Asparagine-linked oligosaccharides are localized to single extracytosolic segments in multi-span membrane glycoproteins

Carolina LANDOLT-MARTICORENA and Reinhart A. F. REITHMEIER*

MRC Group in Membrane Biology, Departments of Medicine and Biochemistry, University of Toronto, Toronto, Ontario, Canada, M5S 1A8

A comprehensive survey of mammalian multi-span (polytopic) membrane proteins showed that asparagine(N)-linked oligosaccharides are localized to single extracytosolic segments. In most membrane proteins this is because potential consensus sites for N-glycosylation (Asn-Xaa-Ser/Thr, X ≠ Pro) are not found in multiple extracytosolic segments. In functional proteins where consensus N-glycosylation sites are contained within more than one extracytosolic segment, only the first segment contains N-linked carbohydrate. An exception is the α -subunit of the Na⁺ channel, which consists of a duplicated structure containing two glycosylated segments. The average size of established N-glycosylated loops connecting two transmembrane segments is 62 residues, with the smallest glycosylated loop being 33 residues in size. N-glycosylated sites are more highly conserved than non-

glycosylated (primarily cytosolic) sites and are more common toward the N-terminus of the membrane domain of multi-span membrane proteins. The optimal conditions for glycosylation of consensus sites within an extracytosolic domain of a multi-span membrane protein are (i) the acceptor site is well-spaced (greater than 10 residues) from the transmembrane domain, (ii) the loop is greater than 30 residues in size and (iii) the segment is the first in the protein to contain a suitable extracytosolic consensus site. The localization of N-linked oligosaccharide chains to a single protein segment suggests either glycosylation of multiple loops may compromise protein folding or function, or only a single polypeptide domain can be optimally glycosylated during biosynthesis *in vivo*.

INTRODUCTION

N-glycosylation is a co-translational event in which a high-mannose oligosaccharide structure is transferred *en bloc* from a dolichol donor to the nascent polypeptide (Kornfeld & Kornfeld, 1985). This transfer is mediated by an oligosaccharyltransferase localized to the ER lumen (Kaplan et al., 1987). A protein complex containing the oligosaccharyltransferase activity and ribophorins I and II has been isolated, and a putative dolichol-binding site has been identified in ribophorin I (Kelleher et al., 1992). Evidence suggests that the oligosaccharyltransferase forms part of the protein processing complex and is closely associated with the protein translocation channel (translocon) (Kaplan et al., 1987; Simon & Blobel, 1991). Oligosaccharide chains are attached to asparagine residues within the consensus sequence Asn-Xaa-Ser/Thr, where Xaa is any amino acid other than proline (Kornfeld & Kornfeld, 1985). In addition, a consensus sequence followed by a proline residue is unlikely to be glycosylated (Gavel & von Heijne, 1990; Roitsch & Lehle, 1989).

Multi-span membrane proteins constitute an important class of glycoproteins that function primarily as receptors, channels or transporters (Reithmeier & Deber, 1992). The overwhelming majority of multi-spanning membrane proteins are N-glycosylated, exceptions being the α -subunits of cation-transporting ATPases and the α_1 -subunit of the Ca²⁺ channel, which are tightly associated with N-glycosylated subunits. Although the presence of a consensus glycosylation site within the amino acid sequence of a membrane protein is suggestive of a glycoprotein, not all consensus sites are necessarily utilized. Sites must be disposed to the luminal side of the endoplasmic reticulum membrane to be glycosylated, and not all luminal sites may be suitable acceptors. As such, N-linked oligosaccharide

chains are useful topological markers for the extracytosolic side of membrane proteins. A role for N-glycosylation has been implicated in directing protein folding, regulating cell-surface expression or increasing the half-life of proteins (Paulson, 1989). It has been demonstrated for several membrane proteins that the N-linked oligosaccharide is not required for protein function [Band 3 (Casey et al., 1992); muscarinic acetylcholine receptor (van Koppen & Nathanson, 1990); cyclic GMP-gated channel from rod photoreceptors (Wohlfart et al., 1989); human cystic-fibrosis transmembrane conductance regulator (CFTR) (Gregory et al., 1991); Na⁺/Ca²⁺ cardiac exchanger (Hryshko et al., 1993); Na⁺-dependent glucose carrier (Hediger et al., 1991; Hirayama & Wright, 1992)], suggesting a general biosynthetic role or a function distinct from substrate transport.

RESULTS AND DISCUSSION

In order to understand the nature and specificity of the N-glycosylation process, a survey of glycosylation sites in an extensive population of multi-span membrane proteins was carried out. The survey includes about 300 multispan membrane glycoproteins, the majority of which (approx. 230) are members of the G-protein-coupled (TM7) receptor family. Table 1 lists mammalian multi-span membrane proteins in which the utilization of specific extracytosolic sites has been demonstrated experimentally. The protein database in Table 1 consists of over 100 different membrane proteins which are grouped into functional families. The proteins are divided into three groups based on their distribution of extracytosolic consensus N-glycosylation sites. The first group (Table 1; G1) consists of 106 proteins that contain consensus sites only in a single extracytosolic segment, thereby restricting glycosylation to this domain. The second

Abbreviations used: AE, anion-exchanger; TM, transmembrane; CFTR, cystic-fibrosis transmembrane conductance regulator; SRP, signal-recognition particle.

*To whom all correspondence should be sent, at the following address: MRC Group in Membrane Biology, Department of Medicine, Room 7344, Medical Sciences Building, University of Toronto, Toronto, Ontario, Canada, M5S 1A8.

Table 1 Sites of asparagine-linked glycosylation in mammalian multi-span membrane proteins

	Protein†	Consensus sites‡	Location§	Conserved	Utilized¶	Experimental evidence**	Reference
GI*	Glucose transporter (GLUT1) {GTR1\$Human}	N45 N411	TM1/2 TM11	5/5 (14/14) 5/5 (12/14)	+	(f)	Asano et al. (1991)
	Erythroid anion transporter (Band 3/AE1) {B3AT\$Human}	N593 N642	TM6/7 TM7/8	3/5 (3/11) 4/5 (4/11)	— +	(c)	Drickamer (1978)
	Erythroid anion transporter (Band 3/AE1) {B3AT\$Trout}	N545 N568	TM5/6 TM5/6	1/5 (5/11) 1/5 (5/11)	+	(c)	Hübner et al. (1992)
	Anion transporter (AE2) {B3AR\$Human}	N441 N858 N867 N881	Nt TM5/6 TM5/6 TM5/6	4/4 (4/11) 4/4 (5/11) 4/4 (7/11) 4/4 (5/11)	— (+) (+) (+)	(b)	Lindsey et al. (1990)
	Anion transporter (AE3) {B3A3\$Mouse}	N868	TM5/6	2/2 (6/11)	+	(a)	Kobayashi et al. (1994)
	Rod photoreceptor Na ⁺ /Ca ⁺ —K ⁺ exchanger {unavailable}	N88 N290 N303	Nt Nt Nt	1/1 1/1 1/1	(+) (+) (+)	(d)	Reid et al. (1990)
		N546 N648 N782	TM3/4 TM5/6 TM5/6	1/1 1/1 1/1	— — —	(f)	Reiländer et al. (1992)
	Na ⁺ -dependent glucose carrier {GTNA\$Rabit}	N248 N306	TM5/6 TM6/7	2/2 (4/4) 2/2 (3/4)	+	(f)	Hediger et al. (1991)
	β2 Adrenergic receptor {B2AR\$Mesau}	N6 N15 N410 N414	Nt Nt Ct Ct	4/4 (4/10) 4/4 (7/10) 4/4 (4/10) 4/4 (4/10)	+	(f)	Rands et al. (1990)
	Muscarinic receptor {ACM2\$Rat}	N2 N3 N6 N108 N284 N432 N444	Nt Nt Nt TM3 TM5/6 TM7 Ct	3/3 (7/15) 3/3 (6/15) 3/3 (6/15) 3/3 (10/15) 3/3 (5/15) 3/3 (5/15) 3/3 (5/15)	(+) (+) (+)	(f)	van Koppen and Nathanson (1990)
	P-glycoprotein (PGP) {MDR1\$Human}	N21 N91 N94 N99 N296 N494 N656 N704 N809 N1034	Nt TM1/2 TM1/2 TM1/2 TM4/5 TM6/7 TM6/7 TM6/7 TM8/9 Ct	1/2 (1/5) 2/2 (3/5) 2/2 (3/5) 2/2 (3/5) 1/2 (3/5) 1/2 (3/5) 1/2 (1/5) 2/2 (4/5) 2/2 (4/5) 1/2 (2/5)	— + + +	(f)	Schinkel et al. (1993)
	CFTR {CFTR\$Human}	N48 N306 N386 N396 N432 N597 N801 N894 N900 N1148 N1229 N1454	Nt TM4/5 TM6/7 TM6/7 TM6/7 TM6/7 TM6/7 TM7/8 TM7/8 TM12 Ct Ct	2/4 2/4 3/4 4/4 4/4 4/4 2/4 4/4 4/4 4/4 4/4 2/4	— — — — — — — (+) (+)	(f)	Cheng et al. (1990)
	Dopamine transporter {NTTD\$Rat}	N44 N181 N188 N196 N204	Nt TM3/4 TM3/4 TM3/4 TM3/4	2/2 2/2 1/2 2/2 2/2	— + + + +	(f)	G. Uhl, personal communication

Table 1 (cont.)

Protein†	Consensus sites‡	Location§	Conserved	Utilized¶	Experimental evidence**	Reference
Rat brain Na ⁺ /L-glutamate co-transporter {unavailable}	N205 N215	TM3/4 TM3/4	3/3 2/3	(+) (+)	(a)	Danbolt et al. (1992)
Acetylcholine-receptor α-subunit {ACHAS\$Torca}	N142	Nt	5/5 (30/30)	+	(a,d)	Nomoto et al. (1986)
Acetylcholine-receptor γ-subunit (ACHG\$Torca)	N68 N141 N306 N431	Nt Nt TM3/4 TM3/4	1/5 (1/30) 5/5 (30/30) 5/5 5/30 2/5 (2/30)	++ ++ — —	(a,d) (c)	Nomoto et al. (1986); Gehle and Sumikawa (1991)
K ⁺ -channel {CIK3\$Human}	N57 N227 N469 N486 N499	Nt TM1/2 Ct Ct Ct	3/3 3/3 3/3 3/3 3/3	— (+) — — —	(a)	Roseberg and East (1992)
Rod photoreceptor cyclic GMP gated channel {unavailable}	N90 N91 N177 N327 N423	Nt Nt TM1 TM5/6 TM5/6	2/2 (2/3) 2/2 (2/3) 2/2 (2/3) 2/2 (2/3) 2/2 (2/3)	— — — ++ —	(a,d) (e)	Wohlfart et al. (1989) Wohlfart et al. (1992)
GII* Cardiac Na ⁺ /Ca ²⁺ exchanger {NACAS\$Canfa}	N41 N128 N133 N157 N656 N866	Nt TM1/2 TM1/2 TM2/3 TM5/6 TM8/9	2/2 2/2 2/2 2/2 2/2 2/2	++ — — — — —	(f)	Hryshko et al. (1993)
Na ⁺ /H ⁺ exchanger (NHE-1) {NAHAS\$Human}	N75 N370 N410	TM1/2 TM9/10 TM10/11	5/6 (6/10) 6/6 (10/10) 6/6 (7/10)	++ — —	(f)	Counillon et al. (1994)
Rhodopsin {OPSD\$Bovin}	N2 N15 N200	Nt Nt TM4/5	9/12 (10/23) 9/12 (13/23) 9/12 (11/23)	++ ++ —	(a) (f)	Fukuda et al. (1979) Doi et al. (1990)
CHIP28 {WCH1\$Human}	N42 N205	TM1/2 TM5/6	2/2 2/2	++ —	(c)	Preston et al. (1994)
GIII* Skeletal-muscle Na ⁺ -channel α-subunit {CINSS\$Rat}	N214 N288 N291 N297 N303 N309 N315 N327 N356 N502 N692 N954 N1184 N1198	TM4 (DI)†† TM5/6 (DI) TM5/6 (DI) TM5/6 (DI) TM5/6 (DI) TM5/6 (DI) TM5/6 (DI) TM5/6 (DI) TM5/6 (DI) TM6/1 (DI/DII) TM3 (DII) TM6/1 (DII/DIII) TM5/6 (DIII) TM5/6 (DIII)	5/5 4/5 1/5 1/5 3/5 4/5 5/5 1/5 5/5 1/5 5/5 1/5 5/5 5/5	— (+) (+) (+) (+) (+) (+) (+) (+) — — — ++ ++	(d) (c)	Kraner et al. (1989) Maceyka et al. (1993)

* GI, proteins with consensus glycosylation sites in only a single extracytosolic domain; GII, proteins containing potential glycosylation sites in multiple extracytosolic domains in which only one segment is glycosylated; GIII, proteins containing potential glycosylation sites in multiple extracytosolic domains in which more than one segment is glycosylated.

† Proteins are listed in homologous gene families with the Swissprotein designation in face ('curly') brackets (where available) and common designation in parentheses. The Swissprotein Data bank was searched with keywords (i.e. 'glycoprotein', 'membrane' and 'transporters') using the Retrieve NIH program to generate a protein database restricted to eukaryotic, multi-spanning, membrane proteins. Additional protein sequences not present in the Swissprotein database were taken from the literature. Isoforms or related family members were included in the database, since they often differed in their glycosylation pattern. A final list of approx. 250 proteins was used for all subsequent analyses.

‡ Location of consensus glycosylation sites (Asn-Xaa-Thr/Ser, Xaa ≠ Pro) in the protein. Residue numbers given are for the primary translation product. Potential sites of glycosylation were identified using the IntelliGenetics Inc. program GeneWorks under a search for consensus glycosylation sites (Asn-Xaa-Ser/Thr, Xaa ≠ Pro). Extracellular consensus sites are in **bold** text.

§ The location of the consensus glycosylation sites with respect to transmembrane (TM) segments. Nt, in N-terminal domain; Ct, in C-terminal domain; TMx/y, between transmembrane segments x and y; TMz, within transmembrane segment z. Current published topological folding models which incorporate available experimental evidence were used to determine the orientation of the consensus sequences with respect to the lipid bilayer (see Figure 1). Putative transmembrane segments within membrane proteins are defined as a hydrophobic amino acid sequence of sufficient length (21 amino acids) to span the lipid bilayer.

|| Conservation of the consensus site in the same gene products in different species. Ratios in parenthesis represent the conservation in all homologous isoforms within a gene family.

group (Table 1; GII) lists proteins containing multiple extracytosolic consensus sites in which it has been shown that only one segment is N-glycosylated. The third group (Table 1; GIII) is composed of only one protein, the α -subunit of the Na^+ channel, which contains two extracytosolic N-glycosylated segments. Experimental evidence and corresponding references for the utilization of consensus sites are summarized in Table 1. Evidence for N-glycosylation that was solely based on putative folding models was not considered. Figure 1 shows models for the topology of the various membrane proteins listed in Table 1 and the locations and utilization of all consensus N-glycosylation sites.

Localization of N-linked oligosaccharide

The survey indicates that N-linked oligosaccharides are localized to single extracytosolic segments in mammalian multi-span glycoproteins, the only exception being the α -subunit of the Na^+ channel. The vast majority of the surveyed proteins fall into the first group (Table 1, GI). In these proteins N-linked oligosaccharide is restricted to a single segment, because consensus acceptor sites are not found on multiple extracytosolic segments. The location of N-glycosylation sites within the anion-exchanger (AE) family supports this contention. Mammalian non-erythroid anion exchangers (AE2 and AE3) and trout erythroid AE1 have a single N-glycosylated loop between the fifth and sixth membrane-spanning segments (Figure 1). This loop contains about 50 residues and is the largest extracellular segment in these proteins. In contrast, other erythroid AE1 forms are N-glycosylated at a single site in the subsequent extracellular loop, which is the largest extracellular domain (~ 42 amino acids) in these AE1 forms. This acceptor site, in trout AE1 and in mammalian AE2 and AE3, has been negated by the presence of a Pro at the Xaa position of the consensus site and, in the case of AE2 and AE3, the absence of the acceptor Asn residue. The compensatory switch of consensus N-glycosylation sites between the third and fourth extracellular domain in AE isoforms supports the model that utilized N-glycosylation sites are restricted to a single extracellular loop.

Four protein families (Table 1, GII) contain members (21/37) which have consensus sites in multiple extracellular domains. The glycosylation pattern of four members of these families ($\text{Na}^+/\text{Ca}^{2+}$ exchanger, NHE-1, rhodopsin and CHIP28) have been determined. In all of these proteins N-linked oligosaccharide is only found on the first acceptor domain. The non-glycosylated sites are either located close to a transmembrane (TM) segment or are contained within a short (< 30 residue) loop.

In canine cardiac $\text{Na}^+/\text{Ca}^{2+}$ exchanger only the consensus site on the first domain is utilized despite the availability of additional sites on two downstream extracellular loops (Figure 1). Both downstream sites are predicted to be viable acceptor sequences (N¹⁵⁷FT and N⁸⁶⁶VT) (Gavel & von Heijne, 1990), located on extracellular loops 11 and 31 residues in size respectively. The observation that neither is utilized argues that translocation of a potential acceptor site to the endoplasmic-reticulum lumen may not be sufficient for transfer of the core oligosaccharide to the

nascent polypeptide. The second loop in the $\text{Na}^+/\text{Ca}^{2+}$ exchanger may be too small to be utilized efficiently, and the acceptor site in the penultimate loop is only eight residues from the bilayer, perhaps too close to be utilized (see below). The three sites are conserved in the human $\text{Na}^+/\text{Ca}^{2+}$ exchanger (Table 1), and a similar pattern of N-glycosylation is predicted.

In the human Na^+/H^+ exchanger (NHE-1) the consensus site at Asn⁷⁵ is the only one utilized (Counillon et al., 1994), although this protein contains an additional downstream extracytosolic consensus site (Asn³⁷⁰). The N-glycosylation site at Asn⁷⁵ is conserved in rat, hamster, rabbit and pig NHE-1, whereas trout NHE-1 has an alternative site (Asn⁴⁹) localized to the same extracellular loop between TM segments 1 and 2. The site equivalent to Asn⁷⁵ is not conserved in NHE-2, -3 and -4, while Asn³⁷⁰ is found in all isoforms of the protein (NHE-1, -2, -3 and -4). The non-glycosylated loop in human NHE-1 (containing Asn³⁷⁰) between TM segments 9 and 10 is only 20 residues in length (Figure 1). The average size of the loop containing Asn³⁷⁰ is 23 residues and, owing to its small size (see below), it is predicted not to be utilized in any of the exchangers. Indeed, the rat NHE-3 protein has been shown to lack N-linked oligosaccharides (Counillon et al., 1994).

In bovine rhodopsin, only Asn² and Asn¹⁵, localized to the N-terminus, are utilized. The non-glycosylated loop (containing Asn²⁰⁰) between TM segments 4 and 5 is 29 residues in length (Figure 1). Asn²⁰⁰ is conserved in only 11 of the 23 rhodopsin isoforms surveyed; 12 of the proteins surveyed therefore lack a consensus N-glycosylation site in this loop and actually fall into Group I. On the basis of bovine rhodopsin, it is unlikely that the distal acceptor sites are utilized in any of the rhodopsin proteins, since they are also localized to small extracellular loops (29 residues in average for all known isoforms).

The presence of N-linked glycosylation sites within the N-terminal domain is a widely conserved property of the TM7 family, with 92% (211/229) of the surveyed proteins exhibiting this feature. Only one-third of the members of this family, like bovine rhodopsin (e.g. human β_2 adrenergic receptor, rat D3 dopamine receptor and rat Substance P receptor) contain additional consensus sites on extracytosolic loops. The glycosylation status of these sites has yet to be determined experimentally. On the basis of the small size of the loops and the data on rhodopsin, these loops are predicted to be poor acceptor sites.

Recent work (Preston et al., 1994) on the N-glycosylation pattern of CHIP28, the water channel of the erythrocyte membrane, has shown that the first exofacial loop (~ 20 residues) contains a single N-glycosylated site located about seven residues from the predicted end of the first TM segment. This loop is, however, poorly glycosylated, with only one-quarter of the mature protein containing N-linked oligosaccharide (Nielsen et al., 1993). A second, larger loop which contains a consensus site (Figure 1) is not N-glycosylated. This suggests that loops distal to a glycosylated segment may not be readily accessible to the oligosaccharyltransferase.

Proteins that are composed of duplicated domains in which the topological features of the N-terminal half are reproduced in

¶ + Indicates an established glycosylation site; — Indicates lack of glycosylation; (+) indicates that the potential glycosylation site is likely to be utilized, since it is localized to a domain shown to be glycosylated.

** Experimental evidence for the glycosylation of specific consensus sequences was obtained from the literature and includes the following types: (a) a shift in the electrophoretic mobility of a protein in SDS-containing gels upon enzymic deglycosylation; (b) inhibition of glycosylation by drugs; (c) specific consensus sequences localized directly by proteolytic cleavage and protein sequencing; (d) lectin binding and protein-fragmentation studies; (e) antibodies directed against synthetic peptides corresponding to putative glycosylated sites in combination with enzymic deglycosylation; (f) site-directed mutagenesis and truncation experiments which delete potential carbohydrate-attachment sites.

†† DI, DII, DIII and DIV correspond to the first, second, third and fourth homologous domains respectively of the α -subunit of the Na^+ channel. The utilization of consensus sites N1184 and N1198 was determined for the *Electrophorus electricus* Na^+ channel. These sites are conserved in the skeletal-muscle Na^+ channel and are assumed to be utilized.

Figure 1 Proposed folding models for the membrane glycoproteins listed in Table 1

The two horizontal broken lines indicate the biliary; above the upper line is extracytosolic and below the lower line is cytosolic. All the potential N-glycosylation sites in each protein are illustrated by circles. Closed circles indicate sites which are utilized. Hatched circles indicate sites which are presumed to be utilized, since they are localized to known N-glycosylated domains. Open circles indicate sites which are not utilized. Numbers listed above certain extramembranous loops indicate the estimated size of the loop residues. References: ¹Pessin and Bell (1992); ²Alper (1991); ³Jennings (1992); ⁴Hübner et al. (1992); ⁵Wright et al. (1992); ⁶Ostrowski et al. (1992); ⁷Yoshimura et al. (1993); ⁸Riordan et al. (1989); ⁹Schinkel et al. (1989); ¹⁰Cheng et al. (1990); ¹¹Uhl, (1992); ¹²Amara (1992); ¹³DeLorey and Olsen (1992); ¹⁴Catterall (1988); ¹⁵Wohlfart et al. (1992); ¹⁶Counillon et al. (1992); ¹⁷Counillon and Pouyssegur (1993); ¹⁸Phillipson and Nicoll (1992); ¹⁹Khorana (1992); ²⁰P. Agre, personal communication.

the C-terminal half (MDR1, CFTR and the α -subunit of the Na⁺ channel) might also be expected to duplicate their N-glycosylation pattern. Surprisingly, the restriction of utilized N-glycosylation sites to one extracytosolic loop extends to proteins with a duplicated structure (e.g. MDR1 and CFTR). For example, of the ten N-glycosylation consensus sites in human MDR1, the three sites which are predicted to be extracellularly disposed are clustered in the first extracellular loop of the N-terminal domain (Figure 1). This glycosylation feature is not matched in the first loop of the second domain of MDR1. Recent work reporting the independent expression of both the N- and C-terminal half molecules of MDR1 demonstrated that only the first half domain is glycosylated in an *in vivo* system (Loo & Clarke, 1994). Mutagenesis studies have demonstrated that all three consensus sites localized to the first 47 residue N-terminal domain are utilized (Schinkel et al., 1993). When the size of the extracellular loop was reduced by 20 residues with the associated removal of two of the three sites of N-glycosylation, utilization of the remaining site was prevented (Schinkel et al., 1993). This result suggests that the size of the extracellular loop is critical for utilization of viable acceptor sites which otherwise fulfill all other criteria for N-glycosylation. The α -subunit of the Na⁺ channel is the sole example to date of a functional protein that is N-glycosylated on two extracellular loops (Table 1). These loops, however, are homologous, residing in duplicated domains I and III (Figure 1). This suggests that when loops are located in independently folded domains, both can be N-glycosylated.

Size of glycosylated loops

N-linked carbohydrate is commonly localized to the largest extracytosolic segment of multi-span membrane proteins; the loops have an average size of 62 residues, with most loops in the range of 40–80 residues (Table 2). N-glycosylated extracytosolic loops in the surveyed proteins have a minimum size of approx. 33 residues (Asn⁴⁵; human GLUT1 glucose transporter). In CFTR the two utilized sites (Asn⁸⁹⁴ and Asn⁹⁰⁰) are localized to a 38-residue loop. As demonstrated for MDR1, a viable acceptor site is no longer utilized when the size of the extracellular domain is reduced to 27 residues, and the acceptor site is moved six residues distal to the bilayer. The reduced size of the acceptor loop agrees well with the threshold size obtained from the documented survey data in Table 2. Small loops (< 30 residues) containing consensus N-glycosylation sites are likely inefficient acceptors of oligosaccharide and are predicted to be poorly N-glycosylated, as in the case of CHIP28.

Distance of glycosylation sites from the TM segment

In multi-spanning membrane proteins with an extracytosolic N-terminal domain (e.g. rhodopsin; Figure 1) N-glycosylation consensus sites are found as close as two residues from the first TM segment, although the two closest *utilized* sites are located 10 (muscarinic acetylcholine receptor α -subunit) and 17 (β 2 adrenergic receptor) residues away from the bilayer. N-glycosylation can even occur at the second residue of an N-terminal extracytosolic segment (e.g. rhodopsin) as long as the site is well-spaced from the TM domain. In internal loops, utilized acceptor sites are found as close as 12 residues upstream (Asn⁹⁰⁰; CFTR) or distal (Asn⁴⁵; human GLUT1) from a TM segment.

A survey (not shown) of 115 human single-span Type I membrane proteins (N-terminus extracellular) yielded compar-

Table 2 Loop sizes of membrane proteins

Protein	Loop size*
Skeletal-muscle Na ⁺ channel	142
DI domain (rat)	
Na ⁺ /glutamate transporter (rat)	98
Skeletal-muscle Na ⁺ channel	87
DIII domain (rat)	
Neurotransmitter transporters†	77 (range 66–89; <i>n</i> = 6)
Na ⁺ /H ⁺ transporter (human)	73
B3LP (AE2) (human)	53
B3AT (AE1) (trout)	49
MDR1 (human)	47
B3AT (AE1) (human)	42
Na ⁺ /glucose co-transporter (human)	42
Facilitative glucose transporters†	42 (range 33–68; <i>n</i> = 5)
K ⁺ channel (CIK3, human)	41
B3A3 (AE3) (mouse)	40
CFTR (human)	38
Average	62

* Predicted size of glycosylated extracellular domains in descending order of size. The size of the extramembranous loops were defined as the number of amino acids between consecutive hydrophobic membrane-spanning segments. Since the boundaries of the TM segments are often not clearly defined, the size of the glycosylated loops are estimates. In this survey, published topological models were utilized to select the TM domains (see Figure 1).

† Indicates that loop size is an average of known isoforms. Identical gene products from various species were not included when calculating the average loop size.

able results. The closest *utilized* consensus site is located 16 residues (leukosialin) from the membrane domain, although potential sites are as close as two residues from the bilayer (hepatocyte growth-factor receptor). These results suggest that a minimum distance of ten residues from the bilayer is required for successful N-glycosylation of acceptor sites which precede a membrane-spanning segment. A survey (not shown) of about 20 single-span Type II (C-terminus extracellular) showed that the closest potential acceptor site is 11 residues (macrophage low-density-lipoprotein receptor) from the membrane though the nearest known utilized site is 18 residues (Fc epsilon receptor) from the bilayer.

Nilsson and von Heijne (1993) reported experimental findings that a minimum spacing of 14–15 residues is necessary for the utilization of an extracytosolic consensus site which precedes the first TM segment in a two-span membrane protein. An extracytosolic consensus site C-terminal to the second TM segment in the same protein requires a minimum distance of 12–14 residues in order to be N-glycosylated (Nilsson & von Heijne, 1993). Although Nilsson and von Heijne (1993) did not discuss a limit size for polypeptide loops flanked by membrane-spanning segments, their results suggest that a loop must exceed 28 residues to be N-glycosylated. These results are consistent with our observations. The sharp threshold observed in the size of known N-glycosylated domains (Table 2), as well as the results reported by Nilsson and von Heijne (1993) suggest that the oligosaccharyltransferase active site resides a fixed distance from the bilayer. The protein loop serves as a scaffolding for the N-glycosylation consensus sequence, breaching the distance from the bilayer to the oligosaccharyltransferase active site.

Location of glycosylation sites

A feature common to most of the proteins examined is the paucity of N-glycosylated sites in the C-terminal half of the membrane domain of the polypeptide (except CFTR and rod photoreceptor cyclic GMP-gated channel; Figure 1). For example, many proteins contain N-glycosylated sites in the N-terminal domain (cardiac $\text{Na}^+/\text{Ca}^{2+}$ exchanger, rod $\text{Na}^+/\text{Ca}^{2+}$ - K^+ exchanger, rhodopsin, β_2 -adrenergic receptor, muscarinic receptor, acetylcholine receptor and K^+ channel; Figure 1) or in the first extracytosolic loop between TM segments 1 and 2 (GLUT1, NHE-1, MDR1 and CHIP28; Figure 1). The clustering of N-glycosylated domains to the N-terminal half of the proteins suggests a general biosynthetic role for the carbohydrate moiety. It has been proposed (Paulson, 1989) that oligosaccharide structures may serve as nucleating points for protein folding, fixing nascent polypeptides into a semi-native conformation. In this sense the oligosaccharyltransferase may act as a chaperone, directing the folding process. Inhibiting the transfer of the core oligosaccharide or the introduction of novel N-glycosylation sites can result in the misfolding of membrane proteins and their accumulation in the endoplasmic reticulum (e.g. haemagglutinin (Gallagher et al., 1988)). Nevertheless, it is unlikely that N-glycosylation plays a critical role in the folding of all membrane proteins, since, in many cases [Band 3 (Casey et al., 1992); muscarinic acetylcholine receptor (van Koppen & Nathanson, 1990); cyclic GMP-gated channel from rod photoreceptors (Wohlfart et al., 1989); human CFTR (Gregory et al., 1991); $\text{Na}^+/\text{Ca}^{2+}$ cardiac exchanger (Hryshko et al., 1993); Na^+ -dependent glucose carrier (Hediger et al., 1991; Hirayama & Wright, 1992)] deglycosylation still yields a functional protein. An alternative explanation for the localization of N-glycosylated domains to the N-terminal region of membrane proteins is that partial folding of a polypeptide places physical constraints on the transfer of the sugar moiety to subsequent acceptor sites. Only consensus sites which are located on leading segments of a nascent protein are suitable acceptor sites for the core oligosaccharide. A survey of secreted and single-span membrane proteins indicates that non-glycosylated sites tend to be found more frequently towards the C-termini of glycoproteins (Gavel & von Heijne, 1990). Protein folding may therefore, in general, impair the N-glycosylation of distal sites.

Conservation of glycosylation sites

It was found that two-thirds of the utilized N-glycosylation sites in the proteins listed in Table 1 were conserved, whereas only one-half of the non-utilized (mainly cytosolic) sites were conserved. Although a specific site may not be conserved, the N-glycosylated nature of the loop is often retained (Table 1). The fact that N-glycosylated sites are more highly conserved than non-glycosylated sites suggests that the N-glycosylated domains play some important role. The lack of multiple N-glycosylated domains suggests that additional extracytosolic consensus sites have been lost during evolution. If only the first suitable extracytosolic loop of a polytopic membrane protein can be optimally N-glycosylated, there would not be any evolutionary pressure to conserve subsequent consensus sites. It is also possible that multiple glycosylated domains are detrimental to protein folding and function.

Biosynthetic considerations

The striking absence of multiple N-glycosylated domains in all but one of the surveyed polytopic membrane proteins suggests

that perhaps only one polypeptide segment can be N-glycosylated during protein synthesis *in vivo*. During the biosynthesis of multi-span membrane proteins it appears that the first TM segment binds to signal recognition particle (SRP) and directs the nascent chain to the endoplasmic-reticulum membrane (Wessels & Spiess, 1988). Subsequent TM segments need not necessarily interact with SRP but may insert directly into the translocon. At some point the TM segments move from the translocon laterally into the lipid bilayer. It has been suggested that a TM segment in a Type I single-span membrane protein remains in the translocon until protein synthesis has been terminated (Thrift et al., 1991). Perhaps only one polypeptide segment can optimally occupy the oligosaccharyltransferase active site for the duration of protein biosynthesis. Glycosylation of multiple consensus sites within this single domain is possible, although utilization of potential acceptor sites on subsequent extracytosolic domains is unlikely. Proteins consisting of duplicated independently folding domains, such as the α -subunit of the Na^+ channel, may engage the translocation machinery via an SRP more than once. This would allow the efficient transfer of oligosaccharides to multiple extracytosolic loops, one in each independent subdomain of the polypeptide.

Cell-free-translation experiments using isolated microsomes have shown that multiple extracytosolic peptide regions can be N-glycosylated under certain conditions (Wessels & Spiess, 1988; Lipp et al., 1989). These experiments showed that, in proteins containing identical consensus sites, the first loop is efficiently glycosylated, whereas subsequent loops are poorly glycosylated (Wessels & Spiess, 1988). Interestingly, elimination of the first glycosylation site allowed for enhanced glycosylation of the second site (Lipp et al., 1989). In this *in vitro* assay, both loops, which contain suitable acceptor sites, potentially have access to the glycosylation machinery. The constructs tested, however, consisted of chimaeric multi-span membrane proteins lacking a normal folded structure and function. Results obtained by *in vitro* studies (e.g. cell-free translation) have been shown not to be directly applicable to the *in vivo* situation. For example, only Asn²⁹³ is N-glycosylated in ovalbumin in the oviduct, and a second distal site (Asn³¹²) is also N-glycosylated *in vitro* (Kaplan et al., 1987). Protein synthesis, folding and glycosylation *in vitro* occurs at a lower rate than *in vivo* perhaps allowing for secondary glycosylation events (Kaplan et al., 1987). *In vitro*, the nascent polypeptide chain is continually exposed to the glycosylation machinery in microsomes. In addition, a puzzling observation of *in vitro* N-glycosylation without translocation has been reported (Lipp et al., 1989). Our survey deals with the normal *in vivo* N-glycosylation of functional membrane proteins rather than an artificial *in vitro* situation.

Structural and functional aspects

The preceding discussion addressed possible biosynthetic limitations which may direct the synthesis of multi-span membrane glycoproteins. The fundamental question remains as to why N-linked oligosaccharides are not found on multiple extracytosolic domains in multi-span membrane proteins. As already mentioned, it is possible that N-glycosylation is important only early in the folding process. Glycosylation of multiple segments may impede the folding process by interfering with TM helix packing. Finally, glycosylation of multiple domains on a polytopic membrane protein may adversely affect protein function. It has been demonstrated for the oligosaccharide chains of haemagglutinin that the sugar interacts with the protein via specific hydrophilic interactions (Wiley & Skehel, 1987). Overshielding of the protein surface by oligosaccharides on

multiple extracellular loops might seriously hamper protein function. Most of the surveyed proteins act as receptors, channels or transporters of polar substrates, and the clustering of multiple polar sugar moieties around the opening to a channel might compromise substrate access to the translocation site. The N-glycosylated domain is therefore expected to be remote from the transport channel.

Conclusions

In conclusion, N-linked oligosaccharide chains are localized to single extracytosolic segments in multi-span membrane proteins. In order to be efficiently N-glycosylated, a polypeptide loop in a membrane protein must: (1) contain a consensus N-glycosylation site (Asn-Xaa-Ser/Thr, Xaa $\neq$ Pro); (2) be luminally disposed in the endoplasmic reticulum; (3) be greater than 30 residues in size; (4) have the consensus site more than ten residues away from the bilayer; and (5) be the first such loop exposed to the glycosylation machinery in the endoplasmic reticulum lumen. These conclusions are open to further testing by introducing suitable N-glycosylation sites into one or more loops or by changing the size of extracytosolic loops containing N-glycosylation acceptor sites in membrane proteins. The effect of hyper-glycosylation of multiple loops on the biosynthesis and functioning of membrane proteins could also be tested. The introduction of novel N-glycosylation sites into individual putative extracytosolic loops and assaying for their utilization may be a viable method of studying the topology of multi-span membrane proteins. The inability to glycosylate a consensus site, however, does not necessarily preclude its extracytosolic localization. The results presented here suggest that, in such studies, N-glycosylation sites should be carefully designed to allow for the optimum utilization of the acceptor site(s). In addition, these studies should be done in an *in vivo* system using functional proteins in order to ensure that proper folding has occurred.

Dr. G. Uhl, Dr. P. Agre and Dr. W. Agnew are thanked for providing information prior to publication. D. B. Williams, C. M. Deber, D. H. MacLennan, V. Sarabia, L. Tam and J. H. M. Charuc are thanked for their comments on the manuscript. C. L.-M. was supported by an MRC (Medical Research Council of Canada) Studentship. Financial support was provided by a Group Grant from MRC.

REFERENCES

- Alper, S. L. (1991) *Annu. Rev. Physiol.* **53**, 549–564
- Amara, S. G. (1992) *Nature (London)* **360**, 420–421
- Asano, T., Katagiri, H., Takata, K., Lin, J.-L., Ishihara, H., Inukai, K., Tsukuda, K., Kikuchi, M., Hirano, H., Yazaki, Y. and Oka, Y. (1991) *J. Biol. Chem.* **266**, 24632–24636
- Casey, J. R., Pirraglia, C. A. and Reithmeier, R. A. F. (1992) *J. Biol. Chem.* **267**, 11940–11948
- Catterall, W. A. (1988) *Science* **242**, 50–61
- Cheng, S. H., Gregory, R. J., Marshall, J., Paul, S., Souza, D. W., White, G. A., O'Riordan, C. R. and Smith, A. E. (1990) *Cell* **63**, 827–834
- Counillon, L. and Pouyssegur, J. (1993) *Molecular Biology and Function of Carrier Proteins* (Reuss, L., Russell, J. M. and Jennings, M. L., eds.), pp. 170–185, The Rockefeller University Press, New York
- Counillon, L., Reithmeier, R. A. F. and Pouyssegur, J. (1994) *Biochemistry*, in the press
- Danbolt, N. C., Storm-Mathisen, J. and Kanner, B. I. (1992) *Neuroscience* **51**, 295–310
- DeLorey, T. M. and Olsen, R. W. (1992) *J. Biol. Chem.* **267**, 16747–16750
- Doi, T., Molday, R. S. and Khorana, H. G. (1990) *Proc. Natl. Acad. Sci. U.S.A.* **87**, 4991–4995
- Drickamer, L. K. (1978) *J. Biol. Chem.* **253**, 7242–7248
- Fukuda, M. N., Papermaster, D. S. and Hargrave, P. A. (1979) *J. Biol. Chem.* **254**, 8201–8207
- Gallagher, P., Henneberry, J., Wilson, I., Sambrook, J. and Gething, M.-J. (1988) *J. Cell. Biol.* **107**, 2059–2073
- Gavel, Y. and von Heijne, G. (1990) *Protein Eng.* **3**, 433–442
- Gehle, V. M. and Sumikawa, K. (1991) *Mol. Brain. Res.* **11**, 17–25
- Gregory, R. J., Rich, D. P., Cheng, S. H., Souza, D. W., Paul, S., Manavalian, P., Anderson, M. P., Welsh, M. J. and Smith, A. E. (1991) *Mol. Cell. Biol.* **11**, 3886–3893
- Hediger, M. A., Mendlein, J., Lee, H.-S. and Wright, E. M. (1991) *Biochim. Biophys. Acta* **1064**, 360–364
- Hirayama, B. A. and Wright, E. M. (1992) *Biochim. Biophys. Acta* **1103**, 37–44
- Hryshko, L. V., Nicoll, D. A., Weiss, J. N. and Philipson, K. D. (1993) *Biochim. Biophys. Acta* **1151**, 35–42
- Hübner, S., Michel, F., Rudloff, V. and Appelhans, H. (1992) *Biochem. J.* **285**, 17–23
- Jennings, M. L. (1992) *The Kidney: Physiology and Pathophysiology*, 2nd edn. (Seldin, D. W. and Giebisch, G., eds.), pp. 503–535 Raven Press Ltd., New York
- Kaplan, H. A., Welpy, J. K. and Lennarz, W. J. (1987) *Biochim. Biophys. Acta* **906**, 161–173
- Kelleher, D. J., Kreibich, G. and Gilmore, R. (1992) *Cell* **69**, 55–65
- Khorana, H. G. (1992) *J. Biol. Chem.* **267**, 1–4
- Kobayashi, S., Morgans, C. W., Casey, J. R. and Kopito, R. R. (1994) *J. Neurosci.*, in the press
- Kornfeld, R. and Kornfeld, S. (1985) *Ann. Rev. Biochem.* **54**, 631–664
- Kraner, S., Yang, J. and Barchi, R. (1989) *J. Biol. Chem.* **264**, 13272–13280
- Lindsey, A. E., Schneider, K., Simmons, D. M., Baron, R., Lee, B. S. and Kopito, R. R. (1990) *Proc. Natl. Acad. Sci. U.S.A.* **87**, 5278–5282
- Lipp, J., Flint, N., Haeuptle, M.-T. and Dobberstein, B. (1989) *J. Cell. Biol.* **109**, 2013–2022
- Loo, T. W. and Clarke, D. M. (1994) *J. Biol. Chem.* **269**, 7750–7755
- Maceyka, M., Stephan, M., Ukome, C. and Agnew, W. S. (1993) *Biophys. J.* **64**, M-Pos 376
- Nielsen, S., Smith, B. L., Christensen, E. I., Knepfer, M. A. and Agre, P. (1993) *J. Cell Biol.* **120**, 371–383
- Nilsson, I. and von Heijne, G. (1993) *J. Biol. Chem.* **268**, 5798–5801
- Nomoto, H., Takahashi, N., Nagaki, Y., Endo, S., Arata, Y. and Hayashi, K. (1986) *Eur. J. Biochem.* **157**, 233–242
- Ostrowski, J., Kjelsberg, M. A., Caron, M. G. and Lefkowitz, R. J. (1992) *Annu. Rev. Pharmacol. Toxicol.* **32**, 167–183
- Paulson, J. C. (1989) *Trends Biochem. Sci.* **14**, 272–276
- Pessin, J. E. and Bell, G. I. (1992) *Annu. Rev. Physiol.* **54**, 911–930
- Philipson, K. D. and Nicoll, D. A. (1992) *Curr. Opin. Cell Biol.* **4**, 678–683
- Preston, G. M., Jung, J. S., Guggino, W. B. and Agre, P. (1994) *J. Biol. Chem.* **269**, 1668–1673
- Rands, E., Candelore, M. R., Cheung, A. H., Hill, W. S., Strader, C. D. and Dixon, R. A. F. (1990) *J. Biol. Chem.* **265**, 10759–10764
- Reid, D. M., Friedel, U., Molday, R. S. and Cook, N. J. (1990) *Biochemistry* **29**, 1601–1607
- Reiländer, H., Achilles, A., Friedel, U., Maul, G., Lottspeich, F. and Cook, N. J. (1992) *EMBO J.* **11**, 1689–1695
- Reithmeier, R. A. F. and Deber, C. M. (1992) *The Structure of Biological Membranes* (Yeagle, P., ed.), pp. 337–393, CRC Press, Boca Raton, FL
- Riordan, J. R., Rommens, J. M., Kerem, B.-S., Alon, N., Rozmahel, R., Gselsczak, Z., Zielenksi, J., Lok, S., Plavsic, N., Chou, J.-L., Drumm, M. L., Iannuzzi, M. C., Collins, F. S. and Tsui, L.-C. (1989) *Science* **245**, 1066–1073
- Roitsch, T. and Lehle, L. (1989) *Eur. J. Biochem.* **181**, 525–529
- Rosenberg, R. L. and East, J. E. (1992) *Nature (London)* **360**, 166–171
- Schinkel, A. H., Kemp, S., Dollé, M., Rudenko, G. and Wagenaar, E. (1993) *J. Biol. Chem.* **268**, 7474–7481
- Simon, S. M. and Blobel, G. (1991) *Cell* **65**, 371–380
- Thrift, R. N., Andrews, D. W., Walter, P. and Johnson, A. E. (1991) *J. Cell Biol.* **112**, 809–821
- Uhl, G. R. (1992) *Trends Neurosci.* **15**, 265–268
- van Koppen, C. J. and Nathanson, N. M. (1990) *J. Biol. Chem.* **265**, 20887–20892
- Wessels, H. P. and Spiess, M. (1988) *Cell* **55**, 61–70
- Wiley, D. C. and Skehel, J. J. (1987) *Annu. Rev. Biochem.* **56**, 365–394
- Wohlfart, P., Müller, H. and Cook, N. J. (1989) *J. Biol. Chem.* **264**, 20934–20939
- Wohlfart, P., Haase, W., Molday, R. S. and Cook, N. J. (1992) *J. Biol. Chem.* **267**, 644–648
- Wright, E. M., Hager, K. M. and Turk, E. (1992) *Curr. Opin. Cell Biol.* **4**, 696–702
- Yoshimura, A., Kuwazuru, Y., Sumizawa, T., Ichikawa, M., Ikeda, S., Uda, T. and Akiyama, S. (1989) *J. Biol. Chem.* **264**, 16282–16291