Proc. Natl. Acad. Sci. USA
Vol. 76, No. 8, pp. 3765–3768, August 1979
Biochemistry

# Amino acid sequence and physicochemical similarities between streptococcal M protein and mammalian tropomyosin

(staphylococcal protein A/actin/myosin/conformation/rheumatic fever)

BARBARA HOSEIN, MACLYN MCCARTY, AND VINCENT A. FISCHETTI*

The Rockefeller University, New York, New York 10021

*Contributed by Maclyn McCarty, May 16, 1979*

**ABSTRACT** The amino-terminal sequences of two peptides of type 24 streptococcal M protein show similarities with that of rabbit skeletal muscle tropomyosin, having up to 40% identical residues and probabilities of occurring by chance as low as $P < 10^{-5}$. In addition, a hexapeptide (Glu-Ala-Glu-Lys-Ala-Ala) that is found five times in the M24 protein was shown to be identical to a sequence in tropomyosin. Similarities are also seen in the amino acid compositions and physicochemical properties of the two proteins. The amino-terminal sequences of peptides from another bacterial surface protein, staphylococcal protein A, are highly correlated with segments of two other myofibrillar proteins, rabbit actin ($P < 10^{-7}$) and rabbit myosin A1 light chain ($P < 10^{-6}$). The data presented suggest that a close structural relationship exists between mammalian muscle proteins and the biologically active surface proteins of staphylococci and streptococci. In addition, the correlation between sequences in M protein and tropomyosin represents direct evidence of a structural similarity at a molecular level between a streptococcal protein and a mammalian muscle component and may therefore prove relevant to the pathogenicity of the streptococcus.

Group A streptococci possess antiphagocytic surface antigens, known as M proteins, that are primarily responsible for the virulence of these organisms (1). Based on extensive immunological data, it has been clearly shown that resistance to streptococcal infection is dependent on neutralization of the antiphagocytic effect of M proteins by type-specific opsonic antibodies (1, 2). Immunological cross reactions do occur between certain of the more than 70 immunologically distinct M types (3, 4), but the antibodies involved rarely afford crossprotection (3). Recent peptide analyses (5) and sequence studies (6, 7) on a limited number of M protein types revealed that, in spite of a common biological activity, M proteins are composed of different primary structures. Despite these immunological and structural differences, all M protein molecules studied to date share certain noteworthy chemical and physical properties (8). Their extreme thermal stability (9), their highly elongated shape (10, 11), and their appearance as a network of projections on streptococcal cell walls (12) suggested that M proteins might have properties in common with previously characterized fibrous proteins.

This report describes striking chemical, physical, and sequence similarities between streptococcal M protein and rabbit skeletal muscle tropomyosin. In addition, computer analysis of the partial sequence of staphylococcal protein A, another biologically active, Gram-positive, surface protein, having properties in common with M protein, reveals significant homology with two other myofibrillar proteins, actin and myosin A1 light chain.

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

## RESULTS AND DISCUSSION

M proteins characteristically contain unusually high levels of lysine, aspartate or asparagine, glutamate or glutamine, leucine, and alanine but no cysteine and only negligible amounts of glycine, proline, and aromatic amino acids (7, 13–16). Such a distribution of amino acids is typical of $\alpha$ fibrous proteins in general (17); of these, skeletal muscle tropomyosin bears the greatest resemblance in composition to M proteins.

Tropomyosin is a component of the thin filament in muscle cells that binds actin and troponin during calcium-dependent muscle contraction (18). It is nearly 100% $\alpha$-helical and exists in a two-chain, coiled-coil conformation (19). The physical properties of rabbit skeletal muscle tropomyosin correspond closely with those of M proteins (Table 1). Both proteins exhibit size, shape, and charge similarities. The thermal and pH stability of tropomyosin are general features of myofibrillar proteins that are usually attributed to their high $\alpha$-helical content (24). Whether this same characteristic explains the unusual stability of M protein under hot acidic conditions remains to be seen.

In view of the physicochemical similarities between M proteins and tropomyosin, the primary structures of these proteins were compared. Published partial sequence analysis of one M protein, type 24, has shown that it is a molecule with internally repeating sequences (7). Cyanogen bromide cleavage of this molecule yielded seven peptides (7). The partial sequences of two of the peptides (CB-1 and CB-2) were found to be identical to each other and to the amino-terminal region of the uncleaved M24 molecule through at least the first 27 residues (7). The amino-terminal sequences of the remaining five peptides (CB-3 to CB-7) were identical to each other through the 20th residue but completely different from the amino-terminal region of the other two peptides (7). The sequence of rabbit skeletal muscle $\alpha$ tropomyosin (25) is repetitive throughout its entire length, although the repeating units are not identical to one another as in M-24 protein. There is a nearly regular 42-residue pattern that repeats seven times within the tropomyosin molecule (26), upon which is superimposed a seven amino acid periodicity with a regular pattern of nonpolar amino acids in positions 2 and 5 of every heptapeptide (25).

In comparing the sequence of the M-24 peptides (7) with that of rabbit skeletal tropomyosin (25), it was noted that M-24 peptides CB-3 through CB-7 all contain a hexapeptide, Glu-Ala-Glu-Lys-Ala-Ala (residues 14–19), which is identical to segment 115–120 of tropomyosin. Although this hexapeptide occurs five times in the M-24 molecule, it is found in no other protein sequence listed in the *Protein Segment Dictionary 76* (27). Alignment of the CB-5 peptide with tropomyosin showed additional identities clustered around the hexapeptide in positions 8, 11, 21, and 23 (Fig. 1 *upper*), bringing the overall

* To whom reprint requests should be addressed.

Table 1. Physicochemical properties of M protein and tropomyosin

| Property | M protein | Tropomyosin |
|---|---|---|
| Molecular weight | 30,000–35,000 (6, 13) | 35,000 monomer (20)<br>70,000 dimer |
| Isoelectric point | 5.3 (21) | 5.1 (22) |
| Axial ratio | 15:1 to 25:1 (10, 11) | 24.5:1 (20) |
| Solubility | Soluble in aqueous solutions (23) | Soluble in water (22) except at isoelectric point |
| Heat stability | Stable at 100°C (9) | Stable at 100°C (22) |
| Stability in acid | Stable at pH 2 (9) | Stable at pH 2 (24) |
| Stability in base | Stable at pH 10 (10) | Stable at pH 9.5 (24) |

identity of CB-5 to segment 102–126 of tropomyosin to 40%. CB-5 was also observed to contain an internally repeated segment, Leu-Glu-Ala-Glu-Lys-Ala (residues 13–18 and 20–25) which occurs with an interval of seven amino acids between the first residues of the repeated segments. Because tropomyosin contains a seven amino acid periodicity throughout the molecule, this similarity further strengthens the relationship between M protein and this mammalian muscle protein. Fig. 1 *lower* compares the sequence of M-24 peptide CB-1, which occurs twice in the M-24 molecule, with three other segments of tropomyosin. Segment 202–228 exhibits 26% identity with CB-1 and, in addition, contains one example of conserved aromatic amino acids (residue 20) and two of conserved basic residues (residues 4 and 16). Furthermore, tropomyosin segments 122–148 and 73–99 both exhibit 19% identity with the CB-1 segment. However, due to the internal repetitive nature of tropomyosin, it is difficult to assess the full significance of the homology of several segments of this molecule with a segment of M protein.

In addition to direct alignment, the correlation between M protein peptides and tropomyosin was also tested by computer analysis using the program SEARCH (28) (carried out by M. O. Dayhoff, W. C. Barker, and L. T. Hunt). Of the 100,000 segments analyzed, the M-24 peptide CB-1 retrieved segment 202–228 of tropomyosin with a probability of $P < 10^{-5}$, a correlation that is unlikely to have occurred by chance (29). The CB-5 segment received a less significant score ($P < 10^{-3}$) for its correlation with segment 102–126 of tropomyosin because 8 of the 10 identities in this segment are Ala or Glu, identities that are statistically less significant than more highly conserved amino acids (28). However, if one judges the degree of homology on the frequency of identical residues, segment CB-5 could certainly be considered homologous with segment 102–126 of tropomyosin. Proteins that are more than 35% identical are usually considered to be obviously homologous (30). In fact, functionally homologous proteins, such as cytochromes *c* or globins may have as few as 15–30% identical residues at corresponding positions, as compared to about 5% expected in comparisons between random sequences (31). The finding of similarities between M peptides and tropomyosin that are 40% and 26% identical would indicate that the peptides exhibit strong homology to the tropomyosin segments. But one must bear in mind that these comparisons are between peptides and not whole proteins. However, because the known sequence of M24 accounts for nearly 50% of the molecule, the observed homology remains significant.

Staphylococcal protein A is another bacterial surface protein that resembles M protein in antiphagocytic activity (32), amino acid composition (33), extended shape (34), and internally repetitive sequences (33, 35). Protein A is composed of four highly homologous regions, each consisting of 58 amino acid residues, which are capable of binding the Fc region of IgG (35). These regions (D, A, B, and C) are arranged consecutively from the amino-terminal part of the protein; and it has been suggested that the residual carboxyl-terminal end of approximately 150 residues binds to the bacterial cell wall structure.

Based on computer analysis, we have found that protein A shows high degrees of homology with rabbit actin (36) and myosin (A1 light chain) (37) (Fig. 2). The probabilities of the homologies for protein A-fragment A and actin ($P < 10^{-6}$) having occurred by chance are even less than those observed for M protein and tropomyosin. The exceedingly high correlation between protein A and actin is primarily due to the conservation of the aromatic residues phenylalanine and tyrosine in positions 13 and 14, because only 20% of the residues are identical (Fig. 2). The correlation with myosin light chain,

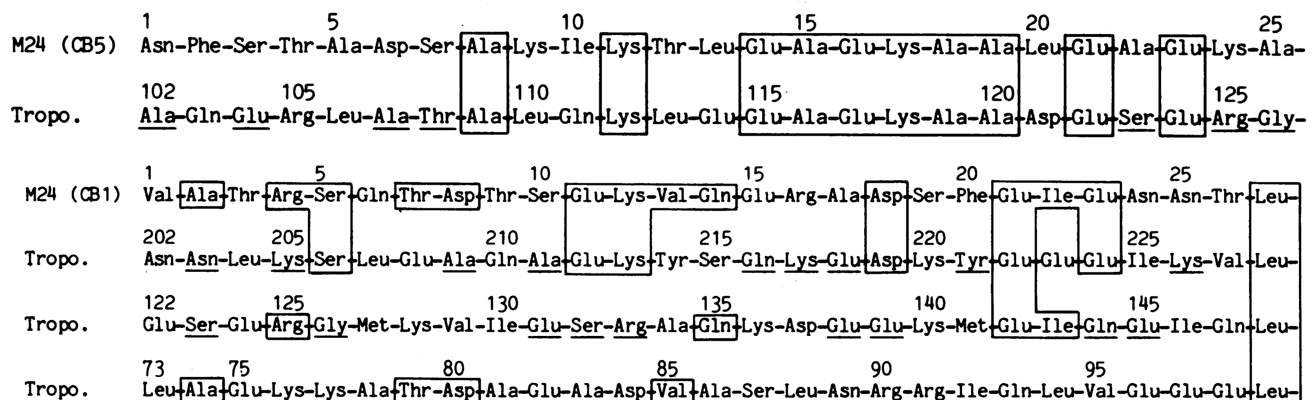


FIG. 1. (*Upper*) Sequence comparison between the amino-terminal region of M24 peptide CB-5 (7) and a segment of rabbit skeletal tropomyosin (Tropo.) (25). Identical residues are boxed. Underlined residues signify conservative substitutions (28). (*Lower*) Sequence comparison between the amino-terminal region of M24 peptide CB-1 (7) and three regions of rabbit skeletal tropomyosin (Tropo.) (25). Those residues of tropomyosin that are identical to CB-1 are boxed. Underlined residues signify conservative substitutions (28).

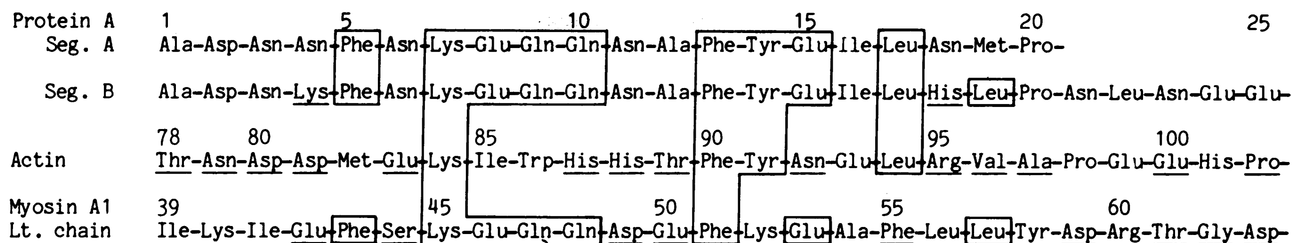


FIG. 2. Sequence comparison between the amino-terminal region of staphylococcal protein A fragments A and B (33) and segments of rabbit skeletal muscle actin (36) and myosin (A1 lt. chain) (37). Identical residues are boxed. Underlined residues signify conservative substitutions (28).

however, features the conservation of two phenylalanine residues, a tetrapeptide identity (Lys-Glu-Gln-Gln), and 32% of identical residues (Fig. 2). Interestingly, of the six residues in the amino-terminal segment of protein A-fragment B that interact with the Fc region of immunoglobulins (residues 5, 6, 9, 13, 17, and 18) (38), three (Phe-5, Gln-9, and Phe-13) are identical in the corresponding positions of myosin light chain and two (Asn-6 and Leu-17) are conservative substitutions. The subsequent residues (residues 26–58) of the protein A-fragment B molecule (39) show little similarity to either actin or myosin (data not shown). Therefore, it may be concluded either that only part of a gene may have been involved in a common evolutionary origin, as has been suggested for H5 histones and murine leukemia virus phosphoprotein P12 (40), or that local conformational requirements led to a convergence of sequences.

In addition to the similarities noted between protein A and actin and myosin A1 light chain, *Escherichia coli* L7/L12 ribosomal proteins (41) have been shown to have sequence similarities with myosin (A1 light chain). Also, the repetitive sequence and conformational properties of *E. coli* lipoprotein (42) have led to a suggestion that it may be similar to tropomyosin. Therefore, the relationship between M protein peptides and tropomyosin appears to be only one example of a general correspondence between bacterial structural proteins and mammalian myofibrillar proteins.

The finding of sequence similarities between M protein peptides and tropomyosin may be of biological significance in view of the association of the group A streptococcus with rheumatic heart disease (43). However, owing to the multiplicity of streptococcal factors implicated in rheumatic fever (43), no simplistic model of pathogenesis can be devised based on this current observation. But nonetheless, this report provides direct evidence of structural similarity, at a molecular level, between a streptococcal protein and a mammalian muscle component. Whether this similarity proves to be the basis for some of the observed crossreactions between streptococcal proteins and heart or skeletal muscle tissues (44, 45) must await detailed immunochemical analysis.

How the structure of M protein enables it to impede phagocytosis is still not apparent, but at least one observation suggests that this may be a property also common to some fibrous proteins. For instance, amyloid fibers have been shown to resist phagocytosis by human leukocytes unless opsonized by specific antisera (46), just as the antiphagocytic effect of M proteins can be overcome only by type-specific opsonic antisera (2). If the fibrous conformation of M proteins determines their antiphagocytic activity, then this may explain how such an immunologically heterogeneous group of molecules can perform the same biological function.

We thank Drs. M. O. Dayhoff, W. C. Barker, and L. T. Hunt for their computer analyses. These studies were made possible by a grant from the American Heart Association (78–1099) and funds from the Institute Merieux. Support also came from U.S. Public Health Service Grants AI 11822 and HL 03919. V.A.F. is a recipient of a U.S. Public Health Service Research Career Development Award.

1. Lancefield, R. C. (1959) *J. Exp. Med.* **110,** 271–292.
2. Lancefield, R. C. (1962) *J. Immunol.* **89,** 307–313.
3. Wiley, G. G. & Bruno, P. N. (1968) *J. Exp. Med.* **128,** 959–968.
4. Fischetti, V. A. (1977) *J. Exp. Med.* **146,** 1108–1123.
5. Fischetti, V. A. (1978) *J. Exp. Med.* **147,** 1771–1778.
6. Beachey, E. H., Seyer, J. M. & Kang, A. H. (1979) in *Streptococcal Diseases and the Immune Response,* eds. Zabriskie, J. B. & Read, S. E. (Academic, New York), in press.
7. Beachey, E. H., Seyer, J. M. & Kang, A. H. (1978) *Proc. Natl. Acad. Sci. USA* **75,** 3163–3167.
8. Fox, E. N. (1974) *Bacteriol. Rev.* **38,** 57–86.
9. Lancefield, R. C. (1928) *J. Exp. Med.* **47,** 91–103.
10. Fox, E. N. & Wittner, M. K. (1969) *Immunochemistry* **6,** 11–24.
11. Pappenheimer, A. M., Jr., Williams, J. W. & Zittle, C. A. (1942) *J. Immunol.* **43,** 61–63.
12. Swanson, J., Hsu, K. C. & Gotschlich, E. C. (1969) *J. Exp. Med.* **130,** 1063–1091.
13. Fischetti, V. A., Gotschlich, E. C., Siviglia, G. & Zabriskie, J. B. (1976) *J. Exp. Med.* **144,** 32–53.
14. Straus, D. C. & Lange, C. F. (1972) *Infect. Immun.* **5,** 927–932.
15. Fox, E. N. & Wittner, M. K. (1965) *Proc. Natl. Acad. Sci. USA* **54,** 1118–1125.
16. Vosti, K. L., Johnson, R. H. & Dillon, M. F. (1971) *J. Immunol.* **107,** 104–114.
17. Fraser, R. D. B. & MacRae, T. P. (1973) *Conformation in Fibrous Proteins and Related Synthetic Polypeptides* (Academic, New York), pp. 403–551.
18. Mannherz, H. G. & Goody, R. S. (1976) *Annu. Rev. Biochem.* **45,** 427–465.
19. Crick, F. H. C. (1953) *Acta Crystallogr.* **6,** 689–697.
20. Holzer, A. R., Clark, R. & Lowey, S. (1965) *Biochemistry* **4,** 2401–2411.
21. Lancefield, R. C. & Perlman, G. E. (1942) *J. Exp. Med.* **96,** 71–82.
22. Bailey, K. (1948) *Biochem. J.* **43,** 271–287.
23. Zittle, C. A. (1942) *J. Immunol.* **43,** 31–46.
24. Lowey, S. (1965) *J. Biol. Chem.* **240,** 2421–2427.
25. Stone, D., Sodek, J., Johnson, P. & Smillie, L. B. (1975) in *Proc. IX FEBS Meetings,* ed. Biro, E. N. A. (North-Holland/Am. Elsevier, New York), Vol. 31, pp. 125–136.
26. McLachlan, A. D., Stewart, M. & Smillie, L. B. (1975) *J. Mol. Biol.* **98,** 281–291.
27. Dayhoff, M. O., Hunt, L. T., Barker, W. C. & Orcutt, B. C. (1976) *Protein Segment Dictionary 76,* (Natl. Biomed. Res. Found., Washington, DC).
28. Dayhoff, M. O. (1976) *Atlas of Protein Sequence and Structure* (Natl. Biomed. Res. Found., Washington, D.C), Vol. 5, Suppl. 2, pp. 3–8 and 311.

29. Barker, W. C. & Dayhoff, M. O. (1977) *Comp. Biochem. Physiol.* **57B**, 309–315.
30. Schwartz, R. M. & Dayhoff, M. O. (1978) *Origin of Life*, Proceedings of the Second International Society for the Study of the Origin of Life Meeting, the Fifth International Conference on the Origin of Life Meeting (Japan Scientific Societies, Tokyo, Japan), pp. 457–469.
31. Ycas, M. (1976) *J. Mol. Evol.* **7**, 215–244.
32. Dossett, J. H., Kronvall, G., Williams, R. C. & Quie, P. G. (1969) *J. Immunol.* **103**, 1405–1410.
33. Sjodahl, J. (1976) *FEBS Lett.* **67**, 62–67.
34. Bjork, I., Petersson, B.-A. & Sjoquist, J. (1972) *Eur. J. Biochem.* **29**, 579–584.
35. Sjodahl, J. (1977) *Eur. J. Biochem.* **73**, 343–351.
36. Collins, J. H. & Elzinga, M. (1975) *J. Biol. Chem.* **250**, 5915–5920.
37. Frank, G. & Weeds, A. G. (1974) *Eur. J. Biochem.* **44**, 317–334.
38. Diesenhofer, J., Jones, T. A. & Huber, R. (1978) *Hoppe-Seyler's Z. Physiol. Chem.* **359**, 975–985.
39. Sjodahl, J. (1977) *Eur. J. Biochem.* **78**, 471–490.
40. Henderson, L. E., Gilden, R. V. & Oroszlan, S. (1979) *Science* **203**, 1346–1348.
41. Amons, R., van Agthoven, A., Pluijms, W. & Moller, W. (1978) *FEBS Lett.* **86**, 282–284.
42. DiRienzo, J. M., Nakamura, K. & Inouye, M. (1978) *Annu. Rev. Biochem.* **47**, 481–532.
43. McCarty, M. (1972) in *Streptococci and Streptococcal Diseases*, eds. Wannamaker, L. W. & Matsen, J. M. (Academic, New York), pp. 517–526.
44. Kaplan, M. H. (1963) *J. Immunol.* **90**, 595–606.
45. Zabriskie, J. B. & Freimer, E. H. (1966) *J. Exp. Med.* **124**, 661–678.
46. Zucker-Franklin, D. (1970) *J. Ultrastruct. Res.* **32**, 247–257.