3a, as reported by Huber and his coworkers [(*9*) and figure 3 in (*9*)].

Implications of the model. Modeling of the C5a structure from C3a has produced a structure whose core is completely identical in the two proteins (Fig. 2). This strong correlation between residue identity and buried position suggests that the structures of these two molecules are closely conserved. It bestows a high level of confidence in the chosen sequence alignment (Fig. 1) and in the validity of the C5a model structure. In fact, the degree of conservation in the inside of these two molecules is even greater than is observed for closely related members of the serine protease (*12–14*) or aspartyl protease families (*24*). Certainly, the high divergence of the sequences of C3a and C5a (~64 percent different) indicates that sufficient evolutionary time has elapsed for changes to appear inside these molecules.

While the internal residues are all conserved in these two anaphylatoxins, the outside residues are almost all different. This result shows graphically that in nature the identical structural framework may occur in two proteins, yet they look to the outside world like very different molecules. This is quite appropriate, since C3a and C5a have very distinct functions that are mediated through very different parts of the two structures and interact with different cell receptors in performing their respective cellular functions.

In the absence of an experimental structure from C3a, the NH_2-terminal region of C5a was constructed as a helix and docked onto a patch of the rest of the molecule, which was chosen by its conservation between human and porcine C5a and by its hydrophobic nature. The resulting ensemble of four α helices forms a short-coiled coil of α helices. Such coiled helix interactions are observed in a variety of proteins with different numbers of helices, such as bacterial rhodopsin (*25*) and hemerythrin (*26*).

The apparent inability of the C3a sequence to form the same contact between the NH_2-terminal helix and the rest of the molecule as is proposed here for C5a does not exclude the possibility that an alternative interaction between an NH_2-terminal helix and the helical segments of the central portion of the molecule may occur in C3a. Somewhat different interactions between neighboring helices have been observed in the many different globin structures (*27*) and a similar change may be occurring between C3a and C5a.

Biological studies have shown that porcine C5a cross-reacts with human C5a receptors with similar activity to that of human C5a (*20*). This result implies that the receptor-binding residues in human and porcine C5a should be fairly well conserved. Therefore, the comparison of the human and porcine C5a sequences was reexamined in the light of the complete four-helix structure proposed for C5a (Fig. 5). The molecule continues to show that the changes between these two sequences are distributed throughout the surface of the structure. However, a single contiguous conserved region of surface can be found which includes residues 5[3] to 7[5], 9[7] to 15[13] and 20[19] through 26[25]. This area of the surface is a candidate for the receptor binding site that lies in the 69 residue NH_2-terminal fragment of C5a (*7*). Interestingly, this region of the surface is not very far from the COOH-terminal Arg residue which is also important for biological activity and thus probably interacts with the receptor. As was discussed earlier, the conformation of the COOH-terminal residues of C5a both in solution and when bound to the receptor is likely to be different from that in the C3a crystal (*9*). Preliminary modeling of the COOH-terminus shows that the Arg 76[74] side chain can approach closer to the NH_2-terminal region described above than the conformation indicated by the C3a crystal structure (Fig. 5). This would form a more compact but probably still discontinuous receptor binding site on C5a.

Thus, the model structure of C5a suggests a large number of directed experiments which may provide new information about the functional properties of C5a and its interaction with neutrophil receptors.

References and Notes

1. T. E. Hugli, in *CRC Critical Reviews in Immunology* (Chemical Rubber, Cleveland, 1981), vol. 3, pp. 321–366.
2. H. N. Fernandez and T. E. Hugli, *J. Biol. Chem.* **253**, 6955 (1978).
3. The residue numbers used in this work are presented in two ways. First the sequence numbers relative to the C3a molecule are given to facilitate comparison to the structure of C3a and its sequence. In brackets are the sequence numbers for C5a itself, to relate to the C5a literature. Both numbering systems are presented in Fig. 1 which can be used to interconvert.
4. C. Gerard and T. E. Hugli, *J. Biol. Chem.* **255**, 4710 (1978); B. Zimmermann and W. Vogt, *Hoppe Seyler's Z. Physiol. Chem.* **365**, 151 (1984).
5. H. N. Fernandez and T. E. Hugli, *J. Immunol.* **117**, 1688 (1978).
6. D. E. Chenoweth, B. W. Erickson, T. E. Hugli, *Biochem. Biophys. Res. Commun.* **86**, 227 (1979).
7. D. E. Chenoweth and T. E. Hugli, *Mol. Immunol.* **17**, 151 (1980).
8. E. P. Paques, H. Scholze, R. Huber, *Hoppe Seyler's Z. Physiol. Chem.* **361**, 977 (1980).
9. R. Huber, H. Scholze, E. P. Paques, J. Deisenhofer, *ibid.*, p. 1389.
10. W. J. Browne *et al.*, *J. Mol. Biol.* **42**, 65 (1969).
11. D. M. Shotton and H. C. Watson, *Nature (London)* **225**, 811 (1970).
12. J. Greer, *Proc. Natl. Acad. Sci. U.S.A.* **77**, 3393 (1980).
13. ———, *J. Mol. Biol.* **153**, 1027 (1981).
14. ———, *ibid.*, p. 1043.
15. L. Jurasek, R. W. Olafson, P. Johnson, L. B. Smillie, in *Proteolysis and Physiological Regulation* (proceedings of the Miami Winter Symposium), D. W. Ribbons and K. Brew, Eds. (Academic Press, New York, 1976), vol. 11, pp. 93–123.
16. T. Blundell, B. L. Sibanda, L. Pearl, *Nature (London)* **304**, 273 (1983); B. L. Sibanda *et al.*, *FEBS Lett.* **174**, 102 (1984).
17. N. E. Corbin and T. E. Hugli, *J. Immunol.* **117**, 990 (1976).
18. J. W. Jacobs *et al.*, *Biochemistry* **17**, 5031 (1978).
19. H. Domdey, K. Wiebauer, M. Kazmaier, V. Muller, K. Odink, G. Fey, *Proc. Natl. Acad. Sci. U.S.A.* **79**, 7619 (1982).
20. C. Gerard, D. E. Chenoweth, T. E. Hugli, *J. Immunol.* **127**, 1978 (1981).
21. P. Y. Chou and G. D. Fasman, *Annu. Rev. Biochem.* **47**, 251 (1978).
22. A. T. Hagler, P. S. Stern, R. Sharon, J. M. Becker, F. Naider, *J. Am. Chem. Soc.* **101**, 6842 (1979).
23. P. Dauber, D. Osguthorpe, A. T. Hagler, *Biochem. Soc. Trans.* **10**, 312 (1982).
24. J. Greer, unpublished results.
25. R. Henderson and P. N. T. Unwin, *Nature (London)* **257**, 28 (1975).
26. R. E. Stenkamp, L. C. Sieker, L. H. Jensen, *Acta Crystallogr.* **B38**, 784 (1982).
27. A. M. Lesk and C. Chothia, *J. Mol. Biol.* **136**, 225 (1980).

28. I thank Dr. George W. Carter and Karl W. Mollison for discussion, Drs. J. L. Fox and W. Mandecki for help with the Chou and Fasman calculations, and Drs. J. Deisenhofer and R. Huber for providing the refined coordinates for C3a.

22 August 1984; accepted 19 February 1985