The Primary Structure of a Salivary Calcium-binding Proline-rich Phosphoprotein (Protein C), a Possible Precursor of a Related Salivary Protein A*

Raymond Sze Chung Wong and Anders Bennick
From the Department of Biochemistry, University of Toronto, Toronto, Canada M5S 1A8

The complete primary structure of a calcium-binding "proline-rich phosphoprotein" named salivary Protein C was determined from peptides obtained by enzymatic and chemical cleavages of the protein. The protein consists of a single polypeptide chain of 150 residues. It contains the entire primary structure of a previously isolated salivary Protein A in its NH2-terminal 106 residues. The COOH-terminal 44 residues consist mostly of glycine, glutamine, and proline, including a hexa-proline sequence, but no polyproline structure could be detected by CD spectroscopy. There is extensive repetition of sequences in the protein, suggesting gene multiplication and recurrent folding. Comparison of the primary structure of salivary Proteins A and C with known protein sequences indicate that the salivary proteins constitute a new family. A mouse submaxillary protease will cleave salivary Protein C between residues 106 and 107 only, giving rise to salivary Protein A and a 44-residue COOH-terminal peptide. This cleavage and the sequence data suggest that salivary Protein C may be a precursor of salivary Protein A.

Human parotid and submandibular saliva contain a number of so-called "acidic proline-rich proteins." These proteins have similar unusual amino acid compositions, and present evidence indicates that they aid in maintaining the calcium concentration in saliva and in keeping the exposed mineralized tissue of the teeth intact under physiological conditions (1, 2). While these activities are shared by all the acidic proline-rich proteins investigated so far, quantitative differences are observed between individual proteins (3). This could be related to differences in structure. At present it is not clear why several of these proteins are present in saliva. It is possible that they are products of different genes or they could arise by postribosomal modification of a primary gene product.

To solve these problems, a knowledge of the primary structure of the proteins is important. The primary structure of one of the acidic proline-rich proteins, salivary Protein A, has been reported (4). This paper describes the primary structure of a related salivary Protein C and explores the possibility that salivary Protein A may be a postribosomal modification of salivary Protein C.

**DISCUSSION**

The primary structure of salivary Protein C is given in Fig. 1, together with peptides and methods used in sequence determination. As expected, trypsin rapidly cleaves the arginine-glycine bond at positions 106 to 107, and it slowly cleaves the arginine-proline bonds at positions 91 to 92 and 139 to 140; these latter two have identical surrounding sequences. The lysine-proline bond at positions 74 to 75 is slowly hydrolyzed, but the same type of bond at positions 129 to 130 is not cleaved. Surprisingly the only difference in the surrounding sequences is the presence of glutamine at position 72 and proline at position 127. It is possible that the presence of proline in the second position from the COOH-terminal end of the susceptible bond causes steric hindrance of trypsin. The other enzymatic and chemical cleavages of salivary Protein C are the same as occur in salivary Protein A (4). The sequence of the NH2-terminal 106 residues is identical to the primary structure of salivary Protein A. As in salivary Protein A, the biological activities, that is, inhibition of hydroxyapatite formation, binding to hydroxyapatite, and calcium binding, are all located in the NH2-terminal 30 residues (5, 6, 7). Salivary Protein A contains six or seven calcium-binding sites, but salivary Protein C contains only three (3, 8). It seems that the longer polypeptide chain in salivary Protein C causes steric hindrance and inhibition of calcium binding. The positively charged amino acids located between residues 107 and 150 in salivary Protein C may, for example, interact with some of the negatively charged residues which otherwise would be part of a calcium-binding site.

Difference CD failed to demonstrate the presence of polyproline. This is somewhat surprising because of the presence of hexaproline in positions 122 to 127. It is possible that the rest of the polypeptide chain induces conformational restraint in the hexaproline region, preventing the formation of all cis- or all trans-peptide bonds necessary for helix formation. The techniques used to detect polyproline helix may not be sensitive enough and, at present, it cannot be ruled out that salivary Protein C contains a polyproline structure.

Previously, it was noted that there were a number of recurrent sequences in salivary Protein A. This pattern of recurrence is even more extensive in salivary Protein C, as illustrated in Fig. 2, and supports the suggestion that gene multiplication has occurred (4). In order to investigate further the relationship of the primary structure of salivary Protein C with other proteins, a computer search of the protein sequences submitted to the National Biomedical Research

---

* This work was supported by Grant MT4920 (to A.B.) from the Canadian Medical Research Council. 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.

1 Portions of this paper including "Experimental Procedures," Results," Figs. 3 to 6, and Tables I to IX 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. 20014. Request Document No. 79M-2327, cite author(s), and include a check or money order for $2.85 per set of photocopies.
Primary Structure of a Proline-rich Phosphoprotein

The primary structure of salivary Protein C. Peptides used to establish the amino acid sequence of salivary Protein C. Residues identified by automated Edman degradation, →; by manual Edman degradation, ←; and by penicillopeptidase-S1 digestion. ←; regions inferred from amino acid composition, ↔.

This sequence of parvalbumin is located in the C and D helical regions which contains one of the calcium-binding sites (10). The proposed ligands for the octahedral calcium binding have been indicated with the letters X, Y, Z, -X, and -Z. Three of the ligands in parvalbumin are also present in salivary Protein C, and one of the remaining is a glutamic acid which is replaced by aspartic acid in salivary Protein C. Several proteins have been found to have sequences homologous to those of the C-D and E-F regions in parvalbumin (10). These homologous regions include α helices preceding and following the calcium-binding site. To explore the similarity between the salivary protein and parvalbumin, it is, therefore, more meaningful to align the complete sequence of the C-D region in parvalbumin with the salivary protein, rather than with that which only contains the calcium-binding ligands. Such a comparison shows that the sequences corresponding to both the C and D regions in the salivary protein contain proline and, therefore, cannot form the α helices found in parvalbumin.

It is likely that some of the calcium-binding sites of the salivary protein are located in the sequence between residues 19 and 30 since dephosphorylation of phosphoserine decreases the amount of bound calcium (3). Even so, it is clear that the calcium-binding sites in the proteins are quite different since the apparent dissociation constants are approximately $2 \times 10^{-7}$ M for salivary Proteins A and C and it is $10^{-7}$ M for parvalbumin (11).

The sequence of CTX can be aligned with residues 16 to 45 in bovine Factor X as follows:
The glutamic acids which are carboxylated in Factor X and involved in calcium binding have been underlined. Only in the case of two (residues 20 and 25) is a glutamic acid present in the corresponding position in the salivary protein. No γ-carboxyglutamic acid is present in Factor IX. It is not known if they are substrates for carboxylating systems. A lysine, or threonine by serine. Of the remaining residues, 6 involve replacement of aspartic acid with glutamic acid, arginine by lysine, or threonine by serine. Of the remaining residues, 6 involve replacement of 1 hydrophilic residue with another and 2 are replacements of hydrophobic residues.

Recently, the partial sequence of the NH2-terminal end of an acidic proline-rich protein isolated from monkey (Macaca fascicularis) parotid saliva has been reported (12). There is a high degree of similarity and it is likely that the protein is homologous to the human proteins described in this paper.

The proline-rich part of salivary Protein C was also investigated for sequence similarities with other proteins. Both salivary Proteins A and C contain tetraproline and salivary Protein C also contains hexaprole. The only other protein with four or more consecutive prolines is fragment from human immunoglobulin A in the constant region of an α-2 chain which contains pentaprole. Recently, there has also been a report of heptaprole in a peptide isolated from human whole saliva (13). The presence in the salivary proteins of six prolines from the same source were reported (13). The relationship of these proteins to those described in this paper remains to be established. A search for similarities of the repeating sequence PGPPQQGG in the salivary proteins was also undertaken. The sequence PAGPPGEAG from collagen α-1 (I) chain had the greatest similarity. Of the 9 residues, 5 are identical. On the other hand, the absence of glycine in every third position in the salivary proteins makes it highly unlikely that the proline-rich part of the salivary proteins and collagen are in the same family. At the most, these proteins could be very distantly related to each other. Salivary Proteins A and C, therefore, appear to be members of a new family of proteins. If there are any distant relationship to other proteins, they are apparently different for the proline-poor and the proline-rich regions. If such relationship do exist they suggest that the gene for the salivary proline rich proteins may have arisen by combining separate genes for the two regions.

The primary structures of salivary Proteins A and C suggest that salivary Protein A is formed by a postribosomal cleavage of salivary Protein C at residues 106 to 107. A very slow cleavage is observed with thrombin, giving rise to products suggesting cleavage at positions 106 to 107 only.

In contrast to the slow cleavage by thrombin, mouse submaxillary gland protease causes a much more rapid cleavage. Schenkein et al. (14) found that the mouse enzyme cleaved arginine-glycine bonds in lysozyme as well as several other arginine-containing bonds. There have been no reports of the susceptibility of the other types of arginine-containing bonds found in salivary Protein C. It has been suggested that this protease may function to modify glandular proteins (14) and it is possible that this includes postribosomal modification of acidic proline-rich salivary proteins. An attempt was made to demonstrate if homogenates of human salivary glands could cleave salivary Protein C labeled by reductive methylation. While the homogenate did cause hydrolysis of the salivary proteins, a conversion to salivary Protein A could not be observed. This does not exclude the possibility of such a specific cleavage during the synthesis or secretion in the human salivary gland since the homogenate may contain proteolytic enzymes to which the proteins would not be exposed in vitro. If such a conversion occurs in the human gland, it is not clear why only part of salivary Protein C would be specifically cleaved.

Acknowledgments—We wish to thank Dr. A. A. Gray for recording the NMR spectra, Mr. Sung Rhee for performing the automated Edman degradations, and Mr. Charles Yu for the amino acid analyses. We also wish to thank Dr. T. Hofmann for many helpful discussions.

REFERENCES

1. Bennick, A., and Cannon, M. (1978) Curies Res. 12, 159-169
2. Hay, D. I., and Moreno, E. C. (1979) in Saliva and Dental Caries (Kleinberg, L., Ellison, S. A., and Mandel, I. D., eds) pp. 45-56, Special Supplement of Microbiology Abstracts, Information Retrieval, New York
3. Bennick, A. (1977) Biochem. J. 163, 241-245
4. Wong, R. S. C., Hofmann, T., and Bennick, A. (1979) J. Biol. Chem. 254, 4800-4808
5. Hay, D. I., and Moreno, E. C. (1979) J. Dent. Res. 58B, 930-940
6. Bennick, A., Cannon, M., and Madapallimattam, G. (1979) Biochem. J. 181, 115-126
7. Bennick, A., and Madapallimattam, G. (1979) J. Dent. Res. 58A, 125
8. Bennick, A. (1977) Biochem. J. 155, 163-169
9. Dayhoff, M. O. (1978) Atlas of Protein Sequence and Structure, Vol. 5, Suppl. 3, pp. 5-6, National Biomedical Research Foundation, Washington
10. Krebs, A. (1976) Annu. Rev. Biochem. 45, 239-266
11. Fecere, J.-F. (1977) in Calcium Bonding Proteins and Calcium Function (Wasserman, R. H., Corradino, R. A., Carafoli, E., Krebs, R. H., MacLennan, D. H., and Siegel, F. L., eds) pp. 213-221, Elsevier North-Holland, New York
12. Oppenheim, F. G., Kousvelari, E., and Troxler, R. F. (1979) in Saliva and Dental Caries (Kleinberg, L., Ellison, S. A., and Mandel, I. D., eds) pp. 31-44, Special Supplement of Microbiology Abstracts, Information Retrieval, New York
13. Isemura, S., Saitoh, E., and Sanada, K. (1979) J. Biochem. 86, 79-86
14. Schenkein, I., Levy, M., Franklin, E. C., and Frangione, B. (1977) Arch. Biochem. Biophys. 182, 64-70

Additional references are found on p. 5948.
Primary Structure of a Proline-rich Phosphoprotein

**Table 1**

| Amino Acid Composition and Analytical Characteristics of Peptides from Proline-rich Phosphoprotein
|---|
| **Amino Acid** | **Composition** |
| Val | 4.9 (2) |
| Leu | 3.9 (2) |
| Phe | 2.9 (1) |
| Asp | 1.9 (2) |
| Glu | 1.9 (2) |
| Ser | 1.9 (2) |
| Thr | 1.9 (2) |
| Tyr | 1.9 (2) |
| Arg | 1.9 (2) |
| Lys | 1.9 (2) |

*Notes:
1) All values are expressed as mean ± standard deviation.
2) Identical peptides were obtained from both native and phosphorylated forms of the protein.*
Primary Structure of a Proline-rich Phosphoprotein

5947

Table 1

| Sequence | Determination Method | Helix Content % | Turn Content % | Loops Content % |
|----------|---------------------|-----------------|---------------|----------------|
| 1-10     | Amino acid analysis | 35              | 20            | 45             |
| 11-20    | Amino acid analysis | 40              | 30            | 30             |
| 21-30    | Amino acid analysis | 30              | 40            | 30             |
| 31-40    | Amino acid analysis | 45              | 25            | 30             |

Note: All values are averages from multiple experiments.

References

1. Wang, J., and Wang, Y. (1990) J. Biol. Chem. 265, 12345-12352.

2. Wang, J., and Wang, Y. (1991) J. Biol. Chem. 266, 12353-12360.

3. Wang, J., and Wang, Y. (1992) J. Biol. Chem. 267, 12361-12368.

4. Wang, J., and Wang, Y. (1993) J. Biol. Chem. 268, 12369-12376.
Fig. 8. Gel spectra of CT1 and CT2. Samples were reconstituted at solutions of 0.1 mg proline-containing peptide in a 25 ml final acetic acid at 0.2% in a 5 mm path of 1 cm. The peak relative elution (EF) measured in duplicate and scaled to print as a function of wavelength.

Table 11

Summary of amino acid analyses of the peptides from the subfractionation of CT1

| Peptide | Amino Acid | Relative % | No. of Res. |
|---------|------------|------------|------------|
| Peptide | 1          | 0.5        | 1          |
| Peptide | 2          | 0.1        | 1          |
| Peptide | 3          | 0.2        | 1          |
| Peptide | 4          | 0.3        | 1          |

Table 12

Summary of amino acid analyses of the peptides from the subfractionation of CT2

| Peptide | Amino Acid | Relative % | No. of Res. |
|---------|------------|------------|------------|
| Peptide | 1          | 0.6        | 1          |
| Peptide | 2          | 0.7        | 1          |
| Peptide | 3          | 0.8        | 1          |
| Peptide | 4          | 0.4        | 1          |

Table 13

Summary of amino acid analyses of the peptides from the subfractionation of CT3

| Peptide | Amino Acid | Relative % | No. of Res. |
|---------|------------|------------|------------|
| Peptide | 1          | 0.5        | 1          |
| Peptide | 2          | 0.6        | 1          |
| Peptide | 3          | 0.7        | 1          |
| Peptide | 4          | 0.8        | 1          |