THE JOURNAL OF BIOLOGICAL CHEMISTRY

Vol. 259, No. 6, Issue of March 25, pp. 3686-3693, 1984
Printed in U.S.A.

The Complete Amino Acid Sequence of a Biologically Active 197-residue Fragment of M Protein Isolated from Type 5 Group A Streptococci*

(Received for publication, September 19, 1983)

Belur N. Manjula‡, A. Seetharama Acharya§, Sheenah M. Mische, Thomas Fairwell¶, and Vincent A. Fischetti‖

From The Rockefeller University, New York, New York 10021 and the ¶Molecular Disease Branch, National Heart, Lung and Blood Institute, National Institutes of Health, Bethesda, Maryland 20205

The complete amino acid sequence of a peptic fragment (Pep M5) of the group A streptococcal type 5 M protein, the antiphagocytic cell surface molecule of the bacteria, is described. This fragment, comprising nearly half of the native M molecule, is biologically active in that it has the ability to interact with opsonic antibodies as well as to evoke such an antibody response in rabbits. The sequence of Pep M5 was determined by automated Edman degradations of the uncleaved molecule and its enzymatically derived peptides. The primary peptides for Edman degradation were the arginine peptides obtained by tryptic digestion. The tryptic cleavage of Pep M5 was limited to the arginyl peptide bonds by derivatizing the ϵ-amino groups of lysine residues by reductive dihydroxypropylation. The overlapping peptides were generated by digestion of the unmodified Pep M5 with chymotrypsin, V8 protease, and subtilisin. The sequence thus established for the Pep M5 molecule consists of a total of 197 residues (M_r = 22,705).

The Pep M5 protein contains some identical, or nearly so, repeating sequences: four 7-residue segments and two 10-residue segments. However, extensive sequence repeats of the kind previously reported within the partial sequence of another M protein serotype, namely Pep M24, were absent. The Pep M5 sequence is distinct from, but exhibits some homology with, the partial sequences of two other M protein serotypes, namely, Pep M6 and Pep M24. Furthermore, the 7-residue periodicity of the nonpolar and charged residues, an α-helical coiled-coil structural characteristic that was previously observed within the partial sequences of M proteins, was found to extend over a significant part of the Pep M5 sequence. The implication of these results to the function and immunological diversity in M proteins is discussed.

M protein of the group A streptococcus (*Streptococcus pyogenes*) is closely associated with the virulence of the bacteria because of its unique property of impeding phagocytosis and hence plays a major role in the pathogenicity of these microorganisms (Lancefield, 1962; Fox, 1974). It is present on the cell surface and is a constituent of the cell wall of the streptococcus (McCarty, 1964; Swanson *et al.*, 1969). Following a streptococcal infection, M protein induces the production of protective antibodies which neutralize its antiphagocytic property and hence facilitate phagocytosis of the bacteria. However, a large number of antigenic types of the M protein have been recognized and, although immunological cross-reactions have been observed among some of these, the acquired immunity to group A streptococcal infections in humans is type specific (Lancefield, 1959, 1962; Wiley and Bruno, 1968, 1969; Berger-Rabinowitz *et al.*, 1972; Fischetti, 1978). Thus, variation of its antigenic structure while still retaining the antiphagocytic property appears to be one of the mechanisms by which these bacteria survive under immunological pressure in the human, their natural host.

The native M protein molecule is approximately 58,000 daltons in size (van de Rijn and Fischetti, 1981; Scott and Fischetti, 1983) and exists as an α-helical coiled-coil molecule on the cell surface (Phillips *et al.*, 1981; Fischetti *et al.*, 1982). However, limited proteolysis of the streptococcal cells with pepsin releases a fragment of the native M molecule namely, Pep M protein[1] (Beachey *et al.*, 1977, 1980b; Manjula and Fischetti, 1980a). This fragment retains the ability to interact with opsonic antibodies and is capable of eliciting such a response in rabbits. Hence, this peptic fragment is a biologically active fragment of the native M molecule. Our previous studies (Phillips *et al.*, 1981) have shown that this fragment is derived from the distal portion of the M protein fibrillae on the streptococcal cell wall and accounts for nearly half of the native M molecule.

The amino acid sequence information available to date on the streptococcal M proteins is rather limited. Only partial sequences of the biologically active fragments derived from three serotypes namely, 5, 6,[2] and 24 are presently known (Beachey *et al.*, 1978, 1980a, 1980b; Fischetti *et al.*, 1976, Fischetti and Manjula, 1982; Manjula and Fischetti, 1980a,

* The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

This work is dedicated to the memory of Professor Rebecca Lancefield who contributed immensely to our knowledge of streptococcal M proteins.

‡ Supported by Grant AHA 83-1102 from the American Heart Association with funds contributed in part by the New York Heart Association. An Established Investigator of the American Heart Association.

§ Supported by United States Public Health Service Grant HL-27183. An Established Fellow of the New York Heart Association.

‖ Supported by United States Public Health Service Grants AI-11822 and HL-25219.

[1] The abbreviations used are: Pep M protein, M protein isolated from the streptococci by limited proteolysis with pepsin; Det M6, M6 protein isolated from type 6 streptococci by nonionic detergent extraction; DHP, dihydroxypropyl; HPLC, high performance liquid chromatography; TPCK, L-1-tosylamido-2-phenylethyl chloromethyl ketone; TLCK, N^{α}-*p*-tosyl-L-lysine chloromethyl ketone; PTH: phenylthiohydantoin.

[2] B. N. Manjula and V. A. Fischetti, unpublished observations.

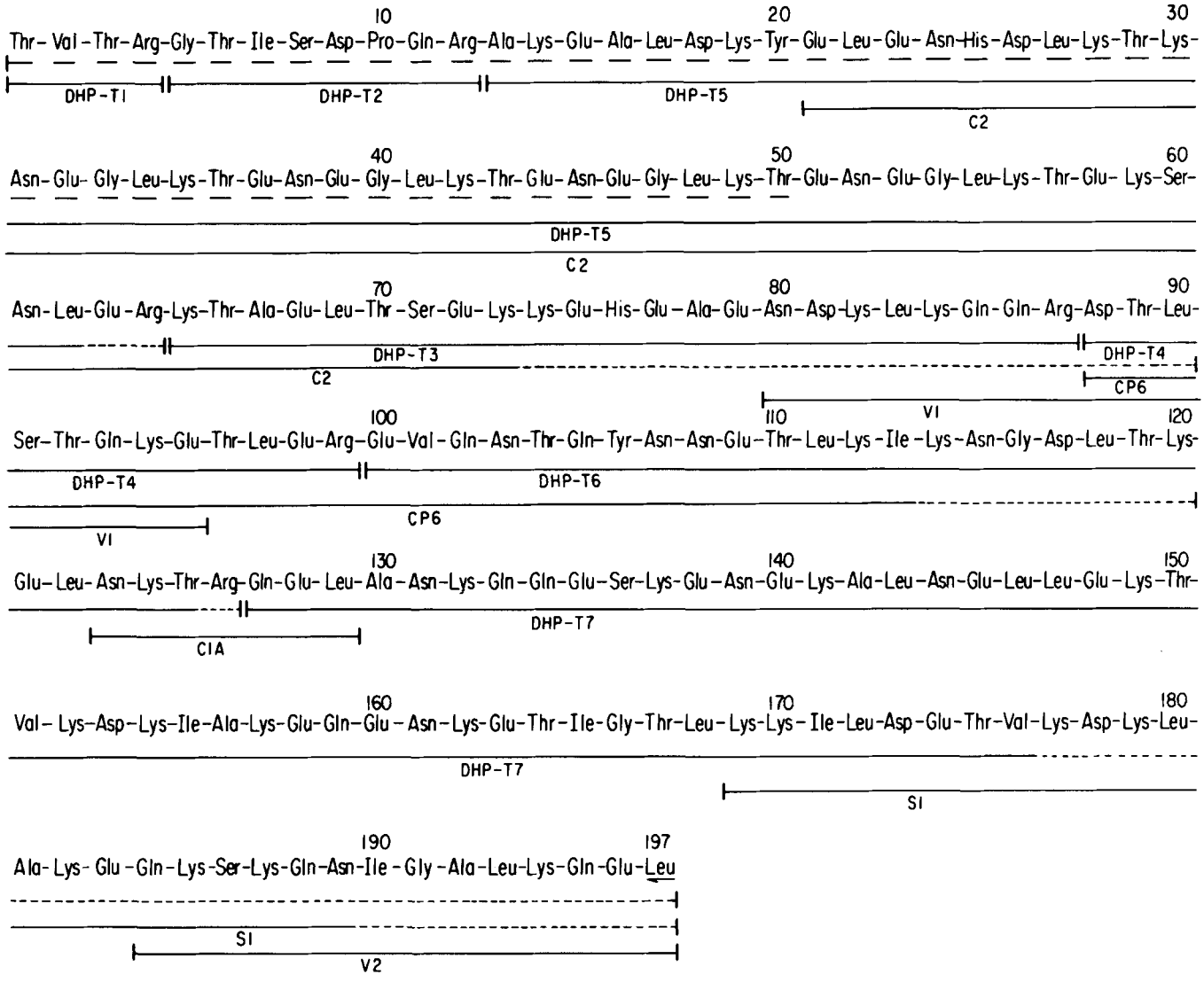

FIG. 1. **Summary of the evidence for the complete sequence of streptococcal Pep M5 protein.** Peptide designations are immediately under the sequence. Sequence determined by Edman degradation of uncleaved Pep M5 (- - -). For peptides, *solid lines* represent the extent of sequence obtained from each peptide and the *dashed lines* indicate the remainder of that particular peptide. ⟵, represents residue determined by carboxypeptidase Y digestion.

1980b, 1982; Manjula *et al.*, 1982, 1983; Seyer *et al.*, 1980). These data, however, have revealed that the partial sequences from the three M protein serotypes are distinct but exhibit a common 7-residue periodicity in the distribution of their nonpolar and charged residues, a feature characteristic of coiled-coil molecules (Manjula and Fischetti, 1980b, Fischetti and Manjula, 1982; Manjula *et al.*, 1983).

A fuller understanding of the structure-function relationships in M proteins warrants knowledge of the complete amino acid sequence of at least a few M protein serotypes. In this report we present the complete amino acid sequence of Pep M5, a biologically active peptic fragment of one of the M protein serotypes namely, the M5 protein. Pep M5 is thus the first such biologically active fragment of an M protein serotype to be sequenced completely. This sequence has now been examined to determine the extent to which the above-mentioned 7-residue periodicity extends within this molecule. In addition, this sequence has been compared with the available partial sequence data on the M6 and M24 molecules to understand the structural basis for the immunological diversity of these M protein molecules.

EXPERIMENTAL PROCEDURES[3]

RESULTS AND DISCUSSION

Amino Acid Sequence of Pep M5—A summary of the data leading to the determination of the complete amino acid sequence of the Pep M5 molecule is delineated in Fig. 1. The sequence of the first 50 residues of Pep M5 was determined directly by Edman degradation of the whole molecule (Manjula and Fischetti 1980b). The remainder of the sequence was assembled from the sequence analysis of 11 of the peptides

[3] Portions of this paper (including "Experimental Procedures," part of "Results and Discussion," Tables 1–11, Figs. 6–9, and additional references are presented in miniprint at the end of this paper. Miniprint is easily read with the aid of a standard magnifying glass. Full size photocopies are available from the Journal of Biological Chemistry, 9650 Rockville Pike, Bethesda, MD 20814. Request Document No. 83M-2711, cite the authors, and include a check or money order for $9.60 per set of photocopies. Full size photocopies are also included in the microfilm edition of the Journal that is available from Waverly Press.

generated by several enzymic digestions of the Pep M5 molecule.

In order to reduce the complexity during sequence analysis, the strategy was to fragment the Pep M5 molecule so as to obtain a small number of large peptides which would be amenable for sequencing by automated Edman degradation. Pep M5 contains 6 arginines, 35 lysines, and 47 glutamates, with no methionine and tryptophan, thus limiting the choice for obtaining large peptides to cleavage at the arginyl peptide bonds. The framework for the sequence presented in Fig. 1 was obtained from the sequences of the arginine peptides of DHP-Pep M5.[4] The overlapping peptides were obtained by chymotryptic, V8 protease, and subtilisin digestions.

The previous knowledge of the NH_2-terminal sequences of the clostripain peptides of Pep M5 (Manjula *et al.*, 1983) was valuable in reinforcing the sequences of the tryptic peptides of DHP-Pep M5. Furthermore, the DHP-lysine residues identified in the peptides of DHP-Pep M5 corresponded to the lysine residues in the clostripain peptides. In addition, during the course of isolation of the overlapping peptides and peptides from the carboxyl-terminal region of the molecule, a large number of other peptides, obtained by independent cleavage methods, were also isolated and sequenced. While these inevitably provided a large amount of redundant information, it enabled us to assign the sequence of each region with great certainty.

Special Features of the Pep M5 Molecule—The amino acid composition of Pep M5 based on the sequence data is shown in Table I. The molecular weight of the protein, based on the sequence data, was calculated to be 22,705. One of the special features of the Pep M5 molecule that is worth noting is the inverse distribution of the acid and the amide forms of aspartic and glutamic acid residues. While the aspartates are present predominantly as amides, the glutamates are present predominantly in the acid form; 65% of the aspartates are present as amides, whereas only 28% of the glutamates are in this form. Also, the total number of acidic residues (Asp + Glu = 43) is equal to the total number of basic residues (Lys + Arg + His = 43). Furthermore, the acidic and basic residues occur quite frequently in pairs.

Sequence Repeats—In an earlier study (Manjula and Fischetti, 1980b), we reported that within the NH_2-terminal region of the Pep M5 molecule a 7-residue sequence, namely Leu-Lys-Thr-Glu-Asn-Glu-Gly occurs tandemly three times (residues 27–33, 34–40, and 41–47), with a single substitution of Lys/Glu at position 30. With the completion of the sequence of Pep M5, it has now been found that this sequence extends tandemly once more (residues 48–54) and is followed by the sequence Leu-Lys-Thr-Glu (residues 55–58) (Fig. 2*A*). Furthermore, as can be seen from Fig. 2*B*, an additional sequence repeat was found in the COOH-terminal quarter of the Pep M5 molecule. The two 10-residue segments, namely residues 150–159 and 175–184, are identical with each other but for a single conservative substitution of an Ile/Leu at 155/180. Moreover, regions 150–172 and 175–197, comprising the above segments, are highly homologous to each other (65% identity and 87% homology).

The only other M proteins for which any significant sequence information is available are Pep M6[2] (Beachey *et al.*, 1980b; Manjula and Fischetti, 1980b) and Pep M24 (Beachey *et al.*, 1978, 1980a). These are biologically active peptic fragments analogous to Pep M5 isolated from type 6 and type 24 streptococcal cells, respectively. Unlike Pep M5, Pep M24 contains 6 methionine residues. Partial sequence analysis of

[4] A. S. Acharya, L. G. Sussman, and B. N. Manjula, manuscript in preparation.

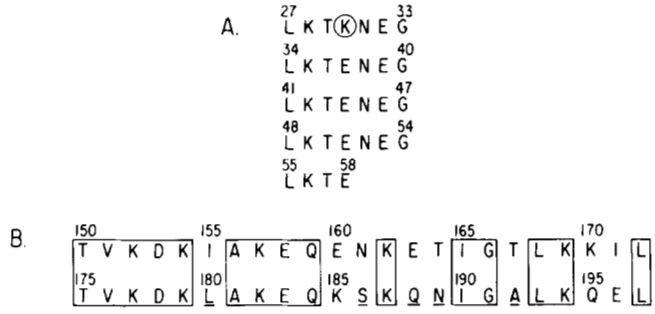

FIG. 2. **Sequence repeats in Pep M5.** *A*, A 7-residue sequence occurs tandemly 4½ times within the region 27–58 of Pep M5, with the exception of a Lys/Glu substitution at position 30 (*circled*). *B*, Homology between segments 150–172 and 175–197 of Pep M5. Identical residues are *boxed* and conservative substitutions (Dayhoff, 1976) are *underlined*.

the cyanogen bromide peptides of Pep M24 has revealed that the sequence of this molecule is highly repetitious (Beachey *et al.*, 1978). It contains at least two identical copies of a 27-residue segment and five identical copies of a 20-residue segment. The sequences of two of the latter fragments, both of which are 35 residues long, are virtually identical with each other, with only three substitutions (Beachey *et al.*, 1980a). Therefore, it is clear from the results presented here that although the Pep M5 protein contains some sequence repeats (Fig. 2), they are not as extensive as those present in the Pep M24 molecule.

7-residue Periodicity in the Pep M5 Protein—A heptad distribution in the nonpolar and charged amino acid residues, a characteristic of α-helical coiled-coil fibrous proteins such as tropomysoin (Stone *et al.*, 1975; McLachlan and Stewart, 1975), myosin rod region (Elzinga and Trus, 1980; McLachlan and Karn, 1982), *Escherichia coli* lipoprotein (McLachlan, 1978), and other fibrous proteins (Parry *et al.*, 1977; Doolittle *et al.*, 1978), was described earlier within the partial sequences of the clostripain peptides of the Pep M5 protein (Manjula *et al.*, 1983). Examination of the complete amino acid sequence of the Pep M5 molecule has now revealed that the repeating heptad periodicity, although not highly regular, is present at least through residue 149 of this molecule. It therefore appears that a significant part of the Pep M5 molecule can exist in the coiled-coil conformation.

Implications of the M Protein Sequence to Its Antigenic Diversity—Since it has long been recognized that the natural host for the group A streptococcus is the human, antigenic variation of the M protein, through immunological pressure, is a possible means by which the bacteria are able to survive in this environment (Lancefield, 1962). The complete amino acid sequence of Pep M5 presented here as well as the partial sequence data available for Pep M6[2] (Beachey *et al.*, 1980b; Manjula and Fischetti, 1980b) and Pep M24 (Beachey *et al.*, 1978, 1980a), clearly indicate considerable divergence in the sequences of the various M protein serotypes. Despite this divergence, a certain degree of homology among these sequences is also apparent (Figs. 3–5). The NH_2-terminal sequences of Pep M5, Pep M6, and Pep M24, although not identical, are highly homologous to one another (Fig. 3). In addition, the various segments of Pep M6 and Pep M24 also show a certain degree of homology with different regions of the Pep M5 molecule (Figs. 4 and 5). Because of the repetitious nature of the Pep M24 molecule, the homology of a given segment of Pep M5 to a repeating region of the Pep M24 would indicate, in essence, homology of a single region of Pep M5 to multiple regions in the Pep M24 molecule. More

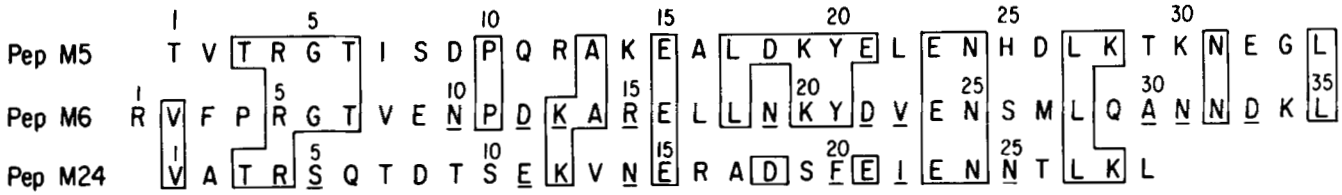

FIG. 3. **Comparison of the NH_2-terminal sequences of Pep M5, Pep M6,[2] and Pep M24 (Beachey *et al.*, 1978).** Identical residues are *boxed* and conservative substitutions (Dayhoff, 1976) are *underlined.*

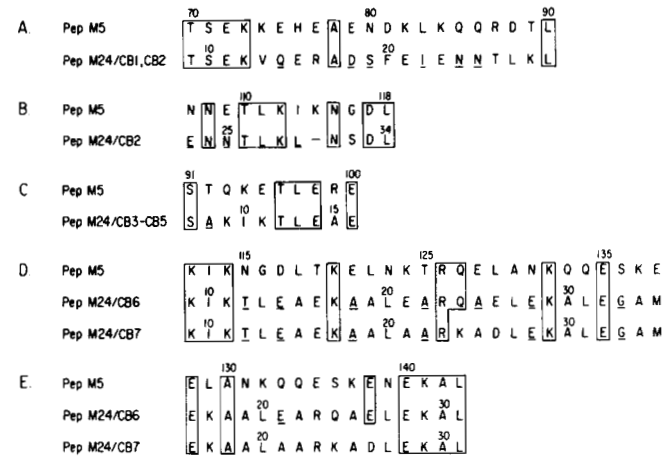

FIG. 4. **Comparison of the partial sequences of Pep M24 (Beachey *et al.*, 1978, 1980a) to various regions of the Pep M5 molecule.** Identical residues are *boxed* and conservative substitutions (Dayhoff, 1976) are *underlined. A,* Segment 9–29 of Pep M24/CB1, CB2 is compared to segment Pep M5/(70–90) (29% identity). *B,* The homology (58% identity) of segments 23–34 of Pep M24/CB2 with Pep M5/(107–118). *C,* The homology (50% identity) of segment 7–16 of Pep M24/CB3-CB5 with Pep M5/(91–100). *D,* The homology of residues 9–35 of Pep M24/CB6 and CB7 with Pep M5/(112–138), revealing 30% and 26% identity, respectively. *E,* The homology (44% identity) of residues 16–31 of Pep M24/CB6 and CB7 with Pep M5/(128–143).

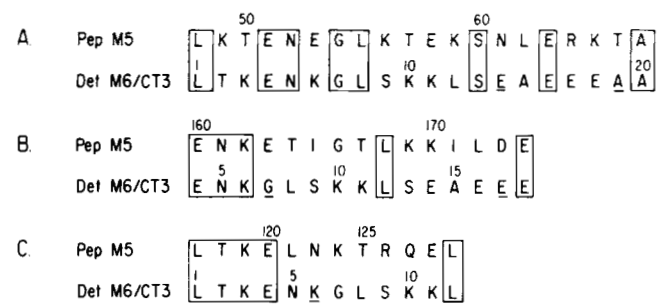

FIG. 5. **Comparison of a region of the M6 molecule namely Det M6/CT3 (Manjula and Fischetti, 1980b) to various regions of Pep M5.** *A,* The homology (40% identity) of residues 1–20 of Det M6/CT3 with Pep M5/(48–67). *B,* The homology (30% identity) of Det M6/CT3(4–18) with Pep M5/(160–174). *C,* The homology (42% identity) of Det M6/CT3(1–12) with Pep M5/(117–129).

sequencing of other M proteins such as Pep M6 and Pep M24 will be required to evaluate the significance of such homology.

The mechanism by which antigenic variations occur among the different M protein serotypes is not clearly understood at present. In addition to the sequence differences among serotypes, limited variations involving conservative as well as nonconservative substitutions have been observed between Pep M5 molecules isolated from two different strains of the same serotype, namely type 5 streptococcus (Beachey *et al.*, 1982; Manjula and Fischetti, 1980a, 1980b; Seyer *et al.*, 1980). Pep M24, being highly repetitious in sequence, appears likely to have evolved by gene duplication. However, as to whether or not this is a primitive structure from which many of the other serotypes evolved requires more sequence as well as genetic data.

Thus, the knowledge of the primary structure of the Pep M5 protein along with the determination of the primary structure of similar biologically active fragments from other M protein serotypes should help us elucidate more clearly the relation between the amino acid sequence and the antiphagocytic property of M protein. This understanding of the structure-function relations in the M protein molecule may lead to a better understanding of the molecular basis of streptococcal diseases such as rheumatic fever and glomerulonephritis.

Acknowledgments—We wish to thank Drs. Maclyn McCarty, Emil C. Gotschlich, and John B. Zabriskie for their encouragement and continued interest in these studies, and Dr. James M. Manning for his valuable comments on the manuscript.

REFERENCES

Beachey, E. H., Stollerman, G. H., Chiang, E. Y., Chiang, T. M., Seyer, J. M., and Kang, A. H. (1977) *J. Exp. Med.* **145,** 1469–1483

Beachey, E. H., Seyer, J. M., and Kang, A. H. (1978) *Proc. Natl. Acad. Sci. U. S. A.* **75,** 3163–3167

Beachey, E. H., Seyer, J. M., and Kang, A. H. (1980a) *J. Biol. Chem.* **255,** 6284–6289

Beachey, E. H., Seyer, J. M., and Kang, A. H. (1980b) in *Streptococcal Diseases and the Immune Response* (Zabriskie, J. B., and Read, S. E., eds) pp. 149–160, Academic Press, New York

Berger-Rabinowitz, S., Ofek, I., and Moody, M. D. (1972) *J. Infect. Dis.* **125,** 339–344

Dayhoff, M. O. (1976) in *Atlas of Protein Sequence and Structure* **5,** Suppl. 2, 3–8 and 311

Doolittle, R. F., Goldbaum, D. M., and Doolittle, L. R. (1978) *J. Mol. Biol.* **120,** 311–325

Elzinga, M., and Trus, B. (1980) in *Methods Peptide* and *Protein Sequence Analysis* (Birr, Chr., ed) pp. 213–224, Elsevier/North-Holland Biomedical Press, Amsterdam

Fischetti, V. A., Gotschlich, E. C., Siviglia, G., and Zabriskie, J. B. (1976) *J. Exp. Med.* **144,** 32–53

Fischetti, V. A. (1978) *J. Exp. Med.* **147,** 1771–1778

Fischetti, V. A., and Manjula, B. N. (1982) *Semin. Infect. Dis.* **4,** 411–418

Fischetti, V. A., Manjula, B. N., Fazio, M., Cohen, C., Flicker, P. F., and Phillips, G. N., Jr. (1982) in *Basic Concepts of Streptococci and Streptococcal Diseases* (Holme, S. E., and Christensen, P., eds) pp. 85–87, Reedbooks Ltd, England

Fox, E. N. (1974) *Bacteriol. Rev.* **35,** 57–86

Lancefield, R. C. (1959) *J. Exp. Med.* **110,** 271–292

Lancefield, R. C. (1962) *J. Immunol.* **89,** 307–313

Manjula, B. N., and Fischetti, V. A. (1980a) *J. Immunol.* **124,** 261–267

Manjula, B. N., and Fischetti, V. A. (1980b) *J. Exp. Med.* **151,** 695–708

Manjula, B. N., and Fischetti, V. A. (1982) in *Basic Concepts of Streptococci and Streptococcal Diseases* (Holme, S. E., and Christensen, P., eds) pp. 79–81, Reedbooks Ltd, England

Manjula, B. N., Mische, S. M., and Fischetti, V. A. (1982) in *Methods in Protein Sequence Analysis* (Elzinga, M., ed) pp. 539–540, Humana Press, New Jersey

Manjula, B. N., Mische, S. M., and Fischetti, V. A. (1983) *Proc. Natl. Acad. Sci. U. S. A.* **80,** 5475–5479
McCarty, M. (1964) in *The Streptococcus, Rheumatic Fever, and Glomerulonephritis* (Uhr, J. W., ed) Williams and Wilkins Co., Baltimore
McLachlan, A. D., and Stewart, M. (1975) *J. Mol. Biol.* **98,** 293–304
McLachlan, A. D. (1978) *J. Mol. Biol.* **121,** 493–506
McLachlan, A. D., and Karn, J. (1982) *Nature (Lond.)* **299,** 226–231
Parry, D. A. D., Crewther, W. G., Fraser, R. D. B., and MacRae, T. P. (1977) *J. Mol. Biol.* **113,** 449–454
Phillips, G. N., Jr., Flicker, P. F., Cohen, C., Manjula, B. N., and Fischetti, V. A. (1981) *Proc. Natl. Acad. Sci. U. S. A.* **78,** 4689–4693
Scott, J. R., and Fischetti, V. A. (1983) *Science (Wash. D. C.)* **221,** 758–760
Seyer, J. M., Kang, A. H., and Beachey, E. H. (1980) *Biochem. Biophys. Res. Commun.* **92,** 546–553
Stone, D. J., Sodex, J., Johnson, P., and Smillie, L. B. (1975) in *Proceedings of the IXth Federation of European Biochemical Societies Meeting* (Biro, E. N. A., ed) **31,** 125–136, Elsevier/North-Holland, New York
Swanson, J., Hsu, K. C., and Gotschlich, E. C. (1969) *J. Exp. Med.* **130,** 1063–1091
van de Rijn, I., and Fischetti, V. A. (1981) *Infect. Immun.* **32,** 86–91
Wiley, G. G., and Bruno, P. N. (1968) *J. Exp. Med.* **128,** 959–968
Wiley, G. G., and Bruno, P. N. (1969) *J. Immunol.* **103,** 149–154
Additional references are found on p. 3692.

Supplemental material to

THE COMPLETE AMINO ACID SEQUENCE OF A BIOLOGICALLY ACTIVE 197 RESIDUE FRAGMENT

OF M-PROTEIN ISOLATED FROM TYPE 5 GROUP A STREPTOCOCCI*

by

Belur N. Manjula, A. Seetharama Acharya, Sheenah M. Mische,

Thomas Fairwell and Vincent A. Fischetti

EXPERIMENTAL PROCEDURES

MATERIALS

Sequanal grade phenylisothiocyanate, heptane, 0.1M Quadrol, ethyl acetate, heptafluoro-butyric acid were products of Beckman Instruments. Glass distilled benzene and butyl chloride were products of Burdick and Jackson. Polybrene and sodium cyanoborohydride were purchased from Aldrich chemical Company; 4-methanesulfonic acid and 6N HCl were from Pierce Chemical Co.; HPLC-grade acetonitrile was from Fisher Scientific Company; TPCK-trypsin and carboxypeptidase Y were from Worthington Biochemicals; S. aureus V8 protease was from Miles Laboratories; TLCK-chymotrypsin, subtilisin (type VIII) and pepetatin were purchased from Sigma Chemical Co.

METHODS

Isolation of Pep M5: Pep M5 was isolated from the type 5 streptococcal cells by limited proteolysis with pepsin, as previously described (Manjula and Fischetti, 1980a). The protein was subjected to fragmentation by several enzymes, and peptides for sequencing were isolated exclusively by HPLC. The criterion of purity of any given peptide was the presence of a single sequence during automated Edman degradation.

Dihydroxypropylation: The ε-amino groups of lysine residues of Pep M5 were blocked by reductive alkylation with glyceraldehyde (footnote 4):

$$R\text{-}NH_2 + OHC\text{-}CHOH\text{-}CH_2OH \xrightarrow[37^{\circ}C]{NaCNBH_3} R\text{-}NH\text{-}CH_2\text{-}CHOH\text{-}CH_2OH$$

Glyceraldehyde — Dihydroxypropylated-Protein

Pep M5 (4mg/ml) in 0.01M sodium phosphate-0.15M sodium chloride, pH 7.4, was reacted with 100mM glyceraldehyde in the presence of 1M sodium cyanoborohydride at 37°C for 1/2h. The excess reagents were removed by dialysis against 0.1M ammonium bicarbonate and the sample was lyophilized. Amino acid analysis of the acid hydrolyzate of the dihydroxypropylated Pep M5 (DHP-Pep M5) revealed the total absence of unmodified lysine residues.

HPLC: Reverse phase HPLC fractionation of peptides was carried out with a Waters HPLC system on either Whatman ODS 2 (Partisil 10) or Waters u-Bondapak C18 column using an acetonitrile-0.05% TFA gradient for elution. An ISCO variable wavelength detector was used to monitor the elution of peptides. Appropriate fractions were pooled and taken to dryness in a Savant Speedvac concentrator centrifuge. The selected peptides are numbered in the order of their elution during fractionation by HPLC. Amino acid compositions are shown only for those peptides for which the sequence data is shown.

Trypsin Peptides of Pep M5: DHP-Pep M5 (2mg/ml) in 0.2M ammonium bicarbonate was digested with TPCK-Trypsin (E/S 1:100) at 37°C for 4 1/2h. The peptide bonds of ε-DHP-lysine residues are resistant to tryptic digestion and therefore the tryptic cleavage is limited to the arginyl peptide bonds in the Pep M5 molecule. The reaction was terminated by freeze drying. Fractionation of the tryptic digest by HPLC is shown in Fig. 6.

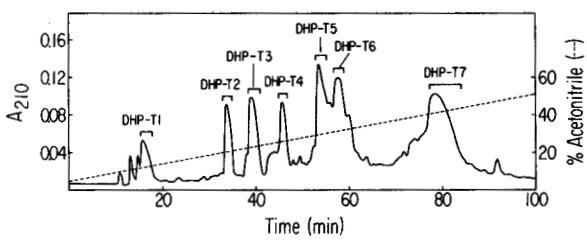

Fig. 6: Fractionation of the tryptic peptides of DHP-Pep M5 by HPLC on a Waters u-Bondapak C18 reverse phase column. Peptides were eluted with a linear gradient (5-70%) of acetonitrile-0.05% TFA, over a period of 140 min. The flow rate was 1ml/min. Fractions were pooled as indicated by the bars.

Chymotryptic Peptides of Pep M5: Pep M5 (1mg/ml) in 0.05M ammonium bicarbonate pH 7.8, was digested with TLCK-chymotrypsin (E/S 1:500) at room temperature for 5h. The digestion was terminated by freeze drying. Fractionation of the chymotryptic peptides by HPLC is shown in Fig. 7A.

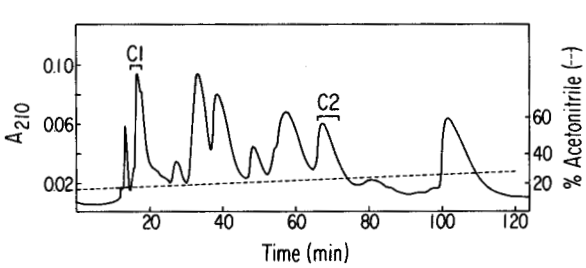

Fig. 7A: HPLC fractionation of the chymotryptic digest of Pep M5. The peptides were fractionated on a Whatman ODS-2 (Partisil 10) reverse phase column. A linear gradient of 15-30% acetonitrile-0.05% TFA over 120 min was employed. All peaks were pooled and their amino acid compositions as well as sequence determined. While most of these results provided very valuable data for confirmation of the sequences of various regions of the Pep M5 molecule established by peptides obtained by other enzymic digestions, peptides C1 and C2 proved useful to establish two overlaps. Peptide C1 was further purified by rechromatography (see Fig. 7B) prior to sequence analysis.

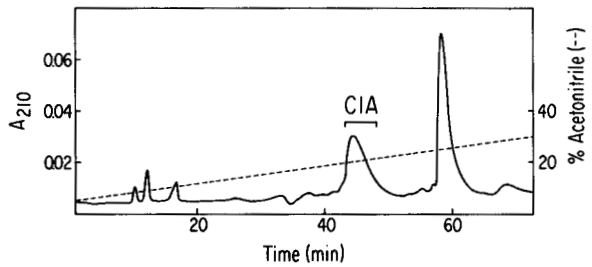

Fig. 7B: Peak C1 from Fig. 7A was rechromatographed on the Whatman ODS-2 column. Gradient: 5-40% acetonitrile-0.05% TFA over 100 min. Peak C1A was isolated for sequence determination.

V8 protease Peptides of Pep M5: Pep M5 (2mg/ml) in 0.05M ammonium bicarbonate was digested with V8 protease (Drapeau, 1977), at an E/S ratio of 1:500, at room temperature for 25h. The digestion was terminated by freeze drying. Fractionation of the V8 digest of Pep M5 by HPLC is shown in Fig. 8.

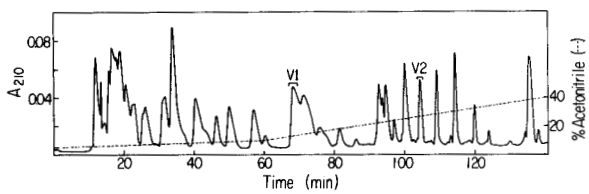

Fig. 8: HPLC of the V8 digest of Pep M5. HPLC conditions are the same as in Fig. 6, but for the gradient being 5-10% acetonitrile-0.05% TFA over the first 60 min and 10-40% of the same solvent over the next 80 min. Based on their amino acid compositions (Table 2), peptides V1 and V2 were selected for sequence analysis.

Subtilisin Peptides of Pep M5: Limited digestion of Pep M5 (2mg/ml) with subtilisin was carried out in 0.05M ammonium bicarbonate (E/S 1:500) for 1h at 0°C. The reaction was terminated by the addition of 1M HCl to reduce the pH to about 1.0 and the digest freeze dried. Fractionation of the digest by HPLC is shown in Fig. 9.

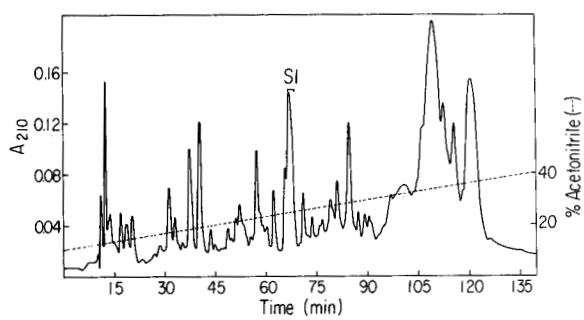

Fig. 9: HPLC of the subtilisin digest of Pep M5. HPLC conditions are the same as in Fig. 6, but for the gradient being 10-40% acetonitrile-0.05% TFA over 140 min. Peptide S1 was used for sequence analysis (see Table 2 for amino acid composition).

Amino Acid Analysis: This was carried out on methanesulfonic acid hydrolyzates of the peptide samples, on a Durrum D500 Amino Acid Analyzer, as described earlier (Manjula and Fischetti, 1980a).

Sequence Analysis: Amino-terminal sequence analysis of peptides (20-70 nmol) was carried out by automated Edman degradation in a Beckman updated 890B sequencer equipped with a Sequemat P6 autoconverter and 80510 programmer. A modified dilute Quadrol program (Beckman 011576) was used for sequencing. Methanolic-HCl was used for the conversion of ATZ-amino acid to the PTH-amino acids. Polybrene (6 mg) was used as carrier to facilitate retention of peptides during sequencing (Klapper et al, 1978), and three complete cycles were run with glycylglycine (50 nmol) before application of the peptide samples. The PTH-amino acids were identified and quantitated by HPLC (Zimmerman et al, 1977) on a Dupont Zorbax ODS column using a Hewlett-Packard Model 1084B instrument. PTH-serine and PTH-threonine were also detected as their dehydro-derivatives.

Some of the residues were also identified by chemical ionization mass spectrometry in a Finnigan 4510 Instrument using isobutane as the reagent gas (Fairwell, 1983).

Carboxypeptidase digestion: The carboxy terminal residue in Pep M5 was determined by digestion with carboxypeptidase Y (Hayashi, 1977) in 0.1 M ammonium acetate buffer, pH 5.5, at E/S = 1:50, at room temperature, in the presence of 17 ug/ml of pepstatin. Norleucine was used as an internal standard. The digestion was terminated by addition of a 5 fold excess of 1M acetic acid followed by lyophilization. The released amino acid was determined by amino acid analysis. The ratio of amino acids released after a 30 min digestion was Leu (1.0), Glu (0.3) and Gln (0.16).

Nomenclature of Peptides: Peptides have been designated according to the enzyme used for cleavage as follows: C, chymotrypsin; CP, clostripain; S, subtilisin and V, S. aureus V8 protease. Peptides arising from trypsin digestion of dihydroxypropylated Pep M5 have been designated DHP-T. To avoid confusion, the numbering used for the peptides corresponds to their positions within the completed sequence of the Pep M5 molecule.

RESULTS AND DISCUSSION

Arginine peptides of Pep M5: In an earlier study (Manjula et al, 1982, 1983) we used the enzyme clostripain to specifically cleave the Pep M5 molecule at its arginine residues. The clostripain peptides were isolated in sufficient yield and their sequences provided very valuable information. However, it was observed that in addition to the major arginine cleavage, some secondary cleavages were also obtained. Due to the high lysine content of Pep M5, these secondary cleavages could have been at the lysyl peptide bonds, as has been reported previously for ribonuclease (Mitchell, 1977). Therefore, in order to establish unambiguously the arginine sites in Pep M5, in the present study we have taken advantage of the specificity of the enzyme trypsin on lysine-modified Pep M5.

In order to limit the tryptic cleavage to the arginyl peptide bonds, the ϵ-amino groups of the lysine residues of the Pep M5 molecule were first modified by reductive dihydroxypropylation with glyceraldehyde (footnote 4) and then the dihydroxypropylated protein, DHP-PepM5, was digested with trypsin. The resulting tryptic peptides namely, DHP-T1 through DHP-T7, were fractionated by HPLC (Fig. 6) and their amino acid compositions are presented in Table 1. The size of the tryptic peptides of DHP-Pep M5 varied from 4 to 72 residues. The amino acid compositions of these peptides together accounted for all the six arginine residues as well as to the total composition of the Pep M5 molecule. DHP-T7 does not contain any arginine residue and hence corresponds to the C-terminal region of the Pep M5 molecule.

Based on the previous knowledge of the sequences of the amino terminal region of the Pep M5 molecule (Manjula and Fischetti, 1980b) and its clostripain peptides (Manjula et al, 1983), DHP-T1, DHP-T2 and DHP-T4 were identified to be residues 1-4, 5-8 and 88-99, respectively. The remaining peptides namely, DHP-T3, DHP-T5, DHP-T6 and DHP-T7 were sequenced by automated Edman degradation (Tables 3-6). It may be added here that the presence of ϵ-DHP-lysine in these peptides did not interfere with their sequencing by Edman degradation. Furthermore, the PTH-derivative of DHP-lysine had a distinct elution time on the Zorbax ODS column and eluted after PTH-proline and before PTH-tryptophan, thus facilitating easy identification of the DHP-lysine residues in the Pep M5 molecule (footnote 4).

DHP-T3 was sequenced through its carboxyl terminal residue (23 residues) and DHP-T6 was sequenced through its penultimate residue (26 residues). The larger peptides DHP-T5 and DHP-T7 were both sequenced through their first 50 residues. Thus the results of the Edman degradation of the tryptic peptides of DHP-Pep M5, together with the previous knowledge of the amino terminal sequence of the whole molecule, provided the sequence of 173 out of the 197 residues (88%) of the Pep M5 molecule.

Completion of the C-terminal sequence: DHP-T7, the 72 residue carboxy terminal peptide, was first sequenced through its amino terminal 50 residues. This corresponds to residues 127 through 176 of the Pep M5 molecule. The sequence of the remainder of this peptide (residues 177-197) was established by the Edman degradation of the subtilisin peptide S1 and the V8 protease peptide V2 (see Table 2 for amino acid compositions). Peptide S1, starting at residue 169, was sequenced through its amino terminal 21 residues to obtain the sequence of residues 169-189 (Table 11). On the other hand, the V8 peptide V2, a 14 residue peptide, was sequenced in its entirety (Table 10), and was found to overlap with the subtilisin peptide S1 at residue 184 and hence provided the sequence of residues 184-197 of the Pep M5 molecule. Thus, the sequences of S1 and V2 together accounted for the region 169-197, and provided a eight residue overlap with the sequenced region of DHP-T7. These data, therefore established the complete sequence of DHP-T7 and hence region 127-197 of the pep M5 molecule. Furthermore, carboxypeptidase Y digestion of uncleaved Pep M5 indicated leucine to be its C-terminal residue. Since the V8 peptide V2 terminates with a leucine residue, it was concluded that residue 197 is the carboxy-terminal residue of the Pep M5 molecule.

Overlaps And Alignment Of Peptides: Alignment of the tryptic peptides of DHP-Pep M5 was facilitated by the sequence analysis of selected peptides obtained by digestion of Pep M5 with chymotrypsin and *S. aureus* V8 protease (See Table 2 for their amino acid compositions), and the previous knowledge of the sequence of region 88-114 from the sequence analysis of the clostripain peptides of Pep M5 (Manjula et al, 1983).

The chymotryptic peptide C2, starting at residue 21, was sequenced through residue 73 (Table 7) and thus provided the overlap for peptides DHP-T5 and DHP-T3 (Arg-64 overlap). The V8 peptide V1 (see Table 2 for amino acid composition) was sequenced in its entirety (Table 9) and was found to correspond to residues 80-95 of Pep M5. Thus, this peptide provided an eight residue overlap for DHP-T3 and DHP-T4 (Arg-87 overlap). The previously reported sequence of the clostripain peptide CP6 (Manjula et al, 1983), which begins at residue 88 and contains an internal arginine residue, provided the overlap for DHP-T4 and DHP-T6 (Arg-99 overlap). The overlap for the peptides DHP-T6 and DHP-T7 was provided by the sequence analysis of the chymotryptic peptide C1A (residues 123-129, Table 8).

REFERENCES

Drapeau, G. R. (1977) Meth. Enzymol. 47, 189-191

Fairwell, T. (1983) Meth. Enzymol. 91, 502-511

Hayashi, R. (1977) Meth. Enzymol. 47, 84-93

Klapper, D. G., Wilde, C. E. and Capra, J. D. (1978) Anal. Biochem. 85, 126-131

Mitchell, W. M. (1977) Meth. Enzymol. 47, 165-170

Nigen, A. M. and Manning, J. M. (1977) Proc. Natl. Acad. Sci. USA 74, 367-371

Zimmerman, C. L., Appella, E. and Pisano, J. J. (1977) Anal. Biochem. 77, 569-573

TABLE 2

AMINO ACID COMPOSITIONS OF THE CHYMOTRYPTIC, V8 PROTEASE- AND SUBTILISIN PEPTIDES OF PEP M5

Amino Acid	C1A	C2	V1	V2	S1
Asp	1.10(1)	9.95(10)	3.00(3)	1.21(1)	3.78(4)
Thr	0.85(1)	7.19(7)	1.60(2)	-	1.33(1)
Ser	-	1.64(2)	0.66(1)	0.69(1)	1.18(1)
Glu	2.32(2)	17.78(18)	4.18(4)	4.34(4)	4.97(5)
Gly	-	3.88(4)	-	1.08(1)	3.83(4)
Ala	-	2.00(2)	-	1.00(1)	2.00(2)
Val	-	-	-	-	1.04(1)
Ile	-	-	-	1.20(1)	2.21(2)
Leu	1.12(1)	9.82(10)	1.99(2)	2.46(2)	3.63(4)
His	-	1.90(2)	-	-	-
Lys	1.14(1)	11.74(12)	3.00(3)	3.26(3)	8.41(8)
Arg	1.00(1)	2.05(2)	0.66(1)	-	-
Total	7	69	16	14	32

Residues in C1A were normalized for Arg=1, those of C2 and S1 were normalized for Ala=2, whereas V1 and V2 were normalized for Lys=3 and Ala=1, respectively.

Table 3

Sequence Data on Peptide DHP-T3

Residues 65-87; 49 nmol

Cycle	Residue	Yield, (nmol)
1	DHP-Lys	NQ
2	Thr	31.06
3	Ala	43.14
4	Glu	39.29
5	Leu	40.47
6	Thr	19.95
7	Ser	7.35
8	Glu	37.78
9	DHP-Lys	NQ
10	DHP-Lys	NQ
11	Glu	25.89
12	His	11.24
13	Glu	20.62
14	Ala	22.22
15	Glu	22.61
16	Asn	12.84
17	Asp	15.44
18	DHP-Lys	NQ
19	Leu	11.58
20	DHP-Lys	NQ
21	Gln	6.82
22	Gln	8.77
23	Arg	1.37

NQ: Not quantitated.

Table 1

AMINO ACID COMPOSITION OF THE TRYPTIC PEPTIDES OF DHP-PEP M5

Amino Acid	DHP-T1	DHP-T2	DHP-T3	DHP-T4	DHP-T5	DHP-T6	DHP-T7	Peptide Total	Sequence
Asp	-	1.04(1)	2.14(2)	1.10(1)	8.10(8)	6.00(6)	8.81(9)	27	9
Asn									17
Thr*	1.79(2)	0.85(1)	1.90(2)	2.73(3)	4.82(5)	3.75(4)	4.21(4)	21	21
Ser	-	0.81(1)	1.01(1)	0.92(1)	0.98(1)	-	1.96(2)	6	6
Glu	-	1.07(1)	7.07(7)	3.14(3)	12.29(12)	5.55(5)	19.08(19)	47	34
Gln									13
Pro	-	0.86(1)	-	-	-	-	-	1	1
Gly	-	0.94(1)	-	-	4.08(4)	1.21(1)	1.94(2)	8	8
Ala	-	-	2.0(2)	-	2.0(2)	-	4.81(5)	9	9
Val	1.00(1)	-	-	-	-	0.84(1)	2.00(2)	4	4
Ile	-	0.91(1)	-	-	-	0.88(1)	4.07(4)	6	6
Leu	-	-	2.01(2)	2.00(2)	8.22(8)	3.41(3)	8.95(9)	24	24
Tyr	-	-	-	-	1.07(1)	1.00(1)	-	2	2
His	-	-	1.09(1)	-	1.11(1)	-	-	2	2
Lys	-	-	-	-	-	-	-	-	35
Arg	1.17(1)	1.00(1)	0.85(1)	0.81(1)	1.0(1)	0.85(1)	-	6	6
Lys*	-	-	5.34(5)	1	9.99(10)	4.38(4)	16.17(16)	36	
Total	4	8	23	12	53	27	72		197

Residues in Peptide DHP-T1 are normalized for val=1, and those of DHP-T2 and DHP-T5 are normalized for arg=1. The residues in DHP-T3, DHP-T4, DHP-T6 and DHP-T7 are normalized for ala=2, leu=2, asp=6, and val=2, respectively.

*Amino acid analysis of DHP-T1 revealed a peak eluting close to histidine and was quantitated using the factor for alanine. since by amino acid composition, DHP-T1 corresponds to residues 1-4 of Pep M5, this new peak was assumed to be α-DHP-Thr. Therefore, the value shown for Thr in DHP-T1 represents one residue each of free Thr and the assumed α-DHP-Thr.
Amino acid analysis of peptides DHP-T3 through DHP-T7 revealed, in addition to the usual amino acids, two other peaks - one major and one minor - eluting just ahead of histidine. The position of the major peak corresponded to that of a standard of ϵ-mono-DHP-lysine (Nigen and Manning, 1977), run under the same conditions. The minor peak (less than 20% of the major ϵ-mono-DHP-lysine peak) is presumably ϵ-di-DHP-lysine. The values shown in the table represent the sum of these two peaks. These peaks were quantitated using the factor for alanine and hence their values are only approximate estimates of the modified-lysine content of the peptides. These values, however, were in fair agreement with the number of lysines determined by sequencing.

Table 4

Sequence data for peptide DHP-T5

Residues 13-64, 69 nmol.

Cycle	Residue	Yield, (nmol)	Cycle	Residue	Yield, (nmol)
1	Ala	53.72	26	Asn	3.38
2	DHP-Lys	NQ	27	Glu	10.63
3	Glu	46.66	28	Gly	3.39
4	Ala	42.88	29	Leu	9.49
5	Leu	39.66	30	DHP-Lys	NQ
6	Asp	21.31	31	Thr	2.77
7	DHP-Lys	NQ	32	Glu	7.31
8	Tyr	19.94	33	Asn	2.23
9	Glu	31.82	34	Glu	6.16
10	Leu	28.80	35	Gly	2.06
11	Glu	28.30	36	Leu	6.08
12	Asn	11.40	37	DHP-Lys	NQ
13	His	NQ	38	Thr	1.84
14	Asp	10.02	39	Glu	4.19
15	Leu	21.14	40	Asn	1.43
16	DHP-Lys	NQ	41	Glu	3.63
17	Thr	6.67	42	Gly	1.31
18	DHP-Lys	NQ	43	Leu	4.46
19	Asn	6.01	44	DHP-Lys	NQ
20	Glu	16.63	45	Thr	1.04
21	Gly	5.16	46	Glu	2.48
22	Leu	13.96	47	DHP-Lys	NQ
23	DHP-Lys	NQ	48	Ser	0.93
24	Thr	4.33	49	Asn	0.88
25	Glu	12.34	50	Leu	2.64

NQ: Not quantitated

The first 20 residues were also identified by chemical ionization mass spectrometry.

Table 5

Sequence Data On Peptide DHP-T6

Residues 100-126; 56 nmol.

Cycle	Residue	Yield, (nmol)
1	Glu	46.60
2	Val	35.93
3	Gln	29.31
4	Asn	22.87
5	Thr	4.45
6	Gln	18.34
7	Tyr	18.45
8	Asn	13.49
9	Asn	13.70
10	Glu	13.07
11	Thr	3.07
12	Leu	9.70
13	DHP-Lys	NQ
14	Ile	5.49
15	DHP-Lys	NQ
16	Asn	2.19
17	Gly	0.88
18	Asp	1.53
19	Leu	1.52
20	Thr	0.34
21	DHP-Lys	NQ
22	Glu	0.84
23	Leu	0.48
24	Asn	0.76
25	DHP-Lys	NQ
26	Thr	0.14

NQ: Not quantitated.

Table 6

Sequence data on peptide DHP-T7

Residues 127-197; 49 nmol.

Cycle	Residue	Yield (nmol)	Cycle	Residue	Yield, (nmol)
1	Gln	38.71	26	DHP-Lys	NQ
2	Glu	40.18	27	Asp	5.92
3	Leu	34.15	28	DHP-Lys	NQ
4	Ala	29.23	29	Ile	5.89
5	Asn	23.97	30	Ala	3.65
6	DHP-Lys	NQ	31	DHP-Lys	NQ
7	Gln	27.92	32	Glu	4.94
8	Gln	26.10	33	Gln	2.47
9	Glu	24.27	34	Glu	4.46
10	Ser	9.95	35	Asn	1.88
11	DHP-Lys	NQ	36	DHP-Lys	NQ
12	Glu	18.6	37	Glu	3.50
13	Asn	11.75	38	Thr	1.22
14	Glu	16.44	39	Ile	2.79
15	DHP-Lys	NQ	40	Gly	1.88
16	Ala	13.89	41	Thr	0.90
17	Leu	14.27	42	Leu	2.82
18	Asn	7.75	43	DHP-Lys	NQ
19	Glu	10.54	44	DHP-Lys	NQ
20	Leu	11.73	45	Ile	2.06
21	Leu	9.91	46	Leu	1.93
22	Glu	8.23	47	Asp	1.00
23	DHP-Lys	NQ	48	Glu	1.24
24	Thr	3.77	49	Thr	0.76
25	Val	7.48	50	Val	1.07

NQ: Not quantitated.

The first 20 residues were also identified by chemical ionization mass spectrometry.

Table 7

Sequence data on peptide C2

Residues 21-90; 30 nmol.

Cycle	Residue	Yield, (nmol)	Cycle	Residue	Yield, (nmol)
1	Glu	23.37	28	Leu	5.30
2	Leu	22.58	29	Lys	4.59
3	Glu	21.09	30	Thr	1.15
4	Asn	13.87	31	Glu	3.31
5	His	NQ	32	Asn	1.97
6	Asp	12.52	33	Glu	3.11
7	Leu	17.91	34	GLY	2.06
8	Lys	15.89	35	Leu	2.96
9	Thr	6.88	36	Lys	3.21
10	Lys	14.03	37	Thr	0.64
11	Asn	9.72	38	Glu	1.84
12	Glu	13.28	39	Lys	2.54
13	Gly	8.75	40	Ser	0.92
14	Leu	12.55	41	Asn	1.38
15	Lys	10.87	42	Leu	2.09
16	Thr	3.56	43	Glu	1.73
17	Glu	9.75	44	Arg	NQ
18	Asn	6.31	45	Lys	2.02
19	Glu	10.28	46	Thr	0.49
20	Gly	6.13	47	Ala	1.40
21	Leu	8.78	48	Glu	1.34
22	Lys	7.07	49	Leu	1.17
23	Thr	1.98	50	Thr	0.40
24	Glu	6.07	51	Ser	0.63
25	Asn	3.81	52	Glu	1.09
26	Glu	5.59	53	Lys	1.13
27	Gly	3.98			

NQ: not quantitated.

The first 20 residues were also identified by chemical ionization mass spectrometry.

Table 8

Sequence data on Peptide C1A

Residues 123-129; 60 nmol.

cycle	Residue	Yield, (nmol)
1	Asn	32.69
2	Lys	14.80
3	Thr	14.21
4	Arg	16.00
5	Gln	24.58
6	Glu	25.00
7	Leu	0.44

Table 9

Sequence data for peptide V1

Residues 80-95; 34 nmol.

Cycle	Residue	Yield, (nmol)
1	Asn	33.88
2	Asp	30.20
3	Lys	20.94
4	Leu	21.80
5	Lys	14.24
6	Gln	16.14
7	Gln	14.11
8	Arg	9.07
9	Asp	9.38
10	Thr	1.90
11	Leu	6.13
12	Ser	1.30
13	Thr	1.00
14	Gln	4.26
15	Lys	4.51
16	Glu	3.15

Table 10

Sequence Data On Peptide V2

Residues 184-197; 22 nmol.

Cycle	Residue	Yield, (nmol)
1	Gln	14.38
2	Lys	15.66
3	Ser	3.39
4	Lys	8.10
5	Gln	5.24
6	Asn	4.38
7	Ile	2.91
8	Gly	0.83
9	Ala	1.65
10	Leu	0.90
11	Lys	0.32
12	Gln	0.15
13	Glu	0.18
14	Leu	0.08

Table 11

Sequence Data On Peptide S1

Residues 169-197; 45 nmol.

Cycle	Residue	Yield, (nmol)
1	Lys	23.30
2	Lys	26.10
3	Ile	32.60
4	Leu	23.30
5	Asp	37.10
6	Glu	36.10
7	Thr	14.30
8	Val	33.90
9	Lys	26.10
10	Asp	25.10
11	Lys	10.70
12	Leu	11.60
13	Ala	11.20
14	Lys	8.80
15	Glu	7.90
16	Gln	4.90
17	Lys	1.90
18	Ser	0.32
19	Lys	0.88
20	Gln	0.46
21	Asn	0.07

The complete amino acid sequence of a biologically active 197-residue fragment of M protein isolated from type 5 group A streptococci.

B N Manjula, A S Acharya, S M Mische, T Fairwell and V A Fischetti

J. Biol. Chem. 1984, 259:3686-3693.