PEPTIDE CHAIN CONFORMATION AND THE GLYCOSYLATION OF GLYCOPROTEINS

John G. Beeley
Department of Biochemistry, University of Glasgow, Glasgow G12 8QQ, U.K.

Received April 21,1977

SUMMARY: Methods for predicting peptide chain conformation have been applied to amino acid sequences adjacant to the carbohydrate attachment sites of glycoproteins containing the N-glycosylamine type of protein-carbohydrate linkage. Of 31 glycosylated residues examined 30 occur in sequences favouring turn or loop structures. Twentytwo of the glycosylated asparagine residues occur in tetrapeptides predicted to have the β-turn conformation. Carbohydrate attachment is therefore associated with peptide sequences which favour the formation of β-turn or other turn or loop structures.

It is well established that there are regularities in the amino acid sequence adjacent to the carbohydrate attachment sites of glycoproteins containing the N-glycosylamine type of protein - carbohydrate linkage. The sequence Asn (Carb.) - X - Ser/Thr is a necessary, but not a sufficient condition for carbohydrate attachment (1). Because of the paucity of high resolution crystallographic analyses of glycosylated proteins little is known about the peptide chain conformation near attachment sites or whether the process of carbohydrate attachment is related to conformation. However, studies on the sequences of glycopeptides from hen ovomucoid suggested a possible relationship between carbohydrate addition and the occurrence of turn conformations in this glycoprotein (2). This report is concerned with the application of predictive methods (3,4) to explore the possibility that glycosylated regions of the peptide chain might generally be associated with specific secondary structures.

The predictive method of Chou & Fasman (3) is based on the observed frequencies of occurrence of each amino acid in α-helical, β-sheet and β-turn regions of proteins whose structures are known. Comparison of the predicted and experimentally determined secondary structures of adenyl kinase has established the value of this approach

ISSN 0006-

(5). In addition to the β-turn which results in reversal of direction of the peptide chain (6) other looser turn or loop structures have been observed to occur widely in globular proteins (4).

METHODS

Amino acid sequences were examined for the presence of turn structures including hairpins, near hairpins, corners and loops by the location of uninterrupted sequences of three or more of any combination of the amino acids listed by Kuntz (4).

The occurrence of β-turns was predicted using the relative probability that a tetrapeptide will form a β-turn (7) given by

$$p_t = f_i f_{i+1} f_{i+2} f_{i+3}$$

where f_i, f_{i+1}, f_{i+2}, f_{i+3} are respectively the occurrence of a certain residue in the 1st, 2nd, 3rd and 4th positions of the β-turn. The averaged probability derived from 408 β-turns in 29 proteins for any given peptide to be in the β-turn is $p = 0.55 \times 10^{-4}$ and Chou & Fasman (8) have shown $p_t > 0.75 \times 10^{-4}$ to be a reasonable cutoff value for β-turn prediction. For all tetrapeptides having $p_t > 0.75 \times 10^{-4}$ the helix, β-sheet and β-turn conformational parameters (3,8) $\langle P_\alpha \rangle$, $\langle P_\beta \rangle$ and $\langle P_t \rangle$ were determined and β-turns predicted if

$$\langle P_\alpha \rangle < \langle P_t \rangle > \langle P_\beta \rangle$$

Using these cut-off values and without consideration of α or β regions Chou & Fasman (8) found that the accuracy of bend residues predicted correctly is $\%_t = 70\%$ and of non-bend residues predicted as non-bends is $\%_{nt} = 71\%$.

RESULTS AND DISCUSSION

Of thirtyone glycosylated asparagine residues in seventeen different polypeptide chains twentytwo would be predicted to occur in β-turn tetrapeptides (Table 1). In only one glycopeptide were the values of p_t for all tetrapeptides including the glycosylated asparagine less than the averaged probability for a peptide to be in a β-turn. These results show that carbohydrate attachment is frequently associated with amino acid sequences which favour the formation of the β-turn conformation.

For the glycosylated tetrapeptides predicted as turns Asn (Carb.) would occur as the fourth residue in four cases, the third residue in eleven and the first residue in seven cases. In several instances there is more than one point at which the turn might

Table 1

Glycoprotein		Glycosylated sequence							$p_t \times 10^{-4}$			
		n-3	n-2	n-1	n				n-3	n-2	n-1	n
Avidin	(17) hen	Leu	Gly	Ser	Asn	Met	Thr	Ile	0.6	1.5	0.1	0.5
Fibrinogen γ-chain	human	Glu	Val	Glu	Asn	Lys	Thr	Ser	0.2	0.7	0.3	1.3
Glycophorin	(26) human	Ser	Gln	Thr	Asn	Asp	Thr	His	0.7	1.2	1.0	0.6
α_1-Glycoprotein	(15) human	Pro	Ile	Thr	Asn	Ala	Thr	Leu	0.2	0.5	0.2	0.6
"	(38) "	Glu	Glu	Tyr	Asn	Lys	Ser	Val	0.3	0.7	0.5	1.2
"	(54) "	Phe	Thr	Pro	Asn	Lys	Thr	Ser	0.2	4.7	0.5	1.3
"	(75) "	Cys	Ile	Tyr	Asn	Thr	Thr	Tyr	0.5	0.4	0.3	1.4
"	(85) "	Gln	Arg	Gln	Asn	Gly	Thr	Ile	0.5	1.2	0.7	0.5
HCG α-subunit	(52) human	Val	Gln	Lys	Asn	Val	Thr	Ser	0.4	0.9	0.1	0.5
"	(78) "	Lys	Val	Glu	Asn	His	Thr	Ala	0.2	0.4	0.3	0.3
β-subunit	(13) "	Arg	Pro	Ile	Asn	Ala	Thr	Leu	0.2	0.4	0.1	0.6
	(30) "	Ile	Thr	Val	Asn	Thr	Thr	Ile	0.1	0.6	0.3	0.6
IgG - H chain	rabbit	Gln	Gln	Phe	Asn	Ser	Thr	Ile	0.4	0.6	0.5	0.8
IgG - L chain	mouse	Ala	Ser	Gln	Asn	Ile	Ser	Asn	0.3	1.3	0.1	0.6
Ovalbumin	hen	Glu	Lys	Tyr	Asn	Leu	Thr	Ser	0.7	0.5	0.2	0.3
Ovomucoid	(10) hen	Arg	Phe	Pro	Asn	Ala	Thr	Asp	0.1	2.0	0.2	0.6
"	(53) "	Phe	Gly	Thr	Asn	Ile	Ser	Lys	0.3	1.2	0.1	0.7
"	(69) "	Val	Pro	Met	Asn	Cys	Ser	Ser	0.2	2.0	0.7	1.1
"	(75) "	Ser	Tyr	Ala	Asn	Thr	Thr	Ser	0.2	0.9	0.5	1.2
Ribonuclease	(21) pig	Ser	Ser	Ser	Asn	Ser	Ser	Asn	1.9	3.4	1.3	2.5
"	(34) "	Ser	Arg	Arg	Asn	Met	Thr	Gln	1.1	0.8	0.1	0.8
"	(76) "	Tyr	Gln	Ser	Asn	Ser	Thr	Met	0.9	2.1	1.0	0.8
Stem bromelain	pine-apple	Pro	Arg	Asn	Asn	Glu	Ser	Ser	1.9	0.7	1.1	1.3
Thrombin	(77) cow	Asn	Tyr	Arg	Asn	Val	Ser	Val	0.9	0.9	0.2	0.5
"	(101) "	Pro	Glu	Ile	Asn	Ser	Thr	His	0.1	0.4	0.4	0.8
"	(376) "	Trp	Asx	Lys	Asn	Phe	Thr	Val	0.4	2.3	0.2	0.2
Thyroglobulin	human	Ala	Leu	Glu	Asn	Ala	Thr	Arg	0.1	0.4	0.1	0.7
Transferrin A	pig	Ser	Arg	Lys	Asn	Arg	Ser	Leu	0.8	1.3	0.5	1.5
" B	"	Thr	Ser	Asp	Asn	Leu	Ser	Ser	1.9	1.8	0.5	0.5
TSH α-subunit	(56) cow	Val	Pro	Lys	Asn	Ile	Thr	Ser	1.2	1.3	0.0	0.4
TSH β-subunit	(23) "	Leu	Thr	Ile	Asn	Thr	Thr	Val	0.1	0.4	0.2	0.6

β-turns in glycosylated sequences. Values of p_t were calculated for all tetrapeptides starting at the glycosylated Asn residue (n) and at one (n-1), two (n-2) and three (n-3) residues towards the N-terminal end. Underlined values indicate $p_t > 0.75$ and $\langle P_\alpha \rangle < \langle P_t \rangle > \langle P_\beta \rangle$. Where more than one peptide satisfies this condition the highest value of P_t is assumed to predict the β-turn beginning.

begin or double turns might be present. While β-turns with Asn in the first or third positions may be favoured sites for carbohydrate attachment it is recognised that there may be errors of ±1 residue in the location of β-turns by predictive methods (3, 7).

Some glycosylated sequences which are not predicted to be included in β-turns may occur in loops or turns of the peptide chain

of slightly different conformation. The location of a variety of turns (hairpin, near hairpin, corners and loops) has been predicted on the basis of the high correlation between their occurrence and uninterrupted sequences of 3 or more of any combination of a series of amino acids (4). Applying this method to the glycosylated sequences given in Table 1 thirty of thirtyone Asn (Carb.) residues would be predicted to occur as part of a turn as defined by Kuntz (4).

The suggestion that glycosylated Asn residues are frequently located within tetrapeptides having a turn conformation is supported by the observation that in crystals of human IgG antibody (Kol) the glycosylated Asn (297) in the heavy chain occurs in a β-turn (4). A β-turn has been observed in bovine pancreatic ribonuclease S involving Asn (34) which is glycosylated in ribonuclease B.

The glycosylation of β-turns and other turn conformations may be favoured because these structures are often associated (4) with the termination of helical or sheet structures at the surface of globular proteins and surface location favours glycosylation. There may in addition be some specific conformational requirement for glycosylation to occur. It is also possible that selective advantage (e.g. protection from proteolysis) may be derived from masking of turn conformations by the addition of carbohydrate.

ACKNOWLEDGEMENTS

I thank Dr. G.D. Fasman & Dr. P.Y. Chou for helpful comment and for allowing use of their data prior to publication.

REFERENCES

1. Marshall, R.D. (1972). Ann. Rev. Biochem., 41, 673-702.
2. Beeley, J.G. (1976). Biochem. J. 159, 335-345.
3. Chou, P.Y., and Fasman, G.D. (1974). Biochemistry, 13, 222-245.
4. Kuntz, I.D. (1972). J. Amer. Chem. Soc., 94, 4009-4012.
5. Schulz, G.E., Barry, C.D., Friedman, J., Chou, P.Y., Fasman, G.D., Finkelstein, A.V., Lim, V.I., Ptitsyn, O.D., Kabat, E.A., Wu, T.T., Levitt, M., Robson, B., and Nagamo, K. (1974). Nature, 250, 140-142.

6. Venkatachalam, C.M. (1968). Biopolymers, 6, 1425-1436.
7. Lewis, P.W., Momany, F.A., and Scheraga, H.A. (1971). Proc. natn. Acad. Sci. U.S.A., 68, 2293-2297.
8. Chou, P.Y., & Fasman, G.D. (1977). Advances Enzymol. In the press.
9. Spiro, R.G. (1973). Advances Protein Chem., 27, 350-456.
10. Marchesi, V.T., Furthmayr, H. and Tomita, M. (1976). Ann. Rev. Biochem. 45, 667-698.
11. Morgan, F.J., Birken, S., and Canfield, R.E. (1975). J. Biol. Chem. 250, 5247-5258.
12. Magnusson, S., Peteron, T.E., Sottrap-Jensen, L. and Claeys, H. (1975). Cold Spring Harbour Conferences on Cell Proliferation, 2, 123-149.
13. Graham. I., and Williams, J. (1975). Biochem. J., 14, 263-279.
14. Huber, R., Deisenhofer, J., Colman, P.M., Matsushina, M., and Palm, W. (1976). Nature, 264, 415-420.
15. Richards, F.M., and Wycoff, H.W. (1971). The Enzymes, 3rd Ed. Vol. 4, pp 647-806. Academic Press, New York.