Biochem. J. (1991) 275, 529–534 (Printed in Great Britain)

Amino acid distributions around *O*-linked glycosylation sites

Iain B. H. WILSON,* Ylva GAVEL†‡ and Gunnar VON HEIJNE‡

*Glycobiology Unit, Department of Biochemistry, University of Oxford, South Parks Road, Oxford OX1 3QU, U.K., †Research Group for Theoretical Biophysics, Department of Theoretical Physics, Royal Institute of Technology, S-100 44 Stockholm, Sweden, and ‡Department of Molecular Biology, Karolinska Institute, Center for Biotechnology, NOVUM, S-141 52 Huddinge, Sweden

To study the sequence requirements for addition of *O*-linked *N*-acetylgalactosamine to proteins, amino acid distributions around 174 *O*-glycosylation sites were compared with distributions around non-glycosylated sites. In comparison with non-glycosylated serine and threonine residues, the most prominent feature in the vicinity of *O*-glycosylated sites is a significantly increased frequency of proline residues, especially at positions −1 and +3 relative to the glycosylated residues. Alanine, serine and threonine are also significantly increased. The high serine and threonine content of *O*-glycosylated regions is due to the presence of clusters of several closely spaced glycosylated hydroxy amino acids in many *O*-glycosylated proteins. Such clusters can be predicted from the primary sequence in some cases, but there is no apparent possibility of predicting isolated *O*-glycosylation sites from primary sequence data.

INTRODUCTION

Proteins that pass through the secretory pathway are subjected to a variety of co- and post-translational covalent modifications that occur within the lumen of the endoplasmic reticulum and the Golgi apparatus. For example, *N*-glycosylation [1] and glypiation (addition of glycophosphatidylinositol membrane anchors) [2] occur co-translationally or within a short time of translocation across the rough-endoplasmic-reticulum membrane. In contrast, *O*-glycosylation [3] and tyrosine sulphation [4] are probably post-translational modifications that occur in the Golgi apparatus.

There are a number of categories of *O*-glycosylation known, which include: (*a*) addition of *N*-acetylgalactosamine to serine or threonine residues of animal proteins, especially mucins [5], (*b*) addition of mannose to serine or threonine residues of fungal, yeast and coral proteins [6–8], (*c*) addition of arabinose to hydroxyproline residues and galactose to serine residues of plant proteins [9], (*d*) addition of galactose to hydroxylysine residues of collagen [10] and C1q [11], (*e*) addition of xyloglucose to serine residues of blood clotting factors [12], (*f*) addition of xylose to serine residues of proteoglycan core proteins [13], and (*g*) addition of *N*-acetylglucosamine to serine and threonine residues of cytoplasmic and nuclear proteins [14]. The present study is concerned with the first of these, with the amino acid distribution around *N*-acetylgalactosamine-attachment sites being examined.

The minimum sequence requirements for *N*-glycosylation of asparagine residues are well established [15,16]. Glycosylated asparagine residues practically always occur in the sequence Asn-Xaa-(Ser/Thr)-Yaa, where Xaa and Yaa can be any amino acid except proline. In contrast, no consensus sequence for *O*-glycosylation has been found. Certain apparent biases in the amino acid distribution around *O*-glycosylated sites have been noted [17,18], with an apparently higher content of proline, serine and threonine in these regions. The general belief is that accessibility, rather than the exact primary sequence, determines whether a particular hydroxyamino acid is *O*-glycosylated or not. It has been proposed that *O*-glycosylated sites occur at reverse turns [19].

However, to our knowledge no careful statistical study on the sequences around *O*-linked glycosylation sites has been made. Experiments addressing the substrate-specificity of *N*-acetylgalactosaminyltransferase (UDP-*N*-acetyl-D-galactosamine:polypeptide *N*-acetylgalactosaminyltransferase, EC 2.4.1.41) have been performed [18,20], but these studies deal with a few acceptor sites, so only limited conclusions about the sequence constraints could be drawn. One theoretical study using secondary-structure prediction methods has been published [19], but only a few *O*-glycosylation sites were analysed. In another theoretical study, a larger sequence collection was used [17], but no thorough analysis of the data was presented. Since the sequences of a number of *O*-glycosylated proteins have been published in recent years, we have undertaken a more comprehensive study to collate the sequence information and to examine it for any primary sequence patterns that might exist.

MATERIALS AND METHODS

Preparation of sequence database

A search for *O*-glycosylated sequences was conducted with the use of the NBRF database [21] and sequences reported in the literature [17,21–74]. The resultant collection of *O*-glycosylated sites is listed in Table 1. The sequences were derived directly from protein, so it is possible to obtain information on post-translational modifications.

There exist a variety of means of identification of *O*-glycosylated residues in proteins. For example, if there is a blank cycle in Edman degradation, and a serine or threonine residue is detected by amino acid analysis or predicted from cDNA sequencing, then the blank position can be expected to represent a modified serine or threonine residue. Since galactosamine is detectable by amino acid analysis, a peptide with a blank position can be confirmed to be *O*-glycosylated.

Other methods for identification of *O*-glycosylation sites include treatment with alkaline sulphite or alkaline borohydride to convert glycosylated serine and threonine into their sulphonyl or reduced derivatives, conversion of dimethylaminoazobenzenethiohydantoin derivatives of amino acids with trifluoroacetic acid, dansyl–Edman degradation, back-hydrolysis of phenylthiohydantoin derivatives, and analysis of peptides following *O*-glycanase treatment.

From the sequence information, a sequence database was prepared of 21-residue segments centred around *O*-glycosylation

* To whom correspondence should be addressed.

Table 1. Collection of *O*-glycosylated sites

Residues in parentheses were excluded from the analysis in order to avoid distortion of the statistics.

Protein, source	*O*-Glycosylated sites	Included in minus database	Reference(s)
Alanine apolipoprotein (apo-CIII), human	74	Yes	[22]
α_1-Microglobulin, human	5	–	[23]
α_2HS-glycoprotein A-chain, human	238, 252	–	[24]
α_2HS-glycoprotein B-chain, human	6	–	[25]
Apolipoprotein E, human	194	Yes	[26,27]
C1 inhibitor, human	26, 42, 66, 70, 74	–	[28]
Chorionic gonadotropin α-subunit (free form), human	39	Yes	[29,30]
Chorionic gonadotropin β-subunit, human	121, 127, 132, 138	–	[31,32]
Echinoidin, sea-urchin	38	Yes	[33]
Erythropoietin, human	126	–	[34,35]
Factor X_1 heavy chain, bovine	300	–	[36]
Glycophorin, pig	1, 6, 11, 13, 21, 23, 30, 31, 41, 48	–	[37]
Glycophorin A, human	(2), (3), (4), (10), (11), (12), (13), (14), (15), (22), (25), 37, 44, 47, 50	–	[38]
Glycophorin B, human	2, 3, 4, 10, 11, 12	–	[39]
Glycophorin C, human	3, 4, 6, 9, 10, 15, 24, 26, 27, 28, 31, 32, 33, 42	–	[39,40,41]
Glycophorin HA, horse	2, 5, 7, 13, 17, 18, 20, 21, 24, 28	–	[42]
Glycoprotein 71, Friend murine leukaemia virus	268, 273, 275, 277, 279	–	[43,44]
Granulocyte-colony-stimulating factor, human	133	–	[45]
Haemopexin, human	1	Yes	[17,46]
HMW kininogen, human	383, 515, 524, 528, 539, 553, 559, 575, 610	Yes	[47]
HMW kininogen fragment 1, bovine	10, 11, (12), 18	–	[48,49]
HMW kininogen light chain, bovine	14, (22), (26), 38, (50), 55, (72)	–	[48,49]
IgA heavy chain, mouse MOPC 47A	218	Yes	[50]
IgA1 heavy chain, human Bur	224, 230, 232, 238, 240	–	[51]
IgD heavy chain, human NIG-65	(109), 110, 113, (126), (127), 131, 132	–	[52]
IgD heavy chain, human WAH	238, 255, 256	–	[53]
IgG heavy chain hinge region, rabbit	4	–	[54]
IgG light chain, human $S_m\lambda$	21	Yes	[55,56]
IgG3 heavy chain, human ZUC	17	–	[57]
Inter-α-trypsin inhibitor, human urine	10	–	[58]
Interleukin 2, human	3	Yes	[59]
Leucine-rich α_2-glycoprotein, human serum	2	Yes	[60]
Leukosialin, rat	6, 8, 13, 16, 18, 20, 21, 22, 26, 27, 29, 30, 33, 101, 106, 113, 117, 118, 119, 167, 169, 173, 180, 182	–	[61]
Pituitary glycopeptide, human	45	–	[62]
Plasminogen, human	345	–	[63,64]
Plasminogen, bovine	339	Yes	[65]
Platelet glycoprotein Ib α-chain, human	292	–	[66,67]
Polysialoglycoprotein, rainbow trout*	17, 18, 22	–	[68,69]
Protein Z, bovine	388	–	[70]
Sex-steroid-binding protein, human plasma	7	Yes	[71]
Somatostatin, catfish	5	Yes	[72]
Submaxillary mucin, tryptic peptides, sheep			
T1	1, 2, 7, 9, 17, 19, 20, 21, 27, 29, 30, 33, 35, 41, 42, 45, 50, 51	–	[73,74]
T4C	1, 6, 11, 14		
T4F	2, 3, 9, 12, 13, 14		

* This protein contains multiple internal repeats. Only the glycosylation sites in the first repeat have been included in the analysis.

sites (O^+ sites). Since some of the proteins in our collection were closely related and had highly similar sequences, a number of segments were excluded in order to avoid distortion of the statistics (see Table 1). This resulted in a database of 174 segments.

Two distinct categories of O^+ sites were delineated. Some proteins are *O*-glycosylated at one or a few isolated sites, whereas other proteins have clusters of several closely spaced *O*-glycosylation sites. In order to compare these groups, the subsamples 'singly glycosylated sites' (34 segments) and 'multiply glycosylated sites' (140 segments) were used. The latter sample contained those 21-residue segments that contain more than one glycosylation site. With this segment length, 93% of the sites from proteins with more than one *O*-glycosylation site fall within the sample of multiply glycosylated peptides. In the sample of singly glycosylated sequences, 24 peptides out of 34 are derived from proteins with only one glycosylation site.

As a control group, a database of 408 non-glycosylated sites (O^- sites) was prepared. In order to make sure that a useful comparison with the O^+ sites could be made, we restricted this sample to luminally oriented non-glycosylated serine and threonine residues from proteins that also have O^+ sites. For such residues, interaction with the *N*-acetylgalactosaminyltransferase that initiates *O*-glycosylation is theoretically possible. Since the

assignment of *O*-glycosylation sites in heavily glycosylated proteins may be ambiguous, only residues from proteins with at most a few well-established sites were included in the O^- sample (see Table 1).

Computational analysis of sequences

The sequences in the O^+ and O^- samples were aligned from their modified serine or threonine residues (position 0), and the positional amino acid distributions were calculated. Additionally, the overall frequency of amino acids immediately surrounding the sites was calculated (positions −3 to +3, excluding position 0). The significance of peaks in the amino acid distributions was estimated by calculating the probability of an equally high (or higher) peak occurring in a sample of random sequences. This calculation was made by using a binomial distribution with amino acid frequencies obtained from a sample of secreted proteins. (This sample was derived from the NBRF database and consists of 387 human proteins that are synthesized with signal sequences. For the purpose of the frequency calculation, the signal sequences were omitted.) The amino acid frequencies around O^- sites do not differ significantly from the frequencies in this control sample (results not shown).

Excluding membrane proteins and a protein with multiple repeats from the dataset, the relative position in the protein sequence was calculated for the O^+ sites in 53 proteins of known length. This was done to determine whether or not the probability of *O*-glycosylation is position-dependent.

In order to examine whether it is possible to predict clusters of glycosylation sites from the primary sequence, we prepared a sample consisting of 15 sequences of 50–60 residues (altogether about 800 residues) centred around heavily glycosylated regions. Some of the multiply glycosylated sites were omitted from this sample because of lack of sequence data extending far enough from the glycosylated region. Furthermore, some sites were replaced with sites from closely related proteins from the primary collection of *O*-glycosylated sequences in order to ensure that the clusters were of sufficient length for the analysis. (For example, the three sites from IgD of human WAH, which were used in the frequency calculations, were replaced with the corresponding sites from human NIG-65 so as to get a cluster of seven sites from the same protein.)

For the sequences in the cluster sample, the average number of serine and threonine residues per unit length was calculated as a function of the position in the sequence by using moving windows of various lengths. For comparison, those sequences from which the O^- sites were derived (altogether about 3000 residues) were analysed by the same method.

RESULTS AND DISCUSSION

Comparison of primary structure around O^+ and O^- sites

The comparison of O^+ and O^- sites reveals some differences in primary and secondary structure. We can now firmly establish earlier tentative conclusions [17,18] that there is a significantly increased frequency of proline, serine and threonine residues around O^+ sites as compared with O^- sites ($P < 10^{-29}$, 10^{-12} and 10^{-18} respectively). Alanine has also a somewhat increased frequency around O^+ sites ($P < 10^{-7}$). In contrast with *N*-glycosylation sites, which have a strong bias towards threonine in the Asn-Xaa-Ser/Thr signal [16], O^+ sites do not have a preference for threonine in position 0.

When the O^+ sample is subdivided into multiple and single glycosylation sites, it becomes apparent that the frequency of serine and threonine residues is significantly increased only around multiply glycosylated sequences (Fig. 1*a*). The increased frequency of hydroxy amino acid residues in multiply glycosylated sequences is due to the fact that 25% of the residues in this sample are glycosylated serine or threonine residues. The frequency of non-glycosylated serine and threonine residues is actually lower for multiply glycosylated sequences (6%) than for single glycosylated sequences (15%), which compares with the control sample, where 13.5% of the residues are serine or threonine.

Proline is the only amino acid residue that has a significantly increased frequency around single glycosylation sites ($P < 10^{-12}$; Fig. 1*b*). Statistically significant peaks in the positional amino acid distribution are found at positions −1 and +3 relative to the *O*-glycosylation site ($P < 10^{-7}$ for both peaks; Fig. 2). These peaks are not correlated: for singly glycosylated sequences with proline at position −1 proline is neither significantly increased nor reduced in frequency at position +3. Similarly, with proline at position +3 there is no increase or reduction in proline frequency at position −1 (results not shown).

Secondary-structure propensity of *O*-glycosylation sites

Previous studies [19] have attempted to correlate the addition of *O*-linked carbohydrate with the presence of reverse turns. A

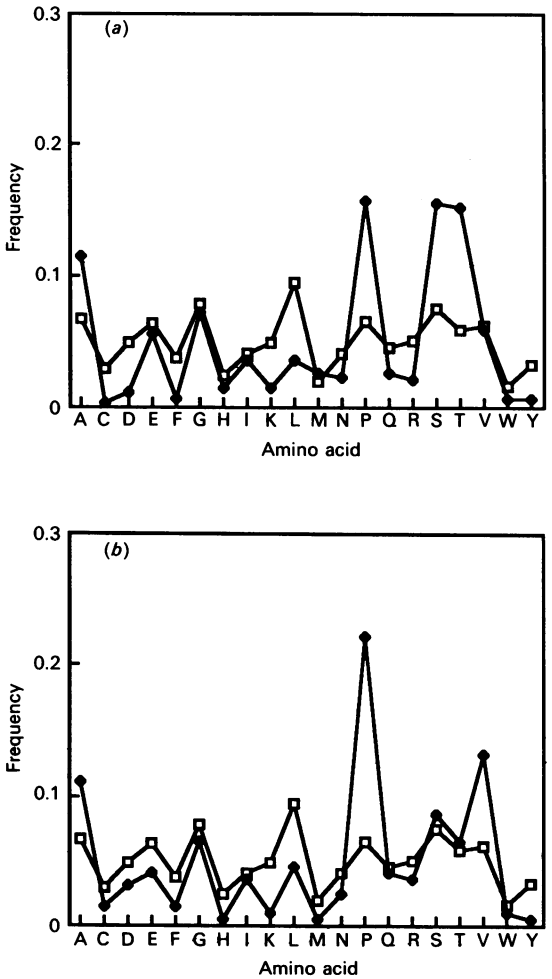

Fig. 1. Average amino acid frequencies in positions −3 to +3 (excluding position 0) surrounding *O*-glycosylation sites

(*a*) ◆, Frequencies around multiply glycosylated sites; □, frequencies in a control sample of exported proteins. (*b*) ◆, Frequencies around singly glycosylated sites; □, frequencies in a control sample of exported proteins.

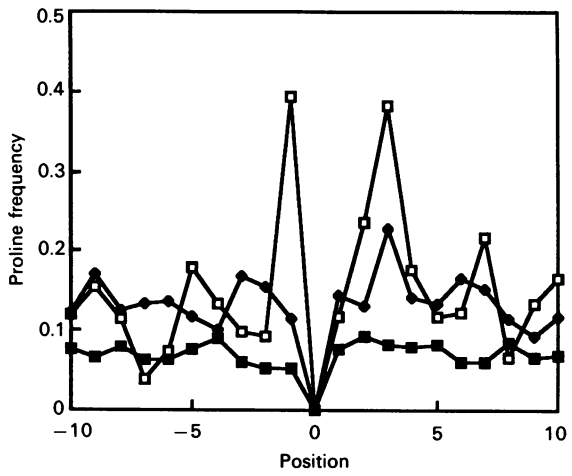

Fig. 2. Positional distributions of proline

■, Position 0 corresponds to a non-glycosylated serine/threonine in an *O*-glycosylated protein; □, position 0 corresponds to a glycosylated serine/threonine in a singly glycosylated sequence; ◆, position 0 corresponds to a glycosylated serine/threonine in a multiply glycosylated sequence.

reverse turn is a bend of the polypeptide backbone that involves four residues [75–78]. Proline, serine and threonine residues are common in such structures. Plots of α-propensity, β-propensity, turn propensity and hydrophobicity are compatible with the amino acid residue distributions (results not shown). Owing to the high frequency of serine and threonine residues, multiply glycosylated sequences have a low average α-propensity and a high turn propensity.

The probability of occurrence of a certain amino acid residue is dependent on the relative position in the turn. Positional preferences have been published [76,77]. According to the data of Chou & Fasman [78], the enhancement in proline frequency in position −1, found in the present study, is compatible with structures where proline and serine or threonine residues occur in a reverse turn. The enhancement in proline frequency at position +3 cannot be similarly interpreted, but could represent a structure where the serine or threonine is just *N*-terminal to the turn. Thus the high frequency of proline residues around the O^+ sites may indicate that *O*-glycosylation tends to occur in regions of the protein that exist as reverse turns, but not necessarily coincident with the turns. Since turns are often located at the protein surface, these results are consistent with the post-translational model of *O*-glycosylation. If *O*-glycosylation takes place in the Golgi apparatus, as most data indicate, then the protein is already folded at this stage and so accessibility would be determining factor.

Prediction of *O*-glycosylation sites

Since the primary sequence patterns around *O*-glycosylation sites are not very specific, it is probably impossible to design a reliable method for prediction of individual sites. With single glycosylation sites, the patterns Pro-(Ser/Thr) and (Ser/Thr)-Xaa-Xaa-Pro are statistically significant but are not sufficiently unique to be of any predictive value. For example, those glycoprotein sequences from which the O^- sites were derived contain 66 such patterns, only six of which are actually glycosylation sites. In multiply glycosylated sequences there appear not to be any features that distinguish serine and threonine residues that are glycosylated from those that remain non-glycosylated.

Although the exact glycosylation pattern of a protein cannot be predicted from the primary sequence, nevertheless the high incidence of serine and threonine residues in heavily glycosylated regions offers possibilities for prediction. However, the expected frequency of serine and threonine residues in an average protein is rather high (13.5%), making it difficult to distinguish random fluctuations from actual multiply glycosylated regions. Therefore the statistical significance of local clusters of serine and threonine residues must be assessed when predictions are made.

Possible candidates for clusters of O^+ sites can be found by scanning the primary sequence for regions that are rich in serine and threonine residues. A convenient way of doing this is to use a hydrophobicity analysis program with a moving window and a hydrophobicity scale where serine and threonine have the value 1.0 and the other residues have the value 0.0. The longer the window the lower is the probability of getting a high random fluctuation in the frequency of serine and threonine residues. However, if the window is very long, short clusters of O^+ sites may be overlooked. With our sample, the optimum window length proved to be about 30 residues. For a 30-residue window, 12 serine and threonine residues (40%) are needed to produce a fluctuation with $P < 10^{-3}$. (This number was calculated by using a binomial distribution with a frequency of 13.5%). With this cut-off, we found significant peaks in ten sequences out of 15 in the cluster sample. In contrast, only one peak was found in the large sample without clustered glycosylation sites that was scanned for comparison. We conclude that significant clusters of serine and threonine residues ($P < 10^{-3}$) seem to be good candidates for *O*-glycosylation.

O-Glycosylation relative to *N*-glycosylation

The *N*-acetylgalactosaminyltransferase is capable of adding *O*-linked carbohydrate not only to consecutive serine and threonine residues (six consecutive residues are *O*-glycosylated in glycophorin B [40]), but also immediately adjacent to *N*-glycosylation sites. In the tripeptide Asn-Xaa-(Ser/Thr), both the asparagine and serine or threonine residues are glycosylated in pig glycophorin (residues 19–21 and 39–41 [38]), human glycophorin C (residues 8–10 [40–42]) and glycoprotein gp71 of Friend murine leukaemia virus (residues 266–268 [44]). Since *O*-glycosylation is probably post-translational, this suggests that the presence of a neighbouring *N*-linked oligosaccharide does not interfere with *O*-glycosylation. It is pertinent to note that, even though most data would indicate that *O*-glycosylation is post-translational [3,79–82], the data are not unanimous, particularly with mucin biosynthesis [83,84].

It has been previously noted that *N*-glycosylation sites have a tendency to occur towards the *N*-terminus of proteins [16]. This does not appear to be the case with *O*-glycosylation sites. The probability of an O^+ site occurring is independent of the relative distance from the *N*- and *C*-termini, with an even distribution of *O*-glycosylation sites along the length of polypeptides (results not shown). This is compatible with *O*-glycosylation being independent of translocation across the endoplasmic-reticulum membrane and occurring after protein folding. It is also interesting to note the possibility that *O*-glycosylated serine residues occur immediately adjacent to the transmembrane domains of proteins such as γ-glutamyltranspeptidase [85].

Conclusion

Our data indicate that some biases exist in the sequence characteristics of regions surrounding *O*-glycosylation sites; in particular, proline is significantly enriched in positions −1 and +3. It appears, though, that these patterns are not sufficiently specific to allow prediction of isolated sites. However, the approximate location of heavily glycosylated regions can still be predicted with some confidence owing to the localized high content of serine and threonine residues.

Since the frequency of serine and threonine residues is enhanced only in multiply glycosylated regions, it can be concluded that such an increased frequency of itself does not constitute a necessary recognition signal for *N*-acetylgalactosaminyltransferase. Rather, it would seem that such heavily glycosylated regions have come direct functional or structural significance, e.g. to serve as 'stalks' in membrane-bound receptors [86,87] and brush-border enzymes [88] or to confer rigidity to the polypeptide backbone [89,90].

This work was supported by a grant from the Swedish Natural Sciences Research Council to G.v.H. I.B.H.W. is supported by the U.K. Medical Research Council and is a member of the Oxford Glycobiology Unit (which is funded by Monsanto Co.).

REFERENCES

1. Kornfeld, R. & Kornfeld, S. (1985) Annu. Rev. Biochem. **54**, 631–664
2. Thomas, J., Dwek, R. A. & Rademacher, T. W. (1990) Biochemistry **29**, 5413–5422
3. Roth, J. (1987) Biochim. Biophys. Acta **906**, 405–436
4. Baeuerle, P. A. & Huttner, W. B. (1987) J Cell Biol. **105**, 2655–2664
5. Sadler, J. E. (1984) in The Biology of Carbohydrates (Ginsburg, V. & Robbins, P. W., eds.), vol. 2, pp. 199–288, Wiley–Interscience, New York
6. Svensson, B., Klarsen, A. & Gunnarsson, A. (1986) Eur. J. Biochem. **154**, 497–502
7. Ballou, C. (1976) Adv. Microb. Phys. **14**, 93–158
8. Meikle, P. N., Richards, G. N. & Yellowlees, D. (1987) J. Biol. Chem. **262**, 16941–19647
9. Allen, A. K., Desai, N. N., Neurberger, A. & Creeth, J. M. (1978) Biochem. J. **171**, 665–674
10. Spiro, R. G. (1973) Adv. Protein Chem. **27**, 349–467
11. Reid, K. B. M. (1979) Biochem. J. **179**, 367–371
12. Hase, S., Ikenaka, T., Takao, T., Shimonishi, Y. & Iwanaga, S. (1989) J. Biol. Chem. **264**, 20320–20325
13. Rodén, L. (1980) in The Biochemistry of Glycoproteins and Proteoglycans (Lennarz, W. J., ed.), pp. 267–371, Plenum Press, New York
14. D'Onofrio, M., Starr, C. M., Park, M. K., Holt, G. D., Haltiwanger, R. S., Hart, G. W. & Hanover, J. A. (1988) Proc. Natl. Acad. Sci. U.S.A. **85**, 9595–9599
15. Marshall, R. D. (1972) Annu. Rev. Biochem. **41**, 673–702
16. Gavel, Y. & von Heijne, G. (1990) Protein Eng. **3**, 433–442
17. Takahashi, N., Takahashi, Y. & Putnam, F. W. (1984) Proc. Natl. Acad. Sci. U.S.A. **81**, 2021–2025
18. Briand, J. P., Andrews, S. P., Cahill, E., Conway, N. A. & Dillaha Young, J. (1981) J. Biol. Chem. **256**, 12205–12207
19. Aubert, J.-P., Biserte, G. & Loucheux-Lefebvre, M.-H. (1976) Arch. Biochem. Biophys. **175**, 410–418
20. Hughes, C. H., Bradbury, A. F. & Smyth, D. G. (1988) Carbohydr. Res. **178**, 259–269
21. George, D. G., Barker, W. C. & Hunt, L. T. (1986) Nucleic Acids Res. **14**, 11–15
22. Brewer, H. B., Shulman, R., Herbert, P., Ronan, R. & Wehrly, K. (1974) J. Biol. Chem. **249**, 4975–4984
23. Takagi, T., Takagi, K. & Kawai, T. (1981) Biochem. Biophys. Res. Commun. **98**, 997–1001
24. Yoshioka, Y., Gejyo, F., Marti, T., Rickli, E. E., Burgi, W., Offner, G. D., Troxler, R. F. & Schmid, K. (1986) J. Biol. Chem. **261**, 1665–1676
25. Gejyo, F., Chang, J.-L., Burgi, W., Schmid, K., Offner, G. D. & Troxler, R. F. (1983) J. Biol. Chem. **258**, 4966–4971
26. Wernette-Hammond, M. E., Lauer, S. J., Corsini, A., Walker, D., Taylor, J. M. & Rall, S. C. (1989) J. Biol. Chem. **264**, 9094–9101
27. Rall, S. C., Weisgraber, K. H. & Mahley, R. W. (1982) J. Biol. Chem. **257**, 4171–4178
28. Bock, S. C., Skriver, K., Nielsen, E., Thogersen, H.-C., Wiman, B., Donaldson, V. H., Eddy, R. L., Marrinan, J., Radziejewska, E., Huber, R., Shows, T. B. & Magnusson, S. (1986) Biochemistry **25**, 4292–4301
29. Cole, L. A., Perini, F., Birken, S. & Ruddon, R. W. (1984) Biochem. Biophys. Res. Commun. **122**, 1260–1267
30. Fiddles, J. C. & Goodman, H. M. (1979) Nature (London) **281**, 351–356
31. Morgan, F. J., Birken, S. & Canfield, R. (1975) J. Biol. Chem. **250**, 5247–4258
32. Kessler, M. J., Mise, T., Ghai, R. D. & Bahl, O. P. (1979) J. Biol. Chem. **254**, 7909–7914
33. Giga, Y., Ikai, A. & Takahashi, K. (1987) J. Biol. Chem. **262**, 6197–6203
34. Broudy, V. C., Tait, J. F. & Powell, J. S. (1988) Arch. Biochem. Biophys. **265**, 329–336
35. Lai, P.-H., Everett, R., Wang, F.-F., Arakawa, T. & Goldwasser, E. (1986) J. Biol. Chem. **261**, 3116–3121
36. Titani, K., Fujikawa, K., Enfield, D. L., Ericsson, L. H., Walsh, K. A. & Neurath, H. (1975) Proc. Natl. Acad. Sci. U.S.A. **72**, 3082–3086
37. Honma, K., Tomita, M. & Hamada, A. (1980) J. Biochem. (Tokyo) **88**, 1679–1691
38. Tomita, M. & Marchesi, V. T. (1975) Proc. Natl. Acad. Sci. U.S.A. **72**, 2964–2968
39. Blanchard, D., Dahr, W., Hummel, M., Latron, F., Beyreuther, K. & Cartron, J.-P. (1987) J. Biol. Chem. **262**, 5808–5811
40. Dahr, W., Hummel, M., Blanchard, D., Beyreuther, K. & Cartron, J. (1985) Biol. Chem. Hoppe-Seyler **366**, 777–778
41. Dahr, W., Beyreuther, K., Kordowicz, M. & Krüger, J. (1982) Eur. J. Biochem. **125**, 57–62
42. Murayama, J.-I., Tomita, M. & Hamada, A. (1982) J. Memb. Biol. **64**, 205–215
43. Geyer, R., Dabrowski, J., Dabrowski, U., Linder, D., Schlüter, M., Schott, H.-H. & Stirm, S. (1990) Eur. J. Biochem. **187**, 95–110
44. Chen, R. (1982) Proc. Natl. Acad. Sci. U.S.A. **79**, 5788–5792
45. Kubota, N., Orita, T., Hattori, K., Oh-eda, M., Ochi, N. & Yamazaki, T. (1990) J. Biochem. (Tokyo) **107**, 486–492
46. Takahashi, N., Takahashi, Y. & Putnam, F. W. (1985) Proc. Natl. Acad. Sci. U.S.A. **82**, 73–77
47. Kellerman, J., Lottspeich, F., Henschen, A. & Muller-Esterl, W. (1986) J. Biochem. (Tokyo) **154**, 471–478
48. Lottspeich, F., Kellermann, J., Henschen, A., Foertsch, B. & Muller-Esterl, W. (1985) Eur. J. Biochem. **152**, 307–314
49. Iwanaga, S., Kato, H., Sugo, T., Ikari, N. & Fujii, S. (1979) in Biological Functions of Proteinases (Holzer, H. & Tschesche, H., eds.), pp. 243–259, Springer-Verlag, Berlin
50. Robinson, E. A. & Appella, E. (1979) J. Biol. Chem. **254**, 11418–11430
51. Putnam, F. W., Liu, Y.-S. V. & Low, T.L. K. (1979) J. Biol. Chem. **254**, 2865–2874
52. Takayasu, T., Suzuki, S., Kametani, F., Takahashi, N., Shinoda, T., Okuyama, T. & Munekata, E. (1982) Biochem. Biophys. Res. Commun. **105**, 1066–1071
53. Takahashi, N., Tetaert, D., Debuire, B., Lin, L.-C. & Putnam, F. W. (1982) Proc. Natl. Acad. Sci. U.S.A. **79**, 2850–2854
54. Smyth, S. & Utsumi, S. (1967) Nature (London) **216**, 332–335
55. Garver, F. A., Chang, L., Mendicino, J., Isobe, T. & Osserman, E. (1975) Proc. Natl. Acad. Sci. U.S.A. **72**, 4559–4563
56. Chandrasekaran, E. V., Mendicino, A., Garver, F. A. & Mendicino, J. (1981) J. Biol. Chem. **256**, 1549–1555
57. Frangione, B. & Milstein, C. (1969) Nature (London) **224**, 597–599
58. Hochstrasser, K., Schönberger, Ö. L., Rossmanith, I. & Wachter, E. (1981) Hoppe-Seyler's Z. Physiol. Chem. **362**, 1357–1362
59. Robb, R. J., Kutny, R. M., Panico, M., Morris, H. R. & Chowdhry, V. (1984) Proc. Natl. Acad. Sci. U.S.A. **81**, 6486–6490
60. Takahashi, N., Takahashi, Y. & Putnam, F. W. (1985) Proc. Natl. Acad. Sci. U.S.A. **82**, 1906–1910
61. Killeen, N., Barclay, A. N., Willis, A. C. & Williams, A. F. (1987) EMBO J. **6**, 4029–4034
62. Seidah, N. G. & Chrétien, M. (1981) Proc. Natl. Acad. Sci. U.S.A. **78**, 4236–4240
63. Hayes, M. L. & Castellino, F. J. (1979) J. Biol. Chem. **254**, 8777–8780
64. Malinowski, D. P., Sadler, J. E. & Davie, E. W. (1984) Biochemistry **23**, 4243–4250
65. Schaller, J., Moser, P. W., Dannegger-Muller, G. A. K., Rosselet, S. J., Kämpfer, U. & Rickli, E. E. (1985) Eur. J. Biochem. **149**, 267–278
66. Lopez, J. A., Chung, D. W., Fujikawa, K., Hagen, F. S., Papayannopoulou, T. & Roth, G. J. (1987) Proc. Natl. Acad. Sci. U.S.A. **84**, 5615–5619

67. Titani, K., Takio, K., Handa, M. & Ruggeri, Z. M. (1987) Proc. Natl. Acad. Sci. U.S.A. **84**, 5610–5614
68. Sorimachi, H., Emori, Y., Kawasaki, H., Kitajima, K., Inoue, S., Suzuki, K. & Inoue, Y. (1988) J. Biol. Chem. **263**, 17678–17684
69. Kitajima, K., Inoue, Y. & Inoue, S. (1986) J. Biol. Chem. **261**, 5262–5269
70. Hojrup, P., Jensen, M. S. & Petersen, T. (1985) FEBS Lett. **184**, 333–338
71. Walsh, K. A., Titani, K., Takio, K., Kumar, S., Hayes, R. & Petra, P. H. (1986) Biochemistry **25**, 7584–7590
72. Andrews, P. C., Pubols, M. H., Hermodson, M. A., Sheares, B. T. & Dixon, J. E. (1984) J. Biol. Chem. **259**, 13267–13272
73. Hill, H. D., Schwayzer, M., Steinman, H. M. & Hill, R. L. (1977) J. Biol. Chem. **252**, 3799–3804
74. Hill, H. D., Reynolds, J. A. & Hill, R. L. (1977) J. Biol. Chem. **252**, 3791–3798
75. Lewis, P. N., Momany, F. A. & Scheraga, H. A. (1971) Proc. Natl. Acad. Sci. U.S.A. **68**, 2293–2297
76. Crawford, J. L., Lipscomb, W. N. & Schellman, C. G. (1973) Proc. Natl. Acad. Sci. U.S.A. **70**, 538–542
77. Chou, Y. & Fasman, G. D. (1974) Biochemistry **13**, 222–244
78. Levitt, M. (1978) Biochemistry **17**, 4277–4285
79. Roth, J. (1984) J. Cell Biol. **98**, 399–406
80. Piller, V., Piller, F., Klier, F. G. & Fukuda, M. (1989) Eur. J. Biochem. **183**, 123–135
81. Morrow, B. & Rubin, C. S. (1987) J. Biol. Chem. **262**, 13812–13820
82. Hanover, J. A., Elting, J., Mintz, G. R. & Lennarz, W. J. (1982) J. Biol. Chem. **257**, 10172–10177
83. Strous, G. J. A. M. (1979) Proc. Natl. Acad. Sci. U.S.A. **76**, 2694–2698
84. Spielman, J., Rockley, N. L. & Carraway, K. L. (1987) J. Biol. Chem. **262**, 269–275
85. Blochberger, T. C., Sabatine, J. M., Lee, Y. C. & Hughey, R. P. (1989) J. Biol. Chem. **264**, 20718–20722
86. Yamamoto, T., Davis, C. G., Brown, M. S., Schneider, W. J., Casey, M. L., Goldstein, J.-L. & Russell, D. W. (1984) Cell **39**, 27–38
87. Kajinami, K., Mabuchi, H., Itoh, H., Michishita, I., Takeda, M., Wakasugi, T., Koizumi, J. & Takeda, R. (1988) Arteriosclerosis **8**, 187–192
88. Hunziker, W., Spiess, M., Semenza, G. & Lodish, H. F. (1987) Cell **46**, 227–234
89. Perkins, S. J., Smith, K. F., Amatayakul, S., Ashford, D., Rademacher, T. W., Dwek, R. A., Lachmann, P. J. & Harrison, R. A. (1990) J. Mol. Biol. **214**, 751–763
90. Gerken, T. A., Butenhof, K. J. & Shogren, R. (1989) Biochemistry **28**, 5536–5543

Received 10 October 1990; accepted 14 November 1990