Proc. Natl. Acad. Sci. USA
Vol. 86, pp. 7366–7369, October 1989
Biochemistry

A 13-amino acid peptide in three yeast glycosyltransferases may be involved in dolichol recognition

(isoprenoid/endoplasmic reticulum/N-glycosylation)

CHARLES F. ALBRIGHT*, PETER ORLEAN, AND PHILLIPS W. ROBBINS

Biology Department and Center for Cancer Research, Massachusetts Institute of Technology, Cambridge, MA 02139

Contributed by Phillips W. Robbins, June 30, 1989

ABSTRACT **A 13-amino acid peptide was identified in three glycosyltransferases of the yeast endoplasmic reticulum. These enzymes, the products of the *ALG1*, *ALG7*, and *DPM1* genes, catalyze the transfer of sugars from nucleotide sugars to dolichol phosphate derivatives. The consensus sequence for the conserved peptide was Leu-Phe-Val-Xaa-Phe-Xaa-Xaa-Ile-Pro-Phe-Xaa-Phe-Tyr. A sequence resembling the conserved peptide was also found in the predicted SEC59 protein, which is suspected to participate in assembly of the lipid-linked precursor oligosaccharide, although its specific function is unknown. All of the identified sequences contain an isoleucine at position 8 and phenylalanine or tyrosine at positions 2, 5, and 12. We believe this peptide may be involved in dolichol recognition for the following reasons. (*i*) The conserved sequence occurs in potential membrane-spanning regions. (*ii*) The ALG7 and DPM1 proteins are known to recognize the isoprenoid region of dolichol phosphate specifically; this recognition presumably occurs in the membrane since dolichol is very hydrophobic. (*iii*) The consensus sequence is similar to a region of two halobacterial proteins implicated in binding of the isoprenoid region of retinal. (*iv*) If the consensus sequence is represented as an α-helix, the conserved residues lie on one face of the helix. An α-helical structure is likely since the conserved regions are in potential membrane-spanning domains.**

N-glycosylation begins with the transfer of a preassembled lipid-linked precursor oligosaccharide (Dol-*PP*-$GlcNAc_2Man_9Glc_3$, where Dol is dolichol) to protein in the lumen of the endoplasmic reticulum (ER) (1, 2). The assembly of the precursor oligosaccharide occurs by an ordered process in which sugars are transferred from nucleotide sugars or dolichol-linked sugars to the growing chain. The biochemistry and genetics of the known enzymes catalyzing this process suggest that the addition of each of the 14 sugars may require a unique enzyme.

To understand better the assembly of the lipid-linked precursor oligosaccharide we are studying this process in *Saccharomyces cerevisiae*, where three glycosyltransferases have been cloned and sequenced. The ALG7 protein catalyzes the formation of Dol-*PP*-GlcNAc from Dol-*P* and UDP-GlcNAc, and this reaction is inhibited by the antibiotic tunicamycin (3, 4). The ALG1 protein catalyzes the formation of Dol-*PP*-$GlcNAc_2Man$ from Dol-*PP*-$GlcNAc_2$ and GDP-Man (5). The DPM1 protein catalyzes the formation of Dol-*P*-Man from GDP-Man and Dol-*P* (6). Dol-*P*-Man serves as the mannosyl donor for the last five mannose residues of the lipid-linked precursor as well as mannose donor in other glycosylation processes (7–9). Antisera to the ALG1 and DPM1 proteins have been used to show that both proteins are integral membrane proteins that are oriented toward the cytoplasmic face of the ER membrane (C.F.A. and P.O., unpublished data). In addition to these proteins, the phenotype of the yeast *sec59* mutant suggests a role for the SEC59 protein in biosynthesis of the lipid-linked oligosaccharide precursor, although its specific function is unclear (10).

The availability of these sequences led us to look for common structural features. We report here the identification of a small region of protein sequence shared by the ALG7, ALG1, DPM1, and SEC59 proteins. The location of this homologous sequence in a membrane-spanning domain, the possible structure of this sequence, and the need for these enzymes to recognize dolichol derivatives suggest this conserved region may be involved in dolichol recognition.

MATERIALS AND RESULTS

A comparison of the deduced amino acid sequences for the ALG7, ALG1, and DPM1 proteins revealed a small region of conserved amino acids; this homology was particularly striking since these proteins have very little primary sequence homology. Table 1 shows this conserved region as well as adjoining amino acids that were predicted to form a membrane-spanning domain. All of the sequences contained isoleucine at position 8. In addition, phenylalanine and tyrosine at positions 2, 5, 10, 12, and 13 were highly conserved. The ALG7 fragment contains aromatic amino acids at all five positions, whereas the ALG1 and DPM1 fragments contain aromatic residues in four of these positions.

We first evaluated the statistical significance of these findings by searching the protein sequence data base using the consensus sequence we derived as a probe (Table 1). The similarity scores and *z*-values were determined for the ALG7, ALG1, and DPM1 regions and compared to the values for the protein data base (Table 1). The similarity score is a measure of the relatedness of two sequences based on empirically determined values for the probability that a given amino acid will be replaced by another amino acid in the course of evolution (11). The ALG7, ALG1, and DPM1 sequences had similarity scores of 67, 47, and 44, respectively. The maximum possible similarity score was 67. Two other protein fragments in the data base scored 45 and no other fragments scored above this value. The *z*-value is a measure of the number of standard deviations from the mean score; *z*-values greater than 6 in pairwise comparisons are conservatively viewed as probably significant (12). The ALG7, ALG1, and DPM1 regions had *z*-values of 8.3, 6.2, and 5.9, respectively. The high *z*-values and the scarcity of related sequences suggested that the occurrence of a region conserved between the ALG7, ALG1, and DPM1 proteins is statistically significant.

We also searched the SEC59 protein sequence with the consensus sequence since the protein has been implicated in biosynthesis of the lipid-linked precursor oligosaccharide. The SEC59 protein encodes a 519-amino acid protein whose

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: ER, endoplasmic reticulum; Dol, dolichol.

*To whom reprint requests should be addressed.

Table 1. Comparison of protein fragments from ALG7, ALG1, DPM1, and SEC59 proteins

Protein	Residues	Fragment sequence	Similarity score	z-value
ALG7	70–92	T-I-G-A-I-P-A-A-V-Y-L-F-V-M-F-I-Y-I-P-F-I-F-Y	67	8.3
ALG1	8–34	W-L-L-A-L-I-I-L-Y-L-S-I-P-L-V-V-Y-Y-V-I-P-Y-L-F-Y-G-N	44	5.9
DPM1	239–260	F-G-A-N-N-L-I-L-F-I-T-F-W-S-I-L-F-F-Y-V-C-Y	45	6.0
Consensus*		L-F-V-X-F-X-X-I-P-F-X-F-Y		
		1 3 5 7 9 11 13		
SEC59	331–346	L-W-H-F-I-L-F-L-L-I-I-P-S-F-Q-M	28	4.3

Fragment sequences are presented in the single-letter amino acid code. Similarity scores were calculated relative to the consensus sequence by using the mutation data matrix. The z-value is (similarity score − average similarity score)/(standard deviation of similarity scores). Version 15.0 of the National Biomedical Research Foundation protein data base was searched using the SEARCH program with the consensus sequence as a probe. The average similarity value for the data base was −13.7 and the standard deviation was 9.7.
*A residue was added to the consensus sequence whenever it was contained by two of the three sequences (ALG7, ALG1, and DPM1).

sequence is unusual in that it contains hydrophobic regions 12–20 amino acids long that are punctuated by clusters of charged residues (10). This search revealed a homologous region with a similarity score of 28 and a z-value of 4.3 (Table 1). This fragment contained the conserved isoleucine at position 8 and three of the five conserved aromatic residues.

We also noticed that many of the sequences with high similarity scores were involved in electron transport; 13% of the top 500 similarity scores were from proteins that comprise the NADH–ubiquinone oxidoreductase complex although these proteins comprised only about 1% of the sequences in the data base. The high similarity scores were at least partially due to the hydrophobic nature of these proteins.

DISCUSSION

The statistical significance of a conserved peptide in three glycosyltransferases of the ER suggests a functional role for this region. Potential functions for this sequence could include a catalytic function, such as substrate recognition, or an ER-specific function, such as localization. The absence of this sequence from other known ER proteins, its location in membrane-spanning domains, and the need for these enzymes to recognize dolichol derivatives all suggest that the conserved region may be involved in dolichol recognition.

There is biochemical evidence that the ALG7 and DPM1 enzymes recognize the isoprenoid region of Dol-*P*. The structure of Dol-*P* is illustrated in Fig. 1. Palamarczyk *et al.* (15) have studied the ability of various isoprenoid compounds to function as acceptors in transfer reactions catalyzed by solubilized yeast enzymes. For the ALG7 protein they found the K_m for Dol-*P* was 31 μM. The K_m increased to 117 μM as the isoprenoid chain length was decreased from 100 to 55 carbons and an isoprenoid derivative with only 35 carbons give rise to very low activity in their assay. Changing the isoprene units from a cis configuration to a trans configuration and changing the α-isoprene unit to an unsaturated unit also greatly reduced transfer activity. For the DPM1 protein, they found the K_m for Dol-*P* was about 1 μM. Decreasing the isoprenoid length to 35 carbons decreased the reaction velocity by 10-fold and changing the α-isoprene unit to an unsaturated unit increased the K_m for Dol-*P* about 10-fold. It is not surprising that the ALG7 and DPM1 enzymes specifically recognize the isoprenoid region of Dol-*P* since the membrane has many potential phosphate acceptors; if the isoprenoid region were not recognized, other phosphate esters might compete for the transfer reactions and possibly interfere with glycosylation.

There is genetic evidence that isoprenoid structure affects glycosylation in mammalian cells. A Chinese hamster ovary cell mutant F2A8 primarily synthesizes a dolichol derivative

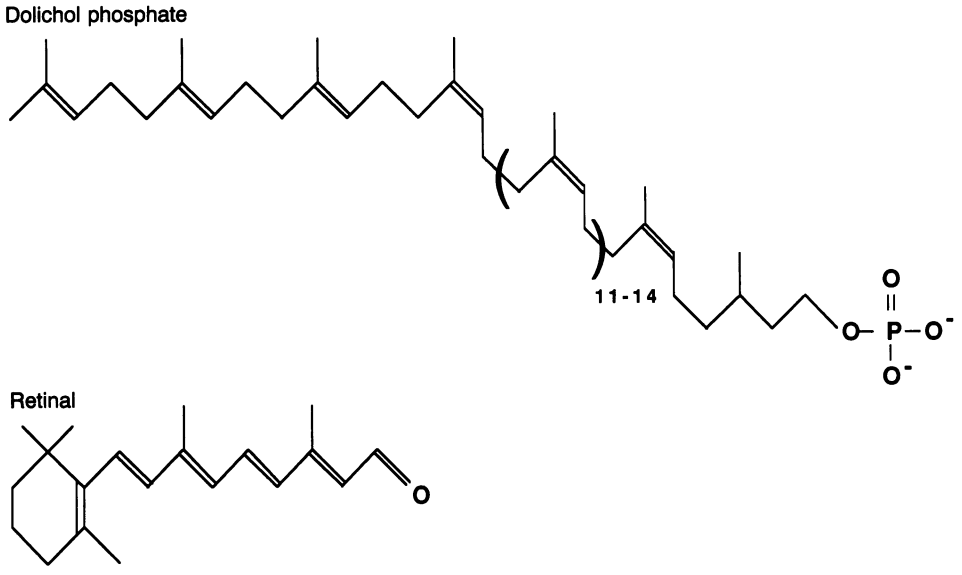

FIG. 1. Structures of Dol-*P* and retinal. The structure of the carbon backbone of Dol-*P* is shown in the upper diagram. Dol-*P* is a polyisoprenoid that typically contains 17–20 isoprene units (for review, see ref. 13). The α-isoprene unit, which is adjacent to the phosphate residue, is saturated and the remaining isoprene units are unsaturated. The three ω-isoprene units are in the trans configuration and the remaining isoprene units are in the cis configuration. This figure is adapted from ref. 14. The structure of retinal is shown in the lower diagram. The linear isoprenoid region is completely unsaturated and in the trans configuration.

with an α-unsaturated isoprene unit, probably due to the loss of polyprenol reductase activity (16). This mutation results in a 7-fold reduction in the incorporation of mannose into lipid-linked oligosaccharides and protein, suggesting that several mammalian enzymes recognize the isoprenoid chain *in vivo* (17).

The similarity of the conserved sequence to an amino acid sequence implicated in retinal binding is also consistent with a role in dolichol recognition. The sequence Trp-Xaa-Xaa-Tyr-Pro-Xaa-Xaa-Trp is found in two halobacterial proteins that catalyze light-driven ion transport using retinal as a chromophore (for review, see ref. 18). This sequence occurs in a membrane-spanning domain as an α-helix and seems to form part of the binding pocket for retinal (18). If the retinal-binding sequence is aligned with the possible dolichol-binding sequence using the proline residue as a reference point, then the dolichol-binding sequence contains phenylalanine residues instead of tryptophan residues and an isoleucine instead of tyrosine. These differences could reflect the structural differences between retinal and dolichol; although retinal contains an isoprenoid chain, it has a higher degree of unsaturation than dolichol (Fig. 1). Even if this alignment is incorrect, it is still interesting that a potential retinal-binding sequence and a potential dolichol-binding sequence are rich in aromatic residues and contain a proline residue.

Since the conserved regions occur within membrane-spanning domains, it seems likely that they form α-helices (19). In this conformation, the conserved residues lie on one face of the helix; this is consistent with a role in dolichol recognition (Fig. 2). The location of the isoleucine at position 8 and the aromatic residues at positions 2, 5, and 12 are particularly interesting since these residues are conserved in all four proteins. In addition, the presence of a proline at residue 9 or 10 is unusual since prolines are significantly underrepresented in membrane-spanning domains (20). In particular, prolines are rarely found in membrane-spanning regions of nontransport proteins whereas membrane-spanning regions of transport proteins nearly all contain proline residues (21). On the basis of this criterion, the ALG1, ALG7, DPM1, and SEC59 proteins could be classified as transport proteins, which is interesting given the topography of glycosylation in the ER. In particular, the DPM1 protein is reported to catalyze the translocation of Dol-*P*-Man across the membrane (22, 23) and products downstream from the ALG1- and ALG7-catalyzed reactions are believed to be translocated across the membrane (for review, see ref 24).

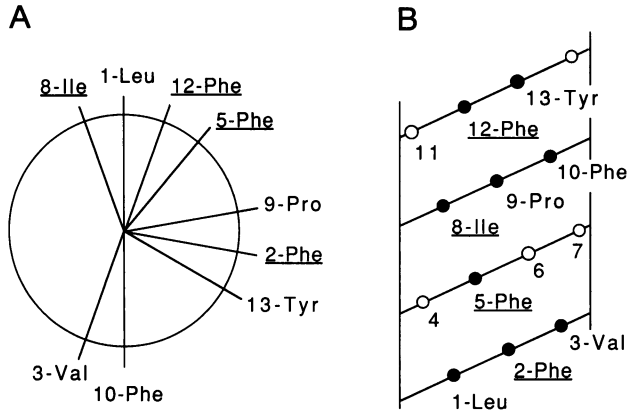

FIG. 2. Representation of consensus sequence as an α-helix. A helical wheel (*A*) and a helical net (*B*) of the consensus sequence are illustrated. Conserved residues are indicated by residue number and amino acid name. Residues that are conserved in all four proteins are underlined.

Although the function of the SEC59 protein is unknown, it seems likely that it is involved in biosynthesis of the lipid-linking precursor oligosaccharide. In particular, loss of SEC59 function results in changes in the pool of lipid-linked precursor oligosaccharides (25) (K. Runge, Ph.D. Thesis, M.I.T.) and the transfer of a reduced number of full-length oligosaccharide chains to proteins (26, 27). Partial restoration of glycosylation was obtained by supplementing membranes containing mutant protein with mannose or GDP-mannose, but not with other sugar nucleotides (10). These results suggest a role for the SEC59 protein in the glycosylation pathway and, hence, the possibility that the SEC59 protein may recognize one or more dolichol derivatives. The lack of total agreement between the SEC59 segment and the glycosyltransferase consensus sequence could reflect a decreased need for the SEC59 protein to recognize the isoprenoid region of dolichol. As the lipid-linked oligosaccharide chain grows, the enzymes may rely more on the oligosaccharide and less on the isoprenoid chain for recognition.

Since ubiquinone contains a polyisoprenoid chain, it would not be surprising if proteins in the NADH–ubiquinone oxidoreductase complex contained sequences related to the glycosyltransferase consensus sequence. However, identification of such a specific region on the basis of sequence comparisons could prove difficult. Many of these proteins are very hydrophobic and, therefore, contain a high background of similar, but probably nonfunctional, sequences. Furthermore, the need to recognize the isoprenoid region of ubiquinone may be minimal since the ubiquinone headgroup could provide a high degree of binding specificity; this is certainly the case with the photosystem II complex that contains two ubiquinone-binding sites (28, 29). This hypothesis is also supported by the finding that long- and short-chain ubiquinones function in electron transport, suggesting that recognition of the isoprenoid chain of ubiquinone is not needed for efficient catalysis (30).

The functional importance of the conserved peptide could be tested by creating mutations in the conserved sequence of ALG7 or DPM1; the kinetic properties of the resulting proteins could be studied *in vitro* and the phenotype conferred by the mutant genes could be studied *in vivo*. In biophysical study of peptides and polyprenols, model membrane systems will be informative. For instance, studies with artificial membranes showed that Dol-*P* induces the formation of nonbilayer structures (31–33). These lipid structures have been shown to promote vesicle fusion *in vitro* (for review, see ref. 34) as well as increase the activity of the mammalian Dol-*P*-Man synthase *in vitro* (for review, see ref. 33). Perhaps, a peptide containing the conserved sequence will partition into nonbilayer regions. Alternatively, the peptide might inhibit the induction of nonbilayer structures by dolichol derivatives. Nonbilayer structures could be deleterious *in vivo* and the possibility has been raised that the widespread induction of nonbilayer organization by dolichols might be prevented by proteins (32).

We thank Dr. F. A. Troy for sharing results prior to publication and for many helpful comments on the manuscript. We also thank Dr. W. L. Adair for reviewing the manuscript. This work was supported by Grant GM-31318 to P.W.R. and Grant CA-26712 to P. Sharp from the National Institutes of Health.

1. Kornfeld, R. & Kornfeld, S. (1985) *Annu. Rev. Biochem.* **54**, 631–664.
2. Kukuruzinska, M. A., Bergh, M. L. E. & Jackson, B. J. (1987) *Annu. Rev. Biochem.* **56**, 915–944.
3. Rine, J., Hansen, W., Hardeman, E. & Davis, R. W. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 6750–6754.
4. Hartog, K. O. & Bishop, B. (1987) *Nucleic Acids Res.* **15**, 3627.
5. Couto, J. R., Huffaker, T. C. & Robbins, P. W. (1984) *J. Biol. Chem.* **259**, 378–382.

6. Orlean, P., Albright, C. & Robbins, P. W. (1988) *J. Biol. Chem.* **263,** 17499–17507.
7. Tanner, W. & Lehle, L. (1987) *Biochem. Biophys. Acta* **906,** 81–99.
8. Fatemi, S. H. & Tartakoff, A. M. (1986) *Cell* **46,** 653–657.
9. Conzelmann, A., Spiazzi, A., Hyman, R. & Bron, C. (1986) *EMBO J.* **5,** 3291–3296.
10. Bernstein, M., Kepes, F. & Schekman, R. (1989) *Mol. Cell. Biol.* **9,** 1191–1199.
11. Dayhoff, M. O., ed. (1978) *Atlas of Protein Sequence and Structure* (Natl. Biomed. Res. Found., Washington, DC), Vol. 5, Suppl. 3.
12. Lipman, D. J. & Pearson, W. R. (1985) *Science* **227,** 1435–1441.
13. Chojnacki, T. & Dallner, G. (1988) *Biochem. J.* **251,** 1–9.
14. Suzuki, S., Mori, F., Takigawa, T., Ibata, K., Ninagawa, Y., Nishida, T., Mizuno, M. & Tanaka, Y. (1983) *Tetrahedron Lett.* **24,** 5103–5106.
15. Palamarczyk, G., Lehle, L., Mankowski, T., Chojnacki, T. & Tanner, W. (1980) *Eur. J. Biochem.* **105,** 517–523.
16. Stoll, J., Rosenwald, A. G. & Krag, S. S. (1988) *J. Biol. Chem.* **263,** 10774–10782.
17. Stoll, J. & Krag, S. S. (1988) *J. Biol. Chem.* **263,** 10766–10773.
18. Oesterhelt, D. & Tittor, J. (1989) *Trends Biochem. Sci.* **14,** 57–61.
19. Engelman, D. M., Steitz, T. A. & Goldman, A. (1986) *Annu. Rev. Biophys. Biophys. Chem.* **15,** 321–353.
20. Michel, H., Weyer, K. A., Gruenberg, H., Dunger, I., Oesterhelt, D. & Lottspeich, F. (1986) *EMBO J.* **5,** 1149–1158.
21. Brandl, C. J. & Deber, C. M. (1986) *Proc. Natl. Acad. Sci. USA* **83,** 917–921.
22. Haselbeck, A. & Tanner, W. (1982) *Proc. Natl. Acad. Sci. USA* **79,** 1520–1524.
23. Haselbeck, A. & Tanner, W. (1984) *FEMS Microbiol. Lett.* **21,** 305–308.
24. Hirschberg, C. B. & Snider, M. D. (1987) *Annu. Rev. Biochem.* **56,** 63–87.
25. Runge, K. (1985) Dissertation (Massachusetts Institute of Technology, Cambridge, MA).
26. Ferro-Novick, S., Novick, P. & Schekman, R. (1984) *J. Cell Biol.* **98,** 35–43.
27. Julius, D., Schekman, R. & Thorner, J. (1984) *Cell* **36,** 309–318.
28. Deisenhofer, J., Epp, O., Miki, K., Huber, R. & Michel, H. (1985) *Nature (London)* **318,** 618–624.
29. Trebst, A. (1987) *Z. Naturforsch. C* **42,** 742–750.
30. Cornell, B. A., Keniry, M. A., Post, A., Robertson, R. N., Weir, L. E. & Westerman, P. W. (1987) *Biochemistry* **26,** 7702–7707.
31. Valtersson, C., van Duyn, G., Verkleij, A. J., Chojnacki, T., de Kruijff, B. & Dallner, G. (1985) *J. Biol. Chem.* **260,** 2742–2751.
32. Knudsen, M. J. & Troy, F. A. *Chem. Phys. Lipids*, in press.
33. Schutzbach, J. S., Jensen, J. W., Lai, C. & Monti, J. A. (1987) *Chem. Scr.* **27,** 109–118.
34. van Duijn, G., Verkleij, A. J., de Kruijff, B., Valtersson, C., Dallner, G. & Chojnacki, T. (1987) *Chem. Scr.* **27,** 95–100.