Pure & Appl. Chem., Vol. 59, No. 11, pp. 1489–1492, 1987.
Printed in Great Britain.

From carbohydrates to cloning: regulation of gene expression

Don M. Carlson
Department of Biochemistry and Biophysics, University of California, Davis, CA 95616

Abstract - Unusual proteins and glycoproteins rich in proline (PRPs) are products of tissue-specific multigene families. Various biological functions such as calcium binding, hydroxylapatite binding, agglutination of oral bacteria, and formation of acquired dental pellicle have been proposed for PRPs in human saliva. Studies on the glycosylation and conformation of these proteins are presented together with the cloning and regulation of several mouse and hampster PRP genes.

Our recent research interests have focused on a group of unusual proteins and glycoproteins high in proline, or the so-called proline-rich proteins (PRPs). The PRPs are products of tissue-specific multigene families (1,2) and the cloning and regulation of several mouse and hamster PRP genes will be discussed in detail later. First I will reflect on some other exciting times in my life, especially in regards to glycoprotein chemistry, biosynthesis and immunochemical properties, starting with the synthesis of GDPglucose (3) and UDPGalNAc (4). I first became interested in glycoproteins as a postdoctoral fellow in Saul Roseman's laboratory. In 1962, Roseman (5) and Warren and Blacklow (6) demonstrated the synthesis of CMP-sialic acid. Subsequently, the first sialyltransferase was identified (7,8). By 1963 and 1964, it was shown that sialic acid was transferred from CMP-sialic acid to asialoglycoproteins such as sheep submaxillary mucin (9) and α_1-acid glycoprotein (10). These and other studies on glycosyltransferases led to the proposal of "multiglycosyltransferase systems" by Roseman (11).

In 1964, I began studies on a mucoid polysaccharide secreted by *Pseudomonas aeruginosa*. Doggett and co-workers had isolated a mucoid-type *Ps.* from the respiratory tract of cystic fibrosis patients (12). This viscous polysaccharide, according to their studies, contained glucose, galactose, glucosamine, galactosamine, sialic acid, and two unidentified substances. The material was negative to the naphthoresorcinol reaction and these investigators concluded that there was no uronic acid present. Subsequent studies from Linker's group (13) and from our laboratory (14) clearly showed that this viscous polymer contained only D-mannuronic acid and L-guluronic acid, the components of alginic acid.

Our interests in the chemistry, biosynthesis and immunochemistry of the carbohydrate components of glycoproteins continued with studies on pig submaxillary mucins (15,16) and thyroxine-binding globulin (17,18). About 1965, Elvin Kabat and his co-workers, mainly Gerald Schiffman and Kenneth Lloyd, published a part of their classical studies on the carbohydrate structures of the ABO-blood group substances. Helen Muir, Karl Meyer, Ward Pigman and others had shown that the carbohydrate moieties of proteoglycans and mucins (which include the blood group substances) were attached to serine and threonine via an O-glycosidic linkage. The sugar chains could be removed by an alkali-catalyzed ß-elimination reaction (Fig. 1), but with the blood group substances extensive degradation of the oligosaccharides occurred as a result of an additional ß-elimination and of the "peeling reaction" (19,20). By using the proper alkali and borohydride conditions, essentially quantitative recoveries of the oligosaccharides from pig submaxillary mucins (16) and blood group H substance (21) were obtained for complete structural studies. I of course was very appreciative when Elvin Kabat referred to these conditions as the "Carlson reaction" (22). The most complex oligosaccharide was the pentasaccharide Oligo-I from A^+-PSM, *i.e.* the mucin from pigs with blood type A. A^--PSM was missing the terminal α-linked GalNAc, the primary determinant of blood group A activity. These structural studies were confirmed by the first report on the *in vitro* synthesis of blood group A antigenic activity (23). The pentasaccharide Oligo-I had both sialic acid and fucose on the same chain, an event unknown before.

Proline-rich proteins - The primary focus of our research changed in 1975 when we became interested in the synthesis of PRPs. The synthesis of the protein portion plays a dominant and initial role in glycoprotein synthesis. As an overall view of protein synthesis, events from gene to protein (and glycoprotein) are shown in a simplistic manner in Fig. 2. The series of reactions involved in regulating gene transcription and RNA processing are receiving the major attention. Regulation occurs at each major step, however. The first step in cloning usually is the synthesis of complementary copies of the mRNAs by reverse transcriptase or a cDNA library. The cDNAs are cloned and sequenced which gives the amino acid sequence of various proteins. Also, cDNAs labeled with ^{32}P are used to search genomic libraries for specific genes.

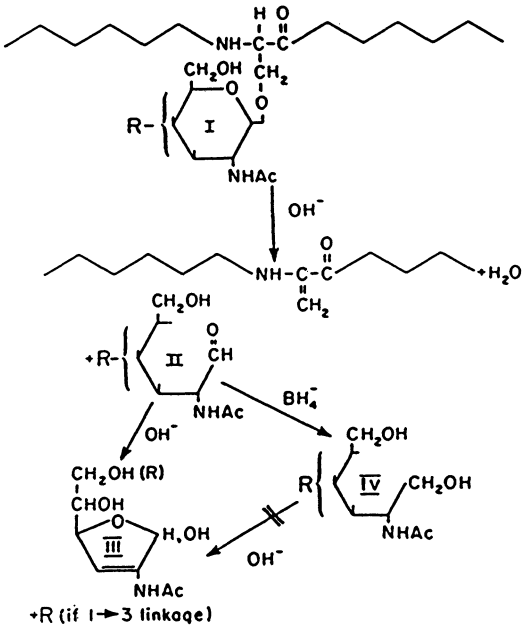

FIGURE 1. Effects of alkali and alkaline borohydride on the N-acetylgalactosaminyl-serine (-threonine) linkage and on N-acetyl-galactosamine.

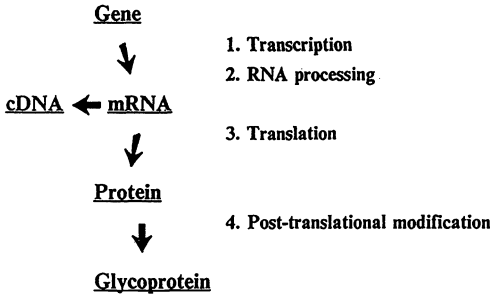

FIGURE 2. Gene expression and cDNA synthesis.

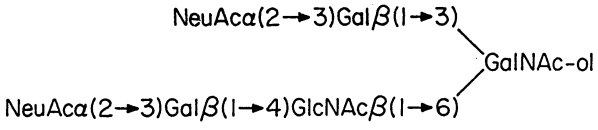

FIGURE 3. The major hexasaccharide from GP66sm.

With this background, we turn to studies on PRPs. In 1970, Bennick and Connell (24) and Oppenheim, Hay and Franzblau (25) isolated and partially characterized a series of PRPs from human saliva, and found that these unusual proteins comprised about 70% of the protein in human saliva. In rats and mice, unlike in humans where PRPs are constitutive, PRPs are generally not detected in the salivary glands of control animals. Treatment with the ß-agonist isoproterenol produces a dramatic induction of the tissue-specific PRP multigene families in rats (26-29,33), mice (1,30-33) and hamsters (2). After 10 days of treatment (2 to 5 mg isoproterenol/day), PRP mRNAs make up about 70% of the total mRNAs in the parotid glands (33). Proline-rich proteins have the following general characteristics:

1. PRPs are composed of 25-45% proline, 18-22% glycine and 18-22% glutamine; amounts of aromatic and sulfur-containing amino acids are either very low or are absent.

2. PRPs are basic or acidic, and may be glycosylated and/or phosphorylated.

3. PRPs are composed of four regions; signal peptide, transition region, repeat region, carboxyl-terminal region (see Fig. 4).

4. PRPs are acid-soluble proteins and glycoproteins (solubilities of PRPs in 10% trichloroacetic acid varies).

The major glycoprotein from the submandibular glands of isoproterenol-treated mice, GP66sm, has been studied in detail (31). GP66sm is 45% proline, 20% glycine, and 20% glutamine. Carbohydrate content is about 20% and molar ratios of GalNAc, GlcNAc, Gal and sialic acid are 1:1:2:2. The major oligosaccharide is a hexasaccharide (Fig. 3). The derived amino acid sequence of GP66sm (Fig. 4) shows a putative N-glycosylation site at Asn-46, but mannose was not detected. GP66sm is not present in submandibular glands of control animals and glycosylation sites are limited to threonine residues located in the repeat region. The hexasaccharide (Fig. 3) apparently is not present in control animals, which means that both a new peptide chain and a new oligosaccharide are synthesized in response to isoproterenol treatment.

The repeat region is highly conserved (2), but some substitutions are present (Fig. 4). For example, substitutions include glutamine and leucine for proline at position 9 of the repeat, alanine for glycine at position 5 and, significant for glycoprotein synthesis, threonine for proline at positions 3, 4 and 12. All substitutions involve only a one base change and all threonines in the repeat region are glycosylated. Glycosylation sites are introduced by the following changes: The repeat sequence PPPPGG---- is encoded CCA CCA CCA CCA GGC GGC-----; a codon change to CCA CCA CCA ACA GGC GGC now encodes PPPTGG---.

```
1M L V V L F T V A L L A L S S                    SIGNAL PEPTIDE

16A Q G P R E E L Q N Q I Q I P N Q R              TRANSITION REGION
34P P P S G F Q P R P P V N G S Q Q G

52P P P P G G P Q P R P P Q G                     REPEAT REGION
66P P P P G G P Q P R P P Q G
80P P P P G G P Q P R P P Q G
94P P P P G G P Q P R P P Q G
108P P P P G G P Q Q R P P Q G
122P P P P G G P Q P R P P Q G
136P P P P G G P Q L R P P Q G
150P P P P A G P Q P R P P Q G
164P P P P A G P Q P R P P Q G
178P P T T - G P Q P R P T Q G
191P P P T G G P Q Q R P P Q G
205P P P P G G P Q P R P P Q G
219P P P P G G P Q P S P T Q G

233P P P T G G P Q Q T P P L A G N T Q G           CARBOXYL TERMINUS
252P P Q G R P Q G P R STOP
                261
```

FIGURE 4. Derived amino acid sequence of GP66sm.

PRP Gene regulation - The upstream regions of the mouse PRP gene MP2 (1) and of the hamster gene H29 (2) each contain putative regulatory units for cAMP induction, an enhancer element and the expected TATA and CAAT boxes (Fig. 5). Isoproterenol stimulates cAMP synthesis which was considered the logical component for modulating transcription of the PRP multigene family. David Ann and Paul Wright in our laboratory have recently confirmed that deletion of the putative cAMP inducible sequence (-639 to -627 bp in MP2) completely obviates the induction of PRP mRNAs in transfected PC-12 cells.

Biological functions - Various functions such as calcium binding, hydroxylapatite binding, agglutination of oral bacteria, and formation of acquired dental pellicle have been proposed for PRPs in human saliva (34). Recently evidence for an unusual protective role or a defense mechanism of PRPs has been presented in response to tannins or tannic acid in the diet (35-37). There is a broad dietary appeal of tannin-rich foodstuffs, and antinutritional effects have been associated with dietary tannins (36,37). Hagerman and Butler (38) investigated interactions of various proteins with tannins and found that proteins high in proline or hydroxyproline have very high affinities for tannins. Subsequent studies showed that rats fed sorghum high in tannins responded similarly to animals injected with isoproterenol (35), *i.e.*, the parotid glands were grossly enlarged and PRP synthesis was dramatically induced. Also, when PRP synthesis was maximal at about 3 days after tannin feeding, an initial weight loss was reversed and the animals gained weight at close to the normal rate. These experiments, and others (36,37), have led us to believe that PRPs in saliva constitute the first-line of defense against tannins ingested, and possibly against other polyhydroxylated phenols.

Hamsters respond differently, however, to tannins in the diet. Like rats and mice, hamsters treated with isoproterenol showed a marked increase in PRPs (39), but there was little if any hypertrophy of the salivary glands. Feeding of tannins <u>did</u> <u>not</u> induce PRPs and hamsters on a diet containing 2% tannin lost weight for 3 days, as did rats and mice,

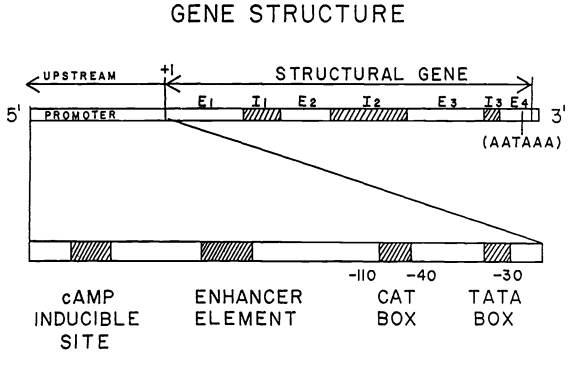

FIGURE 5. Diagram of structure of mouse PRP gene MP2.

but then an unusual growth inhibition occurred (39). Hamsters on the 2% tannin diet failed to gain weight and even at 60 days were essentially the same weight as at 3 days. When the diets were switched, the experimental animals gained weight at almost the normal rate for young hamsters, while the control animals, now on a 2% tannin diet, lost weight.

In about 30 days, both group of animals were close to the same weight. Clearly, the detrimental effects of tannins are reversed by the induction of PRPs in rats and mice, but hamsters lack this induction and are unusually susceptible to tannins. In fact, relatively high tannin levels (5 to 15%) in the diets of mice and rats have little overall effect, but a diet containing 4% tannins fed to hamsters is fatal to most animals within 3 days.

The affinity of the carbohydrate components of glycoproteins for tannins is of interest (36), and a systematic study of the effects of the carbohydrates on the salivary proline-rich glycoprotein GP66sm for the affinity of two condensed tannins has been performed (36,40).

REFERENCES

1. Ann, D.K. and Carlson, D.M. (1985) J. Biol. Chem. **260**, 15863.
2. Ann, D.K., Gadbois, D., and Carlson, D.M. (1987) J. Biol. Chem., in press.
3. Carlson, D.M., Swanson, A., and Roseman, S. (1964) Biochemistry **3**, 402.
5. Roseman, S. (1962) Proc. Natl. Acad. Sci. USA **48**, 437.
6. Warren, L. and Blacklow, R.S. (1962) Biochem,. Biophys. Res. Commun. **7**, 433.
7. Jourdian, G.W., Carlson, D.M., and Roseman, S. (1963) Biochem. Biophys. Res. Commun. **10**, 352.
8. Carlson, D.M., Jourdian, G.W., and Roseman, S. (1973) J. Biol. Chem. **248**, 5742.
9. Carlson, D.M., McGuire, E.J., Jourdian, G.W., and Roseman, S. (1973) J. Biol. Chem. **248**, 5763.
10. Bartholomew, B.A., Jourdian, G.W. and Roseman, S. (1973) J. Biol. Chem. **248**, 5751.
11. Roseman, S. (1970) Chem. Phys. Lipids **5**, 270.
12. Doggett, R.G., Harrison, G.M., Stillwell, R.N., and Wallis, E.S. (1965) J. Bacteriol. **89**, 476.
13. Linker, A. and Jones, R.A. (1966) J. Biol. Chem. **241**, 3845.
14. Carlson, D.M. and Matthews, L.W. (1966) Biochemistry **5**, 2817.
15. Carlson, D.M. (1966) J. Biol. Chem. **241**, 2984.
16. Carlson, D.M. (1968) J. Biol. Chem. **243**, 616.
17. Zinn, A.B., Marshall, J.S., and Carlson, D.M. (1978) J. Biol. Chem. **253**, 6761.
18. Zinn, A.B., Marshall, J.S., and Carlson, D.M. (1978) J. Biol. Chem. **253**, 6768.
19. Zinn, A.B., Plantner, J.J., and Carlson, D.M. (1977) Nature of Linkages Between Protein Core and Oligosaccharides in Glycoconjugates (M. Horowitz and W. Pigman, eds.), pp. 69-85, Academic Press, New York.
20. Lloyd, K.L., Kabat, E.A., Layug, E.J., and Gruezo, F. (1966) Biochemistry **5**, 1489.
21. Iyer, R.N. and Carlson, D.M. (1971) Arch. Biochem. Biophys. **142**, 101.
22. Kabat, E.A., Personal Communication.
23. Carlson, D.M. (1968) Studies on Pig Submaxillary mucin, In: *Biochemistry of Glycoproteins and Related Substances* (E. Rossi and E. Stoll, eds.), pp. 304-307, Karger/Basel, New York.
24. Bennick, A. and Connell, G.E. (1971) Biochem. J. **123**, 455.
25. Oppenheim, F.G., Hay, D.I., and Franzblau, C.P. (1971) Biochemistry **10**, 4233.
26. Muenzer, J., Bildstein, C., Gleason, M., and Carlson, D.M. (1979) J. Biol. Chem. **254**, 5623.
27. Muenzer, J., Bildstein, C., Gleason, M., and Carlson, D.M. (1979) J. Biol. Chem. **254**, 5629.
28. Mehansho, H. and Carlson, D.M. (1983) J. Biol. Chem. **258**, 6616.
29. Ziemer, M.A., Swain, W.F., Rutter, W.J., and Clements, S., Ann, D.K., and Carlson, D.M. (1984) J. Biol. Chem. **259**, 10475.
30. Azen, E.A., Carlson, D.M., Clements, S., Lalley, P.A., and Vanin, E. (1984) Science **266**, 967.
31. Mehansho, H., Clements, S., Sheares, B.T., Smith, S., and Carlson, D.M. (1985) J. Biol. Chem. **260**, 4418.
32. Clements, S., Mehansho, H., and Carlson, D.M. (1985) J. Biol. Chem. **260**, 13471.
33. Ann, D.K., Clements, S., Johnstone, E.M. and Carlson, D.M. (1987) J. Biol. Chem., in press.
34. Bennick, A. (1982) Mol. Cell. Biochem. **45**, 83.
35. Mehansho, H., Hagerman, A., Clements, S., Butler, L.G., Rogler, J., and Carlson, D.M. (1983) Proc. Natl. Acad. Sci. USA **80**, 3948.
36. Mehansho, H., Butler, L.G., and Carlson, D.M. (1987) Ann. Rev. Nutrition, in press.
37. Butler, L.G., Rogler, J.C., Mehansho, H., and Carlson, D.M. (1986) Dietary Effects of Tannins. In: *Plant Flavoniods in Biology and Medicine: Biochemical, Pharmacological, and Structure Activity Relationships*, pp. 141-157, Alan R. Liss, Inc.
38. Hagerman, A.E. and Butler, L.G. (1981) J. Biol. Chem. **256**, 4494.
39. Carlson, D.M., Mehansho, H., and Ann, D.K. (1985) Fed. Proc. **44**, 4015.
40. Asquith, T.N., Mehansho, H., Rogler, J.C., Butler, L.G., and Carlson, D.M. (1985) Fed. Proc. **44**, 1097.