Note

A homology between codon sequence and the linkage in glycoproteins*

M. JETT AND G. A. JAMIESON
The American National Red Cross, Blood Research Laboratory, Bethesda, Maryland 20014 (U. S. A.)
(Received July 11th, 1970; in revised form, February 4th, 1971; accepted February 19th, 1971)

Glycoproteins occur widely in the animal kingdom, but only three fundamental types of linkage have been observed, and their chemical reactivity depends mainly on the amino acid residue constituting the aglycon moiety. These linkages are (*a*) the relatively stable glycosylamine linkage involving the amido group of L-asparagine, observed in all plasma glycoproteins[1]; (*b*) the alkali-labile, glycosidic linkage to the hydroxyl group of L-serine or L-threonine, found in the submaxillary mucins and proteoglycans[1] and in the membrane glycoproteins of the red cell[2] and the platelet[3]; and (*c*) the alkali-stable, glycosidic linkage to the hydroxyl group of 5-hydroxy-L-lysine, found in collagen[4] and basement-membrane glycoprotein[5]. Despite the close structural similarity of L-glutamine to L-asparagine, and its frequent occurrence in purified glycopeptides, glycosylamine linkages to L-glutamine have not been detected, and a previous report that ester bonds to the carboxyl groups of L-glutamic acid and L-aspartic acid occur has been withdrawn[6].

Although there is virtually no structural similarity between the amino acids whose residues constitute the aglycons for the substituted glycosylamines and glycosides, several interesting homologies became apparent when the codons for the various amino acids involved in these two kinds of linkage were compared (see

TABLE I

CODONS FOR L-AMINO ACIDS IMPLICATED IN GLYCOSIDIC LINKAGES

L-Amino acid	*Codon*
Asn	AAU(C)
Ser	AGU(C)
Thr	ACU(C, A, G,
Lys	AAA(G)
Gln	CAA(G)
Pro	CCU(C, A, G)

*This note forms part of a thesis to be submitted for the degree of Doctor of Philosophy in the Department of Biochemistry, Georgetown University Schools of Medicine and Dentistry, Washington, D. C. Contribution No. 198 from the Blood Research Laboratory, The American National Red Cross.

Table I)[7,8]. In particular, it may be seen that one of the codons for L-asparagine (AAU) can, by single base-substitutions, give rise to the codons for L-serine (AGU), for L-threonine (ACU), and for L-lysine (AAA), known to be hydroxylated after incorporation into the protein[9]. However, the codon for L-glutamine (CAA) would require two base-substitutions.

Thus, it appears plausible that the amino acids linked glycosidically in glycoproteins and proteoglycans could have been derived from L-asparagine by single-point mutations of the original codon. As L-glutamine would require two, separate, point mutations, it appears unlikely that it would be found in glycosidic linkage on the basis of this hypothesis.

It may be noted that collagen from the body wall of the earthworm has carbohydrate linked to 5-hydroxy-L-lysine residues[10], whereas, 5-hydroxy-L-lysine is absent from the collagen of earthworm cuticle and the glycosidic linkages are to L-serine and L-threonine[11].

Ribonuclease is the only glycoprotein for which complete amino acid sequences have been established for several species. Porcine ribonuclease contains[12] glycosylated L-asparagine at residues 21, 34, and 76. In bovine ribonuclease, Asn_{21} has been replaced by Ser (and Asn_{76}, by Tyr)[13], and in rat ribonuclease, Asn residues 21 and 76 have been replaced by Ser (and Asn_{34}, by Gly)[14]. These mutations between Asn and Ser are consistent with this hypothesis, and the absence of glycosylated Ser residues may indicate the absence of the appropriate transferases in this tissue.

Several further inferences may be drawn from these two observations. (1) They imply an evolutionary relationship between the various types of glycopeptide linkage, the *N*-L-asparaginyl being the most primitive. (2) They suggest that only one of the two codons possible for L-asparagine, one of the six possible for L-serine, and one of the four possible for L-threonine are encoding for the linkage region in glycoproteins. (3) They would require that glycosidic linkages to 4-hydroxy-L-proline (CCU), which would also require a two-point mutation, will probably not be found in animals. Although such a linkage to L-arabinose has been proposed as being present in plant cell-wall glycoproteins[15], it has not been found in vertebrate collagens in which 5-hydroxy-L-lysine and 4-hydroxy-L-proline occur in approximately equal proportions. Similarly, it would be unlikely that 1-thioglycosides involving the sulfur atom of L-cysteine would occur in glycoproteins (although other types of 1-thioglycoside occur frequently in Nature[16]), as this would also require two, separate point-mutations (Cys = UGU). (4) The other single-point mutations that could arise from the Asn codon would be lethal in terms of glycoprotein biosynthesis, because the derived amino acids would not have active groups capable of glycoside formation. (5) The linking amino acid must form part of the recognition site in glycoprotein biosynthesis, because the monosaccharide forming the linkage differs in each of the three classes, and both types of transferase must be present in those tissues synthesizing glycoproteins in which at least two types of glycopeptide linkage occur, namely, fetuin[11], glomerular basement-membrane[17], human[18] IgA and rabbit IgG immunoglobulins[19], and the membrane glycoproteins[2,3].

Unfortunately, our knowledge of the amino acid and monosaccharide sequences, of the nature of the recognition site in biosynthesis, of biochemical phylogeny, and of structure–function relationships in the glycoproteins and proteoglycans is at present too limited to afford verification of the validity of these speculations.

ACKNOWLEDGMENT

This study was supported by USPHS Grant No. GM 13057.

REFERENCES

1 A. NEUBERGER, A. GOTTSCHALK, AND R. D. MARSHALL, in A. GOTTSCHALK (Ed.), *Glycoproteins, Their Composition, Structure and Function*, Elsevier, Amsterdam, 1966, p. 273.
2 D. B. THOMAS AND R. J. WINZLER, *J. Biol. Chem.*, 244 (1969) 5343.
3 D. S. PEPPER AND G. A. JAMIESON, *Biochemistry*, 9 (1970) 3706.
4 W. T. BUTLER AND L. W. CUNNINGHAM, *J. Biol. Chem.*, 241 (1966) 3882.
5 R. G. SPIRO, *J. Biol. Chem.*, 242 (1967) 813, 1923.
6 A. GOTTSCHALK AND W. KÖNIG, *Biochim. Biophys. Acta*, 158 (1968) 358.
7 M. NIRENBERG, T. COSKEY, R. D. MARSHALL, R. BRIMACOMBE, D. KELLOGG, B. DOCTOR, D. HATFIELD, J. LEVIN, F. RATTMAN, S. PESTKA, M. WILCOX, AND F. ANDERSON, *Cold Spring Harbor Symp. Quant. Biol.*, 31 (1966) 11.
8 H. G. KHORANA, H. BÜCHI, H. GOSH, N. GUPTA, T. M. JACOB, H. KOSSEL, R. MORGAN, S. A. NARANG, E. OHTSUKA, AND R. D. WEELS, *Cold Spring Harbor Symp. Quant. Biol.*, 31 (1966) 39.
9 N. BLUMENKRANTZ, J. ROSENBLOOM, AND D. J. PROCKOP, *Biochim. Biophys. Acta*, 192 (1969) 8.
10 L. MUIR AND Y. C. LEE, *J. Biol. Chem.*, 244 (1969) 2343.
11 R. G. SPIRO, *Ann. Rev. Biochem.*, 39 (1970) 599.
12 R. L. JACKSON AND C. H. HIRS, *J. Biol. Chem.*, 245 (1970) 637.
13 D. G. SMYTH, W. H. STEIN, AND S. MOORE, *J. Biol. Chem.*, 238 (1963) 227.
14 J. J. BIENTEMA AND M. GRUBER, *Biochim. Biophys. Acta*, 147 (1967) 612.
15 D. T. A. LAMPORT, *Biochemistry*, 8 (1969) 1155.
16 D. HORTON AND D. A. HUTSON, *Advan. Carbohyd. Chem.*, 18 (1963) 123.
17 R. G. SPIRO, *J. Biol. Chem.*, 242 (1967) 1923.
18 G. DAWSON AND J. R. CLAMP, *Biochem. J.*, 107 (1968) 341.
19 D. S. SMYTH AND S. UTSUMI, *Nature*, 216 (1967) 332.