Volume 208, number 2 FEBS 4220 November 1986

Type 1 M protein of *Streptococcus pyogenes*

N-terminal sequence and peptic fragments

Ladislav Morávek, Otto Kühnemund+, Jiří Havlíček*, Petr Kopecký and Manfred Pavlík

Institute of Organic Chemistry and Biochemistry, Czechoslovak Academy of Sciences, 166 10 Prague 6, Czechoslovakia, +*Central Institute of Microbiology and Experimental Therapy, Academy of Sciences of the GDR, Jena, GDR* and **Institute of Hygiene and Epidemiology, Šrobárova 48, 100 42 Prague 10, Czechoslovakia*

Received 23 September 1986

Limited proteolysis of the surface of type 1 *Streptococcus pyogenes* by pepsin gives rise to fragment Pep M1 of M_r 20270 as the main product which covers the N-terminal part of the M protein. The amino acid sequence was determined of the N-terminal region of the M protein representing the most exposed part of the molecule on the surface fibrils of streptococcal cells, which seems to be very important for the differentiation of the individual serological types. The sequence differs from the homologous N-terminal sequences of types 5, 6 and 24, and shows a homology with sequences repeating in the chain of type 24. Fragment Pep M1 binds to fibrinogen; the absence of its 30 N-terminal amino acid residues, however, abolishes this interaction which is believed to play a role in the virulence of *S. pyogenes*.

(Streptococcus pyogenes) *M protein* *Amino acid sequence* *Fibrinogen binding* *Synthetic antigen*

1. INTRODUCTION

The M protein which is present in fibrils on the surface of streptococcal cells is the only known factor of the virulence of *Streptococcus pyogenes* [1]. This property of the M protein is neutralized by the production of protective opsonic antibodies against the about 80 known serological types of the M protein of *S. pyogenes*. The single chain molecule is present in the fibrils as a dimer in the form of an α-helical coiled-coil structure [2]. The antiopsonic mechanism of the M proteins has not yet been elucidated; some authors [3] believe that the binding of fibrinogen, a non-type-specific ability of the M protein, plays a role in this mechanism. The isolated M protein was found to be an unsuitable antigen for vaccination against streptococcal infections since the immune cross reactions of some of its types with heart tissues may cause acute rheumatic fever [4]. A solution seems to be offered by the preparation of a synthetic peptide vaccine containing the antigenic determinants of the M protein which would bring about the production of protective antibodies and yet be free of the segments responsible for the formation of cross reacting antibodies and toxic effects [5].

The determination of the general regularities of the molecular architecture of various M protein types, the investigation of the fibrinogen binding region obviously existing in all types and especially the practical aspect of the problem, i.e. the preparation of synthetic vaccines, require information on amino acid sequences of the individual M protein types. At present, sequence data are available for the N-terminal regions of types 5, 6 [6] and 24 [7]; the nucleotide sequence for type 6, permitting the primary structure of its whole polypeptide chain to be derived, was reported recently [8]. We decided to start our comparative studies with type 1.

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/86/$3.50

2. MATERIALS AND METHODS

The cultivation of type 1 *S. pyogenes* strain CNCTC 40/58 and its extraction with limited proteolysis by pepsin as well as the fractionation of the crude extract by affinity chromatography on immobilized fibrinogen, completed by gel chromatography and the identification of the M protein in the fractionated material, are described elsewhere (Kühnemund, O. et al., in preparation). The amino acid analyses were effected by conventional procedures in a Durrum model D-500 amino acid analyzer. The sequence studies were performed by automatic degradation of approx. 5 nmol of sample in an Applied Biosystems (Foster City, CA, USA) model 470 A protein sequencer following a standard program (03RPTH) supplied, as well as the reagents, by the instrument's manufacturer. The average repetitive yield of the degradation was 94.5%. The amino acid phenylthiohydantoins were identified by HPLC on an Ultrasphere ODS column eluted by an acetonitrile gradient in acetate buffer (0.03 M, pH 4.2) using a Beckman liquid chromatography system and a Shimadzu integrator.

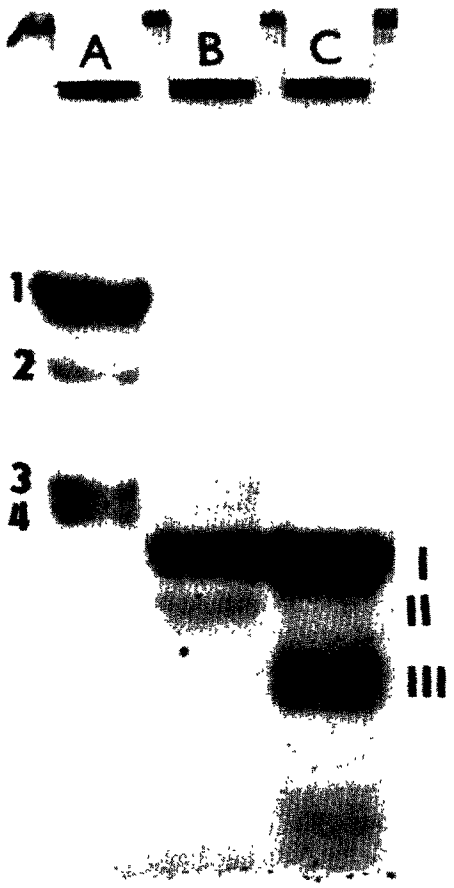

Fig.1. SDS-polyacrylamide gel electrophoresis of the pepsin extract. Lanes: A, M_r standards ($M_r \times 10^{-3}$), band 1 (69), band 2 (45), band 3 (28), band 4 (23); B, crude extract; C, fibrinogen-binding material after affinity chromatography. I–III, bands of peptic fragments.

3. RESULTS AND DISCUSSION

3.1. *Characterization of peptic fragments of M protein*

The result of the basic fractionation of the crude pepsin extract is shown in fig.1. Three large peptic fragments were found in bands I–III, of which component II does not bind to fibrinogen whereas components I and III (the last one, barely visible in the crude extract) are retarded during affinity chromatography.

Component I is the main product of limited peptic hydrolysis and will be referred to as fragment Pep M1. It contains a total of 177 amino acid residues (table 1); its M_r calculated from the amino acid composition is 20270. The N-terminal sequence of Pep M1 is

10
Val-Ala-Gly-Arg-Asp-Phe-Lys-Arg-Ala-Glu-Glu-
20
Leu-Glu-Lys-Ala-Lys-Gln-Ala-Leu-Glu-Asp-
30
Gln-(Arg)-Lys-Asp-Leu-Glu-Thr-Lys-Leu-
Lys-Glu-Leu-Gln-Gln-Asp-Tyr-X-Leu;

the residues in positions 23 and 38 have not been identified unambiguously.

The N-terminal sequence of component II is

Lys-Glu-Leu-Gln-Gln-Asp-Tyr-Asp-Leu-Ala-Lys-
Glu-Ser-Thr-Ser-(Arg)-Lys-(Pro)-Gln-(Gly)-
Ser-Glu-Lys-Lys-Leu;

the residues in brackets have not been identified unambiguously. The sequence of the first 9

Table 1

Amino acid composition of fragments

Amino acid	mol/mol			
	I	Found in sequence of res. 1–30	I – (res. 1–30)	II
Asp	19.0(19)	3	16	15.9
Thr	8.1(8)	1	7	6.7
Ser	13.0(13)	–	13	10.5
Glu	40.8(41)	7	34	35.1
Pro	0.7(1)	–	1	0.7
Gly	4.6(5)	1	4	3.7
Ala	17.9(18)	4	14	13.7
Cys	–	–	–	–
Val	3.1(3)	1	2	2.0
Met	0.9(1)	–	1	0.8
Ile	6.3(6)	–	6	6.3
Leu	24.4(24)	4	20	20.0
Tyr	4.1(4)	–	4	3.3
Phe	1.7(2)	1	1	0.6
His	1.9(2)	–	2	1.7
Lys	19.0(19)	5	14	14.5
Arg	11.0(11)	3	8	8.1
Trp	–	–	–	–
Total	177	30	147	

residues can be aligned with positions 31–39 of fragment Pep M1, the region which follows extending the partial sequence of the type 1 M protein up to residue 55. The formation of component II by cleavage of the bond Leu(30)-Lys(31) of the M protein chain is in agreement with the known specificity of pepsin. The sequence provides evidence that component II, even though it does not bind to fibrinogen, is a part of the M protein corresponding to Pep M1 minus 30 residues at its N-terminus (cf. table 1).

Component III is a minor product of extraction; the possibility cannot be excluded that it arises from secondary cleavage of Pep M1 and hence no conclusions on its interaction with fibrinogen can be made from the course of the affinity chromatography. The mixture of component III with Pep M1 yields a homogeneous sequence, therefore it probably covers the same region of the M protein but is shorter at the C-terminus. The C-terminal parts of the M protein chain are not released into the peptic extract.

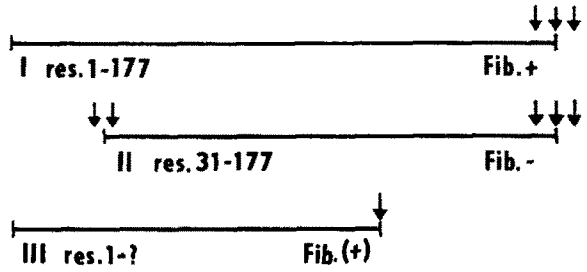

Fig.2. Peptic fragments and their binding to fibrinogen. The number of arrows indicates the preference of peptic cleavage. The binding of fragment III has not been determined beyond doubt (cf. text).

The mutual arrangement of the fragments is shown in fig.2. Fibrinogen interacts with fragment Pep M1 representing on the fibrils the most exposed N-terminal part of the M protein; fragment II, which is shorter by 30 residues at its N-terminus, lacks this binding ability.

3.2. *Homology of N-terminal sequences of M protein types*

A comparison of the N-terminal sequences [6] of types 5, 6 and 24 so far determined reveals a marked homology; an alignment of the sequence of type 1 with these data is shown in fig.3. As is obvious from the top four lines, the homology of type 1 (supporting the assignment of fragment Pep M1 to the N-terminus of the M protein) is restricted to the first four positions only; the arginine residue is conserved in all cases. In contrast, a comparison with the closely related sequences of the internal part of the chain of type 24 [9] indicates a marked relationship with the N-terminus of type 1. A similar relationship with the internal region of type 24 is suggested by our preliminary data (unpublished) on the N-terminus of type 12 M protein.

The binding of the fibrinogen molecule to the extended N-terminal region of the M protein represents a certain restriction to the accessibility of the internal part of its chain. This fact increases the importance of the short exposed N-terminal region of the M protein as a binding site for opsonic antibodies.

```
Type  5   TVTRGTISDPQRAKEALDKYELENHDLKTKNEGLKTENEGLKTENEGLKTENEGL
Type  6   RVFPRGTVENPDKARELLNKYDVENSMLQANNDKLTTENNNLTDQNKNLTTENKNL
Type 24    VATRSQTDTSEKVQERADSFEIENNTLKL
Type  1    VAGRDFKRAEELEKAKQALEDQrKDLETKLKELQQDYDLAKESTSrKpQgSEKKL
     CB3   NFSTADSAKIKTLEAEKAALEAEKADLEKALEGAM
     CB4   NFSTADSAKIKTLEAEKAALEARQADLEKALEGAM
     CB5   NFSTADSAKIKTLEAEKAALEARKADLDDALELEGAM
     CB6   NFSTADSAKIKTLEAEKAALEARQAELEKALEGAM
     CB7   NFSTADSAKIKTLEAEKAALAARKADLEKALEGAM
```

Fig.3. Alignment of N-terminal sequences of types 5, 6 [6], 24 [7] and 1. The positions identical with type 1 are given in bold-face letters; CB3–CB7 are sequences of cyanogen bromide fragments from the internal part of the chain of type 24 [9]. Lower-case letters indicate positions not identified unambiguously.

ACKNOWLEDGEMENT

The authors are grateful to Mr J. Zbrožek for the amino acid analyses.

REFERENCES

[1] Fox, E.N. (1974) Bacteriol. Rev. 38, 57–86.
[2] Phillips, G.N. jr, Flicker, P.F., Cohen, C., Manjula, B.N. and Fischetti, V.A. (1981) Proc. Natl. Acad. Sci. USA 78, 4689–4693.
[3] Whitnack, E. and Beachey, E.H. (1982) J. Clin. Invest. 69, 1042–1047.
[4] Dale, J.B. and Beachey, E.H. (1985) J. Exp. Med. 161, 113–122.
[5] Beachey, E.H. and Stollerman, G.H. (1971) J. Exp. Med. 134, 351–362.
[6] Manjula, B.N., Acharya, A.S., Mische, S.M., Fairwell, T. and Fischetti, V.A. (1984) J. Biol. Chem. 259, 3686–3693.
[7] Beachey, E.H., Seyer, J.M. and Kang, A.H. (1978) Proc. Natl. Acad. Sci. USA 75, 3163–3167.
[8] Hollingshead, S.K., Fischetti, V.A. and Scott, J.R. (1986) J. Biol. Chem. 261, 1677–1686.
[9] Beachey, E.H., Seyer, J.M., Dale, J.B. and Hasty, D.L. (1983) J. Biol. Chem. 258, 13250–13257.