The Amino Acid at the X Position of an Asn-X-Ser Sequon Is an Important Determinant of N-Linked Core-glycosylation Efficiency*

(Received for publication, November 8, 1995, and in revised form, December 28, 1995)

Susan H. Shakin-Eshleman‡§, Steven L. Spitalnik*, and Lakshmi Kasturi‡

From the Case Western Reserve University, School of Medicine, Cleveland, Ohio 44106 and the University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania 19104

N-Linked glycosylation is a common form of protein processing that can profoundly affect protein expression, structure, and function. N-Linked glycosylation generally occurs at the sequon Asn-X-Ser/Thr, where X is any amino acid except Pro. To assess the impact of the X amino acid on core glycosylation, rabies virus glycoprotein variants were generated by site-directed mutagenesis with each of the 20 common amino acids substituted at the X position of an Asn-X-Ser sequon. The efficiency of core glycosylation at the sequon in each variant was quantified in a rabbit reticulocyte lysate cell-free translation system supplemented with canine pancreas microsomes. The presence of Pro at the X position completely blocked core glycosylation, whereas Trp, Asp, Glu, and Leu were associated with inefficient core glycosylation. The other variants were more efficiently glycosylated, and several were fully glycosylated. These findings demonstrate that the X amino acid is an important determinant of N-linked core-glycosylation efficiency.

One of the most common types of protein modification is N-linked glycosylation, in which oligosaccharides are added to specific Asn residues (1, 2). N-Linked glycosylation plays a critical role in the expression of most cell-surface and secreted proteins and is often required for protein stability, antigenicity, and biological function (1, 3–6). The effects of N-linked glycosylation often depend on the number and position of N-linked oligosaccharides added to a protein chain (5, 7–11). This is determined during core glycosylation, in which the oligosaccharide Glc3Man9GlcNAc2 is transferred to a protein by the enzyme oligosaccharyltransferase (2, 12, 13). Oligosaccharyltransferase is integral to the endoplasmic reticulum membrane, and the active site of the enzyme resides near the endoplasmic reticulum membrane on the lumenal side (13–15). Core glycosylation usually occurs co-translationally as the glycosylation site on a nascent protein enters the endoplasmic reticulum lumen (14, 16–18).

Despite the importance of N-linked glycosylation, little is known about the protein signals that control the efficiency of oligosaccharide addition at specific Asn residues. N-Linked glycosylation generally occurs at the sequon Asn-X-Ser or Asn-X-Thr, where X is any amino acid except proline (Asn-X-Ser/Thr) (15, 19, 20). However, because many Asn-X-Ser/Thr sequons in proteins are glycosylated inefficiently (8, 21–26) or not at all (20, 27, 28), other protein signals must also control this process.

We have used rabies virus glycoprotein (RGP)1 as a model system to study the regulation of N-linked core glycosylation (8, 29, 30). Using a rabbit reticulocyte lysate cell-free translation system supplemented with canine pancreas microsomes, we can examine the effects of specific amino acid substitutions on the core-glycosylation efficiency (CGE) of individual sequons in RGP (29). Our results in the cell-free system are similar to those obtained when RGP variants are expressed in transfected Chinese hamster ovary cells (8, 29). In this report we examine the impact of the X amino acid on CGE. To do this we generated a set of RGP variants by site-directed mutagenesis in which each of the 20 common amino acids was substituted at the X position of the sequon Asn37-Leu38-Ser39. We then quantified the CGE at the sequon in each variant using the cell-free system described. Our results demonstrate that the amino acid at the X position is an important determinant of CGE.

MATERIALS AND METHODS

Construction of a Cloning Vector for Cassette Mutagenesis—A cloning vector for cassette mutagenesis was generated from the plasmid that encodes RGP(1–) by the introduction of unique EcoRV and Sad restriction sites on either side of the sequon (Fig. 2). This was accomplished using the polymerase chain reaction-based method, splicing by overlap extension (31), essentially as described (8). Briefly, polymerase chain reaction amplification of RGP cDNA was performed with two separate primer pairs to generate overlapping cDNA fragments containing either EcoRV and Sad restriction sites. The mutagenic primers used for these amplification reactions were 5′-ggatatcactgcagagagct-3′ (antisense primer) and 3′-cAAAGTTGGATACATCTTAGC-5′ (sense primer). The regions corresponding to the RGP sequence are shown in capital letters, and the EcoRV and Sad sites in the primer tails are underlined. Those cDNA fragments were combined in a third polymerase chain reaction. The resulting cDNA fragment was digested with HindIII and XhoI and ligated into the corresponding restriction sites in pRGP(1–) to generate the plasmid pRGP(1–)IES. In that plasmid, the cDNA encoding amino acids 32–46 of RGP(1–) is replaced with a 20-base pair sequence containing the EcoRV and Sad restriction sites (Fig. 2, A and B). The polymerase chain reaction-derived HindIII–XhoI region of pRGP(1–)IES was sequenced to confirm successful mutagenesis. This plasmid was digested with EcoRV and Sad restriction enzymes and gel-purified to generate a vector for oligonucleotide cassette mutagenesis at sequon 1 (Fig. 2C).

Construction of Plasmids Encoding Variants of RGP(1–) with Amino Acid Substitutions at the X Position of Sequon 1—A cassette mutagenesis approach was used to generate plasmids encoding variants of RGP(1–) with amino acid substitutions at the X position of sequon 1. For this construction, sense and antisense oligonucleotides were synthesized, which were complementary to one another except at the nucleotide positions corresponding to the codon for amino acid 38 (Fig. 2D). To introduce the full spectrum of amino acid substitutions at that position, each oligonucleotide was completely degenerate at all

*This work was supported in part by funds from the National Blood Foundation. 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.

‡To whom correspondence should be addressed: Inst. of Pathology, Case Western Reserve University, School of Medicine, 2085 Adelbert Rd., Cleveland, OH 44106. Tel.: 216-844-8006; Fax: 216-844-1810.

§ The abbreviations used are: RGP, rabies virus glycoprotein; CGE, core-glycosylation efficiency.
Role of the X Amino Acid in N-Linked Core Glycosylation

A

B

C

D

E

F

G

H

I

J

K

L

M

N

O

P

Q

R

S

T

U

V

W

X

Y

Z

RESULTS

Construction of Plasmids Encoding RGP Variants with Amino Acid Substitutions at Position 38—The variants used in this study were derived from RGP. Wild type RGP is a 505-amino acid type 1 transmembrane protein with a 439-amino acid extracellular domain and a 44-amino acid cytoplasmic tail (Fig. 1, RGP(WT)) (32). The extracellular domain has three sequences for N-linked glycosylation at Asn37 (sequon 1), Asn247 (sequon 2), and Asn319 (sequon 3) (32). To study the core glycosylation of individual sequons in RGP, sequons 2 and 3 in RGP(WT) were deleted by site-directed mutagenesis by substituting Thr at each Asn-Pro sequence normally present in RGP for Asn-Pro with the sequence Asn-Pro-Arg corresponding to amino acid 38 (Fig. 2). Each thrombin substrate cDNA (RGP(1–)–X38) variant containing the sequence Asn-Pro-Arg was ligated into the RGP(1–)–ES cloning vector at the EcoRV and SacI restriction sites. Each plasmid resulting from this ligation encodes a protein identical to RGP(1–) except for the amino acid at the position corresponding to amino acid 38 (sense sequence). In Vivo Transcription and Expression in the Cell-free System—RNA encoding variants was generated by in vitro transcription with T7 RNA polymerase as described previously (8, 29). In vitro translation was performed using a rabbit reticulocyte lysate system supplemented with [35S]methionine and canine pancreas microsomes (Boehringer Mannheim) as described (29), except that the amount of microsomes was reduced to 1 mg/ml. The reactions were incubated on ice for 20 min in the presence of 200 μCi of [35S]methionine. The microsomes cleave the asparagine-X amino acid on core glycosylation. The amount of RGP incorporated into microsomes while maintaining adequate signal sequence of these proteins and add core oligosaccharides (8, 30). This cell-free system provides a highly reproducible method for quantifying the co-translational CGE of specific sequons in recombinant proteins (30). Analysis of core glycosylation in this system is simpler than in intact cells because oligosaccharide processing is limited and protein variants can be analyzed without immunoprecipitation. Also, unlike analysis in intact cells, alterations in protein stability or expression resulting from amino acid substitutions are uncommon. Our previous studies have demonstrated that the core-glycosylation efficiencies of RGP variants in this system are similar to those observed when the same proteins are expressed in transfected Chinese hamster ovary cells (8, 29).

RGP variants with each of the 20 common amino acids at the X position of sequon 1 were generated by oligonucleotide cassette mutagenesis. A, the DNA and amino acid sequences near sequon 1 in RGP(1–) are shown; the sequon is underlined. B, a cloning vector for cassette mutagenesis, pRGP(1–)–ES, was generated from pRGP(1–) by replacing the sequence for amino acids 32–46 with a novel sequence (upper case letters) containing EcoRV and SacI restriction sites (underlined). C, for cassette mutagenesis the cloning vector was digested with EcoRV and SacI restriction enzymes. D, an oligonucleotide duplex was generated for cassette mutagenesis. The oligonucleotides in the duplex are complementary to one another except at the bases corresponding to the X amino acid in sequon 1 (XXX); both oligonucleotides are fully degenerate at that position. E, ligation of the duplex (upper case letters) into the cloning vector (lower case letters) restores the amino acid coding sequence of RGP(1–), except at the position corresponding to the X amino acid in sequon 1 (underlined). At that position each plasmid encodes one of the 20 common amino acids.

Sad restriction sites on either side of sequon 1 (Fig. 2 and “Materials and Methods”). An oligonucleotide cassette mutagenesis approach was used to generate a set of plasmids encoding variants of RGP(1–) with each of the 20 common amino acids at the X position of sequon 1 (collectively referred to as RGP(1–)–X38 variants) (Fig. 1 and “Materials and Methods”). A variant with Leu at position 38 (corresponding to the sequence normally present in RGP) was among the variants isolated using that approach. DNA sequencing was performed to confirm that each RGP(1–)–X38 plasmid encoded a protein identical to RGP(1–) except for the amino acid at position 38. Expression of RGP Variants in a Cell-free System—Each RGP variant was expressed in a rabbit reticulocyte cell-free translation system supplemented with [35S]methionine and canine pancreas microsomes (33–35). The microsomes cleave the signal sequence of these proteins and add core oligosaccharides (8, 30). This cell-free system provides a highly reproducible method for quantifying the co-translational CGE of specific sequons in recombinant proteins (30). Analysis of core glycosylation in this system is simpler than in intact cells because oligosaccharide processing is limited and protein variants can be analyzed without immunoprecipitation. Also, unlike analysis in intact cells, alterations in protein stability or expression resulting from amino acid substitutions are uncommon. Our previous studies have demonstrated that the core-glycosylation efficiencies of RGP variants in this system are similar to those observed when the same proteins are expressed in transfected Chinese hamster ovary cells (8, 29).
Role of the X Amino Acid in N-Linked Core Glycosylation

FIG. 3. Core glycosylation of RGP variants in the cell-free system. RNA encoding each RGP variant was generated by in vitro transcription and expressed in the cell-free system in the presence of canine pancreas microsomes and [35S]methionine. Translation products were treated with proteinase K and analyzed by SDS-polyacrylamide gel electrophoresis and autoradiography. The amino acid at the X position of sequon 1 in each variant is indicated. The migration positions of the nonglycosylated protein (N) and the protein glycosylated with a single core oligosaccharide (G) are indicated.

FIG. 4. CGE of RGP variants with amino acid substitutions at the X position of sequon 1. The 20 RGP variants generated by cassette mutagenesis were analyzed in the cell-free system as described for Fig. 3 in three independent experiments. Gel autoradiographs from each experiment were exposed in the linear range and analyzed by densitometric scanning. The CGE of each variant was calculated as described in the text for each experiment, and the mean CGE ± 1 S.D. from the three experiments was determined (shown).

DISCUSSION

This report extends previous studies by providing the first comprehensive direct analysis of the impact of the X amino acids on CGE. We demonstrate that the CGE at an Asn-X-Ser sequon in RGP ranges from no glycosylation to full glycosylation, depending on which amino acid is present at the X position. This demonstrates that the X amino acid is an important determinant of CGE.

Because the structure and enzymatic mechanism of oligosaccharyltransferase are not well characterized, currently it is not possible to determine the mechanism by which individual amino acids influence core glycosylation. Several studies suggest that the spatial relationship of the Asn and Ser/Thr residues in a sequon may be critical for oligosaccharide transfer (37–43). Large hydrophilic amino acids (e.g., Trp, Leu, Phe, and Tyr) may inhibit core glycosylation by producing an unfavorable local protein conformation. In contrast, Gly, which is small and does not constrain protein conformation, is associated with efficient core glycosylation. Other factors also appear to be important. The negatively charged amino acids (Asp and Glu) inhibit glycosylation, whereas the positively charged amino acids (Lys, Arg, and His) are favorable. The charge of the X amino acid may influence the ability of oligosaccharyltransferase to bind simultaneously to the sequon and the negatively charged dolichol-PP-oligosaccharide precursor (41, 44). Interestingly, the X amino acids with hydroxy groups (Ser and Thr) and Cys are associated with highly efficient core glycosylation, whereas those with amide groups (Asn and Gln) are associated with suboptimal core glycosylation. Further characterization of

quite translational activity. Under these conditions each translation reaction contains a small amount of protein that is not targeted to microsomes (8, 30). These untargeted proteins are not glycosylated, retain the 19-amino acid N-terminal signal sequence, and migrate between the nonglycosylated and glycosylated forms of RGP synthesized on microsomes in our gel system (8, 30). Because these untargeted proteins can interfere with the quantification of CGE, they were removed from translation reactions by proteinase K digestion prior to gel analysis (36). The extracellular domain of RGP variants is translocated into the microsomal lumen during protein synthesis where it is protected from proteinase K digestion. In contrast, the 44-amino acid cytoplasmic tail (Fig. 1) remains outside of the microsome and is removed by this treatment. Removal of the cytoplasmic tail produces a small shift in the electrophoretic mobility of RGP proteins (data not shown) but does not interfere with the quantification of CGE. Following proteinase K treatment, radiolabeled translation products were analyzed directly (without immunoprecipitation) by gel electrophoresis and autoradiography. A gel autoradiograph showing the translation products of all 20 RGP variants is shown in Fig. 3. The positions of the nonglycosylated protein (N) and the protein glycosylated with a single core oligosaccharide (G) are shown. The total amount of protein produced in each translation can vary from tube to tube reflecting differences in the amount of RNA in each sample. For this reason glycosylation efficiency is determined by comparing the amounts of glycosylated and nonglycosylated protein produced in a single reaction for each variant.

To quantify the CGE at the sequon in each RGP variant, the variants were expressed in the cell-free system in three independent experiments and autoradiographs from each experiment were analyzed by densitometric scanning. The densities of bands representing glycosylated (G) and nonglycosylated (N) proteins were quantified for each variant, and the CGE was calculated as follows: G/(N + G) × 100% (30). The mean CGE ± 1 S.D. was then determined for each variant (Fig. 4). This analysis revealed that the CGE observed for each RGP variant was highly reproducible in this system.

The experiments presented demonstrate that the amino acid at the X position of an Asn-X-Ser sequon can have a profound effect on CGE. These studies confirm that the presence of Pro at the X position completely blocks glycosylation (15, 20, 27, 37). Also, consistent with our findings from earlier studies (8, 29, 30), these data demonstrate that the sequon AsnLeuLeu is glycosylated at an intermediate level (mean CGE = 43%). Remarkably, we find that substitution of Leu with Trp, Asp, or Gln dramatically reduces the efficiency of core glycosylation (mean CGE = 5, 19, and 24%, respectively), whereas substitution of Leu with other amino acids increases CGE to varying degrees (mean CGE ranges from Phe = 70% to Ser = 97%). These results provide the first direct demonstration that amino acids at the X position of an Asn-X-Ser/Thr sequon can influence the efficiency of co-translational core glycosylation.
oligosaccharyltransferase may help clarify the role that individual amino acids play in oligosaccharide formation. Interestingly, the sequon Asn-Cys-Trp (15), Asp (19), and Glu (27) are uncommon at the position in core-glycosylated sequons. Studies of synthetic peptides in membrane preparations also find an inhibitory effect of Asp at the X position (41, 44). The current report provides direct confirmation that Trp, Asp, and Glu at the X position inhibit core glycosylation. Interestingly, the sequon Asn-Cys-Se$^{38}$r (39) in RGP is fully core-glycosylated. The lack of core glycosylation at Asn-Cys-Ser/Thr sequons in other proteins may reflect the potential of certain Cys residues to participate in disulfide bonding (40, 45).

It is important to note that factors other than the X amino acid also influence core glycosylation. For example, the presence of Pro immediately following a sequon can inhibit core glycosylation (20, 37), and the presence of Thr rather than Ser at the hydroxy position favors efficient glycosylation (15, 29, 42). Our previous studies demonstrate that the inhibitory effect of Leu in the sequon Asn-Cys-Leu-Ser (38) in RGP can be overcome by replacing Ser with Thr (29). Studies are currently under way to compare the impact of other X amino acids in Asn-X-Thr versus Asn-X-Ser sequons. Core glycosylation can also be influenced by factors that influence the accessibility of a sequon to oligosaccharyltransferase, such as the position of the sequon in a protein (14, 20, 30, 46, 47) and the folding of the nascent protein chain (45, 48). Core glycosylation is clearly a complex process influenced by a variety of factors. Further characterization of the protein signals that regulate core glycosylation will enhance our understanding of glycoprotein expression and facilitate the design of novel recombinant glycoproteins for research and clinical applications.

Acknowledgments—We thank Joyce Entoft and James Eshleman for helpful discussions.

REFERENCES
1. Varki, A. (1993) Glycoconjugates 3, 97–130
2. Kornfeld, R., and Kornfeld, S. (1985) Annu. Rev. Biochem. 54, 631–664
3. West, C. M. (1986) Mol. Cell. Biochem. 72, 3–20
4. Rademacher, T. W., and Dwek, R. A. (1989) in Carbohydrate Recognition in Cellular Function (Bock, G., and Harnett, S., eds) pp. 241–256, John Wiley & Sons Ltd., Chichester, W. Sussex, UK
5. Shakin-Eshleman, S. H., and Spitalnik, S. L. (1993) Trends Glycosci. Glycotechnol. 5, 355–368
6. Opdenakker, G., Rudd, P. M., Ponting, C. P., and Dwek, R. A. (1993) FASEB J. 7, 1330–1337
7. Gallagher, P., Henneberry, J., Wilson, I., Sambrook, J., and Gething, M. J. (1988) J. Cell Biol. 107, 2059–2073
8. Shakin-Eshleman, S. H., Rixon, A. T., Eshleman, J. R., Wunner, W. H., and Spitalnik, S. L. (1992) J. Biol. Chem. 267, 10690–10698
9. Machamer, C. E., and Rose, J. K. (1988) J. Biol. Chem. 263, 5948–5954
10. Roberts, P. C., Garten, W., and Klenk, H.-D. (1993) J. Virol. 67, 3054–3060
11. Omura, H., Otsu, M., Yashinori, T., Tashiro, Y., and Kikuchi, M. (1992) Eur. J. Biochem. 210, 591–599
12. Hubbard, S. C., and Ivatt, R. J. (1981) Annu. Rev. Biochem. 50, 555–583
13. Kelleher, D. J., Kreibich, G., and Gilmore, R. (1992) Cell 69, 35–65
14. Nilsson, I. M., and von Heijne, G. (1993) J. Biol. Chem. 268, 5789–5801
15. Kaplan, H. A., Wépilly, J. K., and Lennarz, W. J. (1987) Biochem. Biophys. Acta 906, 163–173
16. Bergman, L. W., and Kuehl, W. M. (1977) Biochemistry 16, 4490–4497
17. Rothman, J. E., and Lodish, H. F. (1977) Nature 269, 775–780
18. Glabe, C. G., Hanover, J. A., and Lennarz, W. J. (1988) J. Biol. Chem. 259, 9236–9242
19. Marshall, R. D. (1974) Biochem. Soc. Symp. 40, 17–26
20. Gavel, Y., and von Heijne, G. (1990) Protein Eng. 3, 433–442
21. Plummer, T. H., and Hirs, C. H. W. (1964) J. Biol. Chem. 239, 2530–2538
22. Pohl, G., Kallstrom, M., Bergdorff, N., Wallen, P., and Jornvall, H. (1984) Biochemistry 23, 3701–3707
23. Thim, L., Björn, S., Christensen, M., Nicolaisen, E. M., Lund-Hansen, T., Pedersen, A. H., and Hedner, U. (1988) Biochemistry 27, 7785–7793
24. Stevenson, S. C., Wang, S., Deng, L., and Tall, A. R. (1993) Biochemistry 32, 5122–5126
25. Curling, E. M. A., Hayter, P. M., Baines, A. J., Bull, A. T., Gull, K., Strange, P. G., and Jenkins, N. (1990) Biochem. J. 272, 333–337
26. Inouye, S., Guo, Y., Ling, N., and Shimasaki, S. (1991) Biochem. Biophys. Res. Commun. 179, 352–358
27. Mononen, I., and Karjalainen, E. (1984) Biochim. Biophys. Acta 780, 364–367
28. Nakai, K., and Kanehisa, M. (1988) J. Biochem. (Tokyo) 104, 693–699
29. Kasturi, L., Eshleman, J. R., Wunner, W. H., and Shakin-Eshleman, S. H. (1995) J. Biol. Chem. 270, 14756–14761
30. Shakin-Eshleman, S. H., Wunner, W. H., and Spitalnik, S. L. (1993) Biochemistry 32, 9465–9472
31. Ho, S. N., Hunt, H. D., Horton, R. M., Pullen, J. K., and Pease, L. R. (1989) Gene (Amst.) 77, 51–59
32. Wunner, W. H., Larson, J. K., Dietzschold, B., and Smith, C. L. (1988) Rev. Infect. Dis. 10, Suppl. 4, 771–784
33. Pélham, H. R. B., and Jackson, R. J. (1976) Eur. J. Biochem. 67, 247–256
34. Scheele, G. (1983) Methods Enzymol. 96, 94–111
35. Jackson, R. J., and Hunt, T. (1983) Methods Enzymol. 96, 50–74
36. Perera, E., and Lingappa, V. R. (1985) J. Cell Biol. 101, 2292–2301
37. Bause, E. (1983) Biochem. J. 209, 331–336
38. Bause, E., Hettkamp, H., and Legler, G. (1982) Biochem. J. 203, 761–768
39. Bause, E., and Hettkamp, H., and Legler, G. (1982) Biochemistry 21, 4374–4380
40. Bause, E., and Hettkamp, H. (1979) FEBS Lett. 108, 341–344
41. Bulleid, N. J., Bassil-Duby, R. S., Freedman, R. B., Sambrook, J. F., and Gething, M. J. (1992) Biochem. J. 286, 279–280
42. Bause, E., and Hettkamp, H. (1979) FEBS Lett. 108, 341–344
43. Livi, G. P., Lillquist, J. S., Miles, L. M., Ferrara, A., Sathe, G. M., Simon, P. L., Meyers, C. A., Gorman, J. A., and Young, P. R. (1991) J. Biol. Chem. 266, 15348–15355
44. Landolt-Marticorena, C., and Rethmeier, R. A. F. (1994) Biochem. J. 302, 253–260
45. Allen, S., Naim, H. Y., and Bulled, N. J. (1995) J. Biol. Chem. 270, 4797–4804
The Amino Acid at the X Position of an Asn-X-Ser Sequon Is an Important Determinant of N-Linked Core-glycosylation Efficiency
Susan H. Shakun-Eshleman, Steven L. Spitalnik and Lakshmi Kasturi

J. Biol. Chem. 1996, 271:6363-6366.
doi: 10.1074/jbc.271.11.6363

Access the most updated version of this article at http://www.jbc.org/content/271/11/6363

Alerts:
  • When this article is cited
  • When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts

This article cites 47 references, 17 of which can be accessed free at http://www.jbc.org/content/271/11/6363.full.html#ref-list-1