

# The essential yeast *NLT1* gene encodes the 64 kDa glycoprotein subunit of the oligosaccharyl transferase

Rahul Pathak, Carl S. Parker, Barbara Imperiali\*

Division of Chemistry and Chemical Engineering, California Institute of Technology, Pasadena, CA 91125, USA

Received 27 February 1995; revised version received 3 March 1995

**Abstract** The yeast oligosaccharyl transferase catalyzes the glycosylation of asparagine residues in secreted, vesicular, and membrane proteins. A complex of at least four membrane-bound polypeptides is responsible for oligosaccharyl transferase activity. Amino acid sequences from the 64 kDa glycoprotein subunit of the complex were used to clone the essential *NLT1* (N-linked oligosaccharyl transferase) gene. The Nlt1p gene product is a processed, multiply glycosylated type I membrane protein; it has an extensive amino-terminal soluble domain, a potential hydrophobic transmembrane domain, and a short carboxy-terminal soluble domain. The Nlt1p is significantly similar to the mammalian ribophorin I, a component of the mammalian oligosaccharyl transferase complex, and the enzyme is conserved throughout eukaryotic evolution.

**Key words:** Oligosaccharyl transferase; Protein glycosylation; Endoplasmic reticulum; *Saccharomyces cerevisiae*

## 1. Introduction

The N-linked glycosylation of proteins is an essential processing event in eukaryotic cells; oligosaccharide chains contribute to the folding, solubility, and overall conformation of glycoproteins in addition to participating in protein function [1–3]. The oligosaccharyl transferase catalyzes the transfer of a core oligosaccharide, Glc<sub>3</sub>Man<sub>9</sub>(GlcNAc)<sub>2</sub>, from dolichol pyrophosphate to selected asparagine residues [4] of nascent polypeptides as they are translocated into the lumen of the rough endoplasmic reticulum. The enzyme has recently been purified from detergent extracts of *Saccharomyces cerevisiae* membrane proteins [5–7]. The enzyme was purified either as a tetrameric complex of polypeptides with molecular weights of 64 kDa, 45 kDa, 34 kDa, and 32 kDa, or as a hexameric complex with additional 16 kDa and 9 kDa subunits. The 45 kDa and 32 kDa subunits were identified as the products of the essential *WBP1* and *SWP1* genes. Previous experiments with the cloned *WBP1* and *SWP1* genes had identified Wbp1p and Swp1p as protein subunits of an oligosaccharyl transferase complex; however, simultaneous over-expression of the genes encoding both proteins failed to significantly increase oligosaccharyl transferase activity [8,9]. Presumably, over-expression of the enzyme also requires the genes encoding other subunits of purified enzyme complex. The amino acid sequences of the amino-terminal and internal peptide fragments from the 64 kDa and 34 kDa subunits indicated that they were novel yeast proteins.

The sequences of the identified components of the oligosac-

charyl transferase are well conserved in eukaryotes from yeast to mammals. The mammalian enzyme was previously purified from canine pancreas microsomal membranes [10] as a complex including the proteins ribophorin I, ribophorin II, and OST48. The amino acid sequence of the 32 kDa Swp1p is similar to the sequence of the carboxy-terminal half of the 63 kDa ribophorin II from rat and human cell lines [7], and the overall sequence of the 45 kDa Wbp1p is similar to the OST48 from canine pancreas cells [11].

Here, we describe the novel yeast protein Nlt1p, a third component of the yeast oligosaccharyl transferase. The *S. cerevisiae* *NLT1* gene was cloned using degenerate oligonucleotides derived from extensive amino acid sequencing of the 64 kDa subunit of the purified enzyme complex. The 1.4 kilobase open reading frame of the *NLT1* gene encodes a polypeptide with a predicted unprocessed molecular weight of 54 kDa. The amino acid sequence includes a 22 residue endoplasmic reticulum signal sequence and five potential glycosylation sites in the mature protein. Cleavage of the signal sequence and glycosylation of the acceptor sequences would yield a mature protein of 62–67 kDa, in close agreement with the observed doublet of 64 kDa polypeptide bands in the purified complex [5]. The amino acid sequence predicts a type I membrane protein with an extensive soluble domain and one potential transmembrane hydrophobic domain at the carboxy-terminus of the protein. Deletion of an 800 bp region of the *NLT1* coding region indicated that the gene is essential for vegetative growth of yeast. Consistent with the homology shown by the 45 kDa and 32 kDa yeast subunits to their mammalian counterparts, the amino acid sequence of the novel 64 kDa Nlt1p bears significant similarity to the amino acid sequences of the human and rat ribophorin I proteins [12,13]; the *NLT1* gene is the first gene encoding a ribophorin I homologue to be cloned from yeast.

The cloned *NLT1* gene, together with the *WBP1* and *SWP1* genes and perhaps the gene encoding the 34 kDa subunit, will enable a complete genetic and biochemical understanding of the essential oligosaccharyl transferase enzyme.

## 2. Materials and methods

### 2.1. Yeast strains and media

The following yeast strains were used in this study: SEY 6210/6211 (*MATa leu2-3112 ura3-52 his3-200 trp1-901 lys2-801/+ ade2-101/+ suc2-9 mal*) [14]; PRY46 (*MATa his- pep4-3 prc1-1126 prb1-1122*) (obtained from P. Robbins); YPH274 (*MATa/MATa ade2-101° lade2-101° lys2-801°/lys2-801° his3Δ200/his3Δ200 ura3-52/ura3-52 leu2-Δ1/leu2-Δ1 trp1-Δ1/trp1-Δ1*) [15]; YPH274a (*MATa ade2-101° lys2-801° his3 Δ200 ura3-52 leu2-Δ1 trp1-Δ1*) (obtained from J. Abelson); YPH274α (*MATa ade2-101° lys2-801° his3Δ200 ura3-52 leu2-Δ1 trp1-Δ1*) (obtained from J. Abelson); and RPY5 (*MATa/MATa ade2-101°/ade2-101° lys2-801°/lys2-801° his3Δ200/his3Δ200 ura3-52/ura3-52 leu2-Δ1/leu2-Δ1 trp1-Δ1/trp1-Δ1 Δnlt1::URA3/+*) (this study). Standard yeast media [16] were used.

\*Corresponding author. Fax: (1) (818) 564-9297.

Peptide 64-1	A Q Y E P P A T W E N V D Y K R T I D V
Peptide 64-2	L L P E G A T D H Y F T
Peptide 64-3	R Q T H F V N V L T M L

Fig. 1. Amino acid sequences of the 64 kDa glycoprotein. The yeast oligosaccharyl transferase was purified [5], and the 64 kDa glycoprotein subunit was isolated by SDS-PAGE. The sequence of peptide 64-1 was obtained from the mature amino-terminus of the protein, and the sequences of peptides 64-2 and 64-3 were obtained from two co-migrating cyanogen bromide fragments of the 64 kDa glycoprotein.

## 2.2. Amino acid sequencing of the 64 kDa protein

The yeast oligosaccharyl transferase was purified from approximately 100 g of PRY46 cells [5]. For amino-terminal sequencing, approximately 10 µg of the purified protein was concentrated by chloroform/methanol precipitation [17] and separated by SDS-PAGE [18]. The separated polypeptides were transferred electrophoretically [19] to a ProBlott PVDF membrane (Applied Biosystems). The 64 kDa oligosaccharyl transferase band was carefully cut out and sequenced on an Applied Biosystems automated peptide sequencer. For internal amino acid sequencing, approximately 100 µg of the purified protein was concentrated and separated by SDS-PAGE and then transferred to an Immobilon P<sup>80</sup> membrane (Millipore). The 64 kDa band was cut out and digested in situ with 10 mg/ml cyanogen bromide in 70% formic acid for 24 h at room temperature. The cyanogen bromide solution was removed, and the membrane fragments were washed with 70% isopropanol and 0.2% trifluoroacetic acid. The cyanogen bromide eluate was combined with the isopropanol wash and evaporated under a stream of nitrogen. The resultant peptide pellet was washed with water and then separated on a high resolution SDS-PAGE gel [20]. The peptide fragments were transferred to a ProBlott membrane (ABI), cut out, and sequenced on an Applied Biosystems automated peptide sequencer.

## 2.3. Isolation of the NLT1

Standard DNA techniques were used [21]. Degenerate oligonucleotides with sequences derived from the amino-terminal and internal amino acid sequences of the 64 kDa glycoprotein were synthesized (see Fig. 1). Primer RP1 had the sequence 5'-GCNCAA/GTAT/CGAA/GCCNCCNGCNAC-3', and primer RP6 had the sequence 5'-TAA/GTGA/GTCNGTNGCNCCT/CTCNGG-3'. Genomic DNA prepared from yeast strain SEY 6210/6211 was used as a template for the polymerase chain reaction with primers RP1 and RP6. An 850 bp fragment of the yeast genome was amplified and cloned into pBluescript (Stratagene) to yield plasmid pBluescript-850. Partial DNA sequence of the PCR product was obtained using Sequenase (US Biochemical), and the amino acid translation of the sequence was consistent with the amino acid sequences from the 64 kDa protein. Probes were prepared from the amplified DNA by the polymerase chain reaction with [ $\alpha$ -<sup>32</sup>P]dCTP using pBluescript-850 as a template; the specific activity of each probe was 10<sup>7</sup>–10<sup>8</sup> cpm/µg.

The probes were used to screen a  $\lambda$ YES yeast genomic library [22]. Lawns of *Escherichia coli* Y1090 cells were infected with the phage library at plaque densities of 30,000–50,000 per 150 mm plate. The plaques were transferred to nitrocellulose filters and hybridized with 0.5 × 10<sup>6</sup> cpm/ml of the probe synthesized from pBluescript-850 for 16 h

at 42°C in 0.9 M NaCl, 60 mM sodium phosphate (pH 7.4), and 0.1% SDS. The filters were then washed twice at 55°C with 0.9 M NaCl, 90 mM sodium citrate (pH 7.0), and 0.5% SDS; positive plaques were identified by autoradiography. From a screen of approximately 300,000 plaques, 22 plaques hybridizing with the probe were identified; seven of these plaques were purified after four rounds of screening, and one ( $\lambda$ YES-16) was selected for further study.

A plaque of  $\lambda$ YES-16 phage DNA was suspended in 10 mM Tris-Cl (pH 7.5) and 8 mM MgSO<sub>4</sub>. Recombination competent BNN132 cells [23] were infected with 1000 pfu of phage DNA and plated on LB/ampicillin media. The genomic insert of the  $\lambda$ YES phage is within an 8 kb pYES plasmid sequence flanked by *lox* sites; infection of BNN132 cells expressing the *cre* gene allows recombination to occur at the *lox* sites yielding an intact pYES plasmid and expression of the pYES ampicillin resistance gene. Approximately 100 ampicillin-resistant colonies grew, and plasmid DNA was isolated from one colony. Restriction digestion of the plasmid DNA revealed a 3.7 kb genomic clone flanked by *Xho*I restriction endonuclease sites. The clone was removed from the pYES vector by digestion with *Xho*I and inserted into the *Xho*I site of pBluescript to yield plasmid pBluescript(NLT1). Both strands of the entire 3.7 kb genomic clone were sequenced.

## 2.4. Disruption of the NLT1 gene

The 1.1 kb *Hind*III fragment containing the *URA3* gene from YEp24 was inserted between the *Stu*I and *Msc*I sites within the *NLT1* gene to yield plasmid pBluescript(*Δnlt1*); the modified gene contained an 800 bp deletion within the *NLT1* open reading frame. Diploid yeast strain YPH274 was transformed with a 4.0 kb linear DNA fragment carrying the modified *NLT1* gene [24]. Uracil prototrophs were selected, and successful disruption of one copy of the wild-type *NLT1* gene was verified by Southern blotting. The resulting diploid yeast strain RPY5 was sporulated and subjected to tetrad analysis.

## 3. Results

### 3.1. Isolation of the NLT1 gene

To clone the gene encoding the 64 kDa subunit of the oligosaccharyl transferase, amino acid sequences were obtained from the amino-terminus and from one cyanogen bromide fragment of the 64 kDa subunit [5]. The amino-terminus yielded 20 residues of reliable sequence, peptide 64-1 (see Fig. 1). A 30 kDa cyanogen bromide fragment yielded sequence from two co-migrating glycopeptides in a 4-to-1 molar ratio. The more abundant peptide gave 12 residues of reliable sequence, peptide 64-2; twelve residues of sequence were also obtained from the less abundant peptide, peptide 64-3, near the detection limits of the polypeptide sequencer.

The region of the yeast genome encoding the 64 kDa polypeptide from the amino-terminus to the start of the 30 kDa cyanogen bromide fragment was amplified by the polymerase chain reaction (PCR) with degenerate oligonucleotide primers derived from peptides 64-1 and 64-2. An amplified 850 bp

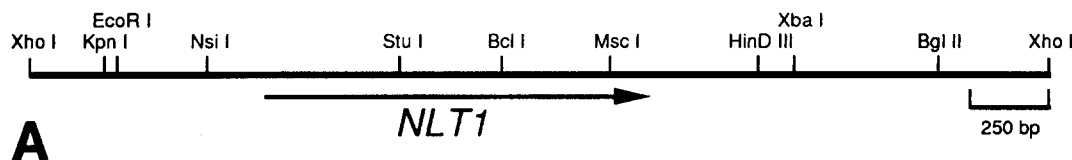


Fig. 2. (A) Restriction map of the *NLT1* locus. The location of the *NLT1* open reading frame is shown with an arrow in the 3.7 kb clone of the *NLT1* locus. Selected restriction enzyme cleavage sites are shown. The *Xho*I sites were introduced during the preparation of the library. (B) Nucleotide sequence of the *NLT1* locus. The entire 3.7 kb of the *NLT1* locus was sequenced on both strands. The sequence in the vicinity of the *NLT1* open reading frame is shown. The numbers above the sequence indicate the nucleotides of the *NLT1* gene starting with the ATG initiation codon. The numbers to the left of the sequence indicate the amino acids of the translation of the open reading frame. Amino acid sequences corresponding to those obtained from the 64 kDa glycoprotein are underlined, and potential N-linked glycosylation acceptor sites are also underlined. An arrow indicates the site of signal sequence cleavage by the endoplasmic reticulum signal peptidase. A fifteen residue potential transmembrane hydrophobic domain was revealed by hydropathy analysis [27] and is indicated by a dashed underline.

AAAATTCAAAAGAAATCTATCAATATAACAAAAACGCCAAATCATCAAAATGATCGGTG  
 TTGAAGAGCTTGGCACTCTTAAAGCGGCTAGTTCATAATGCTATAGGAATTGACTCAAG  
 GAAACGGACTATGTCTTGTACTGAATACTGTCTTCATTTGCCCATAGAATGCATTAGTGT  
 TACTCTTCTTCGCGAGCAGGTAAGTATGCGCGTAATGTTTTATTCTCTGAAAGGTTTC  
 AAAGTATGCAGAACAAATTAATGTTTGTCTTTTATTAAGCAAGCTACTTCTTGACAAGT  
 ACCCGATTGCTTCTTTAGCTCGGAACAAGACGCAAACTACAAAATATAGGTGCTGAAAAA  
 10 30 50  
 ATGAGGCAGGTTTGGTTCTCTTGGATTGTGGGATTGTTCTATGTTTTTCAACGTGTCT  
 1 M R Q V W F S W I V G L F L C F F N V S  
 70 90 110  
 TCTGCTGCCCAATACGAGCCACCTGCGACTTGGGAGAATGTTGATTATAAGAGGACAATA  
 21 S A A O Y E P P A T W E N V D Y K R T I  
 130 150 170  
 GACGTGTCAAACGCTTATATTTTCAGAAACAATCGAAATAACTATCAAAAACATAGCAAGC  
 41 D V S N A Y I S E T I E I T I K N I A S  
 190 210 230  
 GAACCTGCGACTGAATACTTACAGCCTTTGAGAGTGGCATCTTCAGTAAAGTTTCCCTTT  
 61 E P A T E Y F T A F E S G I F S K V S F  
 250 270 290  
 TTTTCAGCCTATTTTACCAACGAGGCAACTTTTTAAATAGCCAATTACTTGCCAATTTCG  
 81 F S A Y F T N E A T F L N S Q L L A N S  
 310 330 350  
 ACTACAGCACCTGGTGACGATGGTGAAAGTGAAATTAGATACGGGATCATTCAATTTCCA  
 101 T T A P G D D G E S E I R Y G I I Q F P  
 370 390 410  
 AATGCAATTTCCCTCAGGAAGAAGTTTCTTTAGTGATTAAAGAGCTTCTATAATACCGTA  
 121 N A I S P Q E E V S L V I K S F Y N T V  
 430 450 470  
 GGTATTCCTTATCTGAGCAGTGTGAATGTGAGAAGAACAACCTATTGTGGGAAACG  
 141 G I P Y P E H V G M S E E Q H L L W E T  
 490 510 530  
 AACAGATTGCCGCTTTCTGCTTACGATACCAAGAAGGCCTTTTACGCTGATTGGTAGC  
 161 N R L P L S A Y D T K K A S F T L I G S  
 550 570 590  
 TCATCATTTGAGGAGTACCACCCCCAAATGACGAGAGTTTACTGGGAAAAGCTAATGGA  
 181 S S F E E Y H P P N D E S L L G K A N G  
 610 630 650  
 AACTCTTTTGTAGTTTGGGCCTTGGGAAGATATTCGAGATTTTCTTCAAACGAAACCTTA  
 201 N S F E F G P W E D I P R F S S N E T L  
 670 690 710  
 GCAATGTGTTTATCCCAACAATGCCCATTTGAATCAGGTAGTGAATTTGAGAAGAGATATT  
 221 A I V Y S H N A P L N Q V V N L R R D I  
 730 750 770  
 TGGCTTTCTCATTGGGCTTCCACAATACAATTTGAGGAATATTATGAATTAACAAACAAA  
 241 W L S H W A S T I Q F E E Y Y E L T N K  
 790 810 830  
 GCCGAAAACCTGTCTAAAGGATTTTCAAGATTAGAATTAATGAACAGATTCAAACCTCAG  
 261 A A K L S K G F S R L E L M K Q I Q T Q  
 850 870 890  
 AATATGAGACAACTCACTTGTACTGTCTTAGACATGCTCCTGCCTGAGGGAGCTACT  
 281 N M R O T H F V T V L D M L L P E G A T  
 910 930 950  
 GATCATTATTTCACTGATTTGGTTGGCCTTGTTCACGTCGCATGCAGAACGTGACCAT  
 301 D H Y F T D L V G L V S T S H A E R D H  
 970 990 1010  
 TTCTTTATAAGACCAAGATTCCCAATCTTTGGAGGTGGAACATAATTTTACTGTCTGGT  
 321 F F I R P R F P I F G G W N Y N F T V G  
 1030 1050 1070  
 TGGACTAATAAATTGTCGATTTCTTGCATGTATCTCTGGCTCAGACGAGAAATTCGTT  
 341 W T N K L S D F L H V S S G S D E K F V  
 1090 1110 1130  
 GCTTCTATCCCAATTCTAAACGGCCACCGGACACTGTATATGATAATGTTGAATTATCG  
 361 A S I P I L N G P P D T V Y D N V E L S  
 1150 1170 1190  
 GTATTTCTTCCGGAAGGGCCGAAATATTCGATATTGATTCTCCAGTCCCTTTTACAAAT  
 381 V F L P E G A E I F D I D S P V P F T N  
 1210 1230 1250  
 GTTCTATAGAAACCCAGAAATCATACTTTGACCTAAATAAAGGTACGTTAAATTAAT  
 401 V S I E T Q K S Y F D L N K G H V K L T  
 1270 1290 1310  
 TTCAGTTACAGAAATTTGATTAGTCAAGTTGCCAATGGCCAAGTCTTGATAAAGTACGAC  
 421 F S Y R N L I S Q V A N G Q V L I K Y D  
 1330 1350 1370  
 TACCCGAAAAGCTCTTTTTCAAGAAGCCTCTGTCTATTGCTTGCTATATTTTACCAGCA  
 441 Y P K S S F F K K P L S I A C X Y I F T A  
 1390 1410 1430  
 CTAATGGGAGTTTTTGTCTTAAAACTTTGAACATGAACGTAACCTAAGTACGTTACGTT  
 461 L M G V F V L K T L N M N V T N \*  
 TGAATGATATATACCTTATGATGCAGAGCAATAAATTCAGTATTTTATAAGTT  
 TTATAGATAAGAATTTTCAATGAATAAAGGTATTTACAATCATTTCTAGACATTCTCG  
 TGGTCTTTTCTTGCACAATCAATGGTCATAGTTTTTTTGTAGTTTAAAGGACAG  
 AAGAATAACAAAGATGTTACGTTTTGAATGTTACCAGACATTCTCAGATTTCCTTTA  
 ACCTAGCGACGCGCACAGGCTCCCAATCCTCTCAGAAAGACCTTGAACGATAATATT

**B**

Fig. 2 (continued).

Human	NEAPAAAGLFLLLLLGTHAFAPGSASSEAPP	30
Rat	NEAPI--VLLLLLVLLALAPTTPGSASSEAPP	28
Nlt1p	NRQVWFSSWIVGLFLCFPNVSSAAQYEPFAT	30
Human	LYNEDVKRTVDLSSHLAKVTAEVVLAHLGG	60
Rat	LVEDVKRTVDLSSHLARVTAEVVLAHLGG	58
Nlt1p	WEHVDVTKRTIDVSNAYISRTITIKNIAS	60
Human	GSTSRATSFLLALEFELRLAHLGVQVQKG	90
Rat	GSTARAISFVLALAEPELESRLAHLGVQVQKG	88
Nlt1p	EPF--ATDYFATAFSGIVSKVSPFSAYFTN	87
Human	KDR--EENMLEVRETRIKGKSG--RPFPTV	115
Rat	EDH--EDNMLEVRETRIKGKSG--RPFPTV	113
Nlt1p	EATFLNSQLLANSITATAGDDGSESIKCTI	117
Human	KLPVALDPGAKISVIVETVYTHVLPYPTQ	145
Rat	KLPVALDPGSKISIVETVYTHVLPYPTQ	143
Nlt1p	QFPFMAISPGQEVSLVSKSFYNTVGCIPYPER	147
Human	ITQSEKQFVVFEGNHYFYSFYPTKTQTMRV	175
Rat	ITQSEKQFVVFEGNHYFYSFYPTKTQTMRV	173
Nlt1p	VIGHSEELQHLLEWETNRCLPLSAYDITKASFTL	177
Human	KLASRNVESYTKLGNPFT--RSEDLLDYG	201
Rat	RLASRNVESHTKLGNPST--RSEDILLDYG	199
Nlt1p	-YGSSTFEEYHFFNDLSLGLKANGNSFEFG	206
Human	FFRDPVAV--SQDTFKVHYENNSPFLTITSM	230
Rat	FFKDIPAV--SQDTFKVHYENNSPFLTITSM	228
Nlt1p	PWEIDTPRFSSMETLAVIYSHNAPLNQVNVNL	236
Human	TRVIEVSHWGN--IAVEENVDLKHTGAVLKG	259
Rat	TRVIEVSHWGN--IAVEENVDLKHTGAVLKG	257
Nlt1p	NRDITWLSHWASTITQFEYVYELTNKAALISK	266
Human	FFSKYDYQRPDS--GISSIRSPFKTILP	285
Rat	FFSKYDYQRPDS--GISSIRSPFKTILP	283
Nlt1p	GFSRDELRKQITQNMROTHFVITVLDMHL	296
Human	AAAQDVYRDEIGNVSTSHLLILDDSVEME	315
Rat	AAAQDVYRDEIGNVSTSHLLILDDSVEME	313
Nlt1p	EGATDHYPTDVLGVSTSH--AEHNDHFF	322
Human	IRPRFPLFGGKTHYIVGYNLPSYEYLY--	343
Rat	IRPRFPLFGGKTHYIVGYNLPSYEYLY--	341
Nlt1p	IRPRFPLFGGKNYNYFTVGM--TNKLSDFLHVS	352
Human	-NLGDDQYALKMRFDVHVFDEQVIDSLTVKI	372
Rat	-NLGDDQYALKMRFDVHVFDEQVIDSLTVKI	370
Nlt1p	SGSDLRKIFVASIPILNGP-PDITVVDNVELSV	381
Human	ILPEGAKNIEIDSPYKISRAPDELHYTYLD	402
Rat	ILPEGAKNIEIDSPYDISRAPDELHYTYLD	400
Nlt1p	FLPEGAETFDIDSPVFFTNVSIETQKSYFD	411
Human	TFGRFPVIVAYK--KNLVEQHIQ--DIVVHYTF	430
Rat	TFGRFPVIVAYK--KNLVEQHIQ--DIVVHYTF	428
Nlt1p	LNGGRVVKLTFSYRNLSIQVANGGVVQVYDY	441
Human	NKVLNLQEPFLLVAAFPYILFFTVIIVVRLD	460
Rat	NKVLNLQEPFLLVAAFPYILFFTVIIVVRLD	458
Nlt1p	PKSSFPKPLPSIACYFPTALHGVVFLKTLN	471
Human	FSITKDPAAEARMKVACITEQVLTTLVNKRRI	490
Rat	FSITKDPAAEARMKVACITEQVLTTLVNKRRL	488
Nlt1p	NNVTN--	476
Human	GLYRHFDFTVNRYKQSRDISTLNSGKKSLE	520
Rat	GLYRHFDFTVNRYKQSRDISTLNSGKKSLE	518
Nlt1p	---	476
Human	TEHKA--LTSEIAL--LQSRLLKTEGSDLCDRVSE	550
Rat	TEHKA--LTSEIAL--LQSRLLKTEGSDLCDRVSE	548
Nlt1p	---	476
Human	NOKLDAQVKELVLKSAVEAERLVAGKLRKD	580
Rat	NOKLDAQVKELVLKSAVEAERLVAGKLRKD	578
Nlt1p	---	476
Human	TYIENEKLS--SGKROELVTKIDHILDAL	607
Rat	TYIENEKLS--SGKROELVTKIDHILDAL	605
Nlt1p	---	476

product was isolated and cloned; a partial sequence of this clone was consistent with the three sequences obtained from the 64 kDa glycoprotein. Probes synthesized from the 850 bp PCR product as a template were used to screen a  $\lambda$ YES yeast genomic library [22]. One hybridizing  $\lambda$  phage with a 3.7 kb genomic insert was selected for further study.

The sequence of the 3.7 kb region revealed one long 1.4 kb potential open reading frame (see Fig. 2A) and one shorter potential open reading frame; the long open reading frame was assigned to the *NLT1* gene. Approximately 1 kb of the sequence upstream of the *NLT1* open reading frame is 98% identical to the nucleotide sequence upstream of the previously sequenced *PRE3* locus [25]. The *PRE3* locus contains the entire sequence of *CEN10*, thereby positioning the start of the *NLT1* open reading frame 1.4 kb downstream of *CEN10* on the right arm of chromosome ten of *Saccharomyces cerevisiae*. In addition, the final 400 bp of the 3.7 kb genomic clone are 100% identical to 400 bp from the previously sequenced *CDC35* locus [26], placing the *CDC35* locus further downstream of the *NLT1* gene also on the right arm of chromosome ten.

### 3.2. The *NLT1* gene encodes the 64 kDa glycoprotein

The 1.4 kb open reading frame of the novel *NLT1* gene encodes a 476 amino acid 54 kDa polypeptide (see Fig. 2B). The amino acid sequence includes the exact sequences of peptides 64-1 and 64-2, and it includes all but two of the amino acids of peptide 64-3 which was sequenced at the limits of detection. The sequence matching peptide 64-3 is found immediately upstream of the sequence matching peptide 64-2; evidently, the cyanogen bromide fragments co-migrated when analyzed by SDS-PAGE because of incomplete cleavage at Met-193, the first residue of cyanogen bromide fragment 64-2. With the exception of two unclear residues out of 44 sequenced, the predicted sequence of Nlt1p from the yeast genome exactly matches that obtained from the 64 kDa glycoprotein subunit of the oligosaccharyl transferase. Based on extensive agreement between the translation of the *NLT1* gene and the amino acid sequences obtained from the 64 kDa glycoprotein, we conclude that Nlt1p is the 64 kDa glycoprotein subunit of the oligosaccharyl transferase.

Hydropathy analysis [27] of the Nlt1p amino acid sequence reveals two potential transmembrane hydrophobic regions. At the carboxy-terminus is a potential transmembrane domain followed by a short soluble tail. At the amino-terminus is a 22 residue signal peptide cleaved by the endoplasmic reticulum signal peptidase. The amino-terminal amino acid sequence for the 64 kDa glycoprotein (see Fig. 1) begins at residue 23 of the sequence derived from the yeast genome. In addition, a predictive weight matrix for the site of signal peptidase cleavage [28] indicates that the most likely site of signal peptidase cleavage is between residues 22 and 23 of the unprocessed protein. The predicted structure of the Nlt1p is that of a type I membrane

←

Fig. 3. Nlt1p is homologous to the mammalian ribophorin I. The sequences of the yeast Nlt1p and the human and rat ribophorin I proteins [12,13] were compared with the program PRETTYPLOT (Computer Group, European Molecular Biology Laboratory, Heidelberg, Germany). Nlt1p shows significant amino acid similarity and identity to the amino-terminal domains of the mammalian proteins.

protein with a large 435 residue soluble domain anchored in the membrane by a carboxy-terminal hydrophobic domain.

The amino acid sequence includes six potential glycosylation acceptor sites (see Fig. 2B); one of these sites is in the amino-terminal signal sequence and would not be present in the mature glycoprotein. Glycosylation of the remaining five sites would yield a mature glycoprotein of 62–67 kDa. The observed doublet of bands near 64 kDa [5] would suggest that the mature Nlt1p is partially glycosylated and functions in vivo as two glycoforms with three or four oligosaccharides.

Nlt1p is conserved in higher eukaryotes. The protein shows 59% amino acid similarity and 32% amino acid identity to the mammalian ribophorin I protein from a human cell line [12]; it also shows 56% similarity and 30% identity to the rat ribophorin I [13] (see Fig. 3). The amino acid similarity extends throughout the sequence of Nlt1p; it is not concentrated in any one region of the protein. However, the mammalian ribophorins are approximately 130 amino acids larger at the carboxy-terminus than Nlt1p. Whereas the mammalian ribophorin I proteins have extensive soluble domains on either side of a potential transmembrane helix, the yeast Nlt1p includes only a single soluble domain on the amino-terminal side of a potential transmembrane anchor. The canine ribophorin I protein was purified as a subunit of the mammalian oligosaccharyl transferase complex, and the other two identified components of the yeast enzyme show homology to the other two subunits of the mammalian enzyme. Hence, the oligosaccharyl transferase is conserved throughout evolution from yeast to mammals.

### 3.3. The *NLT1* gene product is essential for vegetative growth

Yeast strain RPY5 was constructed with one intact and one partially deleted copy of the *NLT1* gene. To construct this strain, the 800 bp fragment between the *StuI* and *MscI* sites within the coding region of the *NLT1* gene was replaced with a 1.1 kb DNA fragment containing the selectable *URA3* gene. Southern blotting confirmed that one genomic copy of the *NLT1* gene was intact and the other contained a stable deletion and insertion (data not shown). The diploid strain was sporulated, and tetrads were dissected; segregants with the *NLT1* deletion were not viable, indicating that the gene is essential for vegetative growth of yeast. Spores containing the deletion germinated and grew to colonies of only 2–4 cells, and spores carrying the deletion and a copy of the *NLT1* gene on an episomal plasmid were viable.

## 4. Discussion

Here we report the cloning and characterization of the yeast *NLT1* gene. This gene encodes the 64 kDa glycoprotein subunit of the oligosaccharyl transferase complex, and it is the third oligosaccharyl transferase gene to be cloned from yeast. The *WBPI* gene encodes the 45 kDa subunit of the complex [9], and the *SWPI* gene encodes the 32 kDa subunit [8]. Preliminary experiments indicate that simultaneous expression of the three genes does not yield a significant increase in transferase activity (R.P. and B.L., unpublished data); evidently, characterization of the enzyme will require the cloning and expression of at least one more gene responsible for oligosaccharyl transferase activity. Previous studies have isolated the enzyme as a tetrameric [5,6] and a hexameric [7] complex of polypeptides. The genes

encoding the 34 kDa and perhaps the 16 kDa and 9 kDa polypeptides which co-purified with oligosaccharyl transferase activity should complete the genetic description of the enzyme.

The amino acid sequence of the 54 kDa Nlt1p is consistent with that of a type I transmembrane protein. A soluble amino-terminal domain comprises most of the protein. This domain is followed by a potential transmembrane helix and a short soluble carboxy-terminal tail. The amino-terminal domain contains four potential glycosylation sites which are consistent with the 64 kDa apparent mobility of the mature glycopeptide and the tight binding of the intact oligosaccharyl transferase complex to a concanavalin A affinity matrix [5]. Both Wb1p and Sw1p have their extensive amino-terminal soluble domains in the lumen of the endoplasmic reticulum and carboxy-terminal membrane anchors; Wb1p has one potential transmembrane domain while Sw1p has three [8,9]. If Nlt1p has a similar orientation relative to the endoplasmic reticulum membrane, then a model of the oligosaccharyl transferase emerges as that of a large, complex luminal domain anchored in the membrane by five or more transmembrane domains. These transmembrane domains could be involved in the translocation of nascent polypeptides from the ribosome across the endoplasmic reticulum membrane or in the recognition of the dolicholpyrophosphoryl oligosaccharide substrate of the enzyme. Although the potential transmembrane domain of Nlt1p does not show homology to the putative dolichol binding consensus amino acid sequence [29], recent studies have shown that the residues of this consensus are not necessary for dolichol binding by the yeast dolichyl-phosphate-mannose synthase [30].

The oligosaccharyl transferase enzyme is conserved throughout eukaryotic evolution from yeast to mammals. The Wb1p subunit is homologous to the 48 kDa subunit of the canine pancreas enzyme [11], and the Sw1p subunit is homologous to the carboxy-terminal half of the ribophorin II subunit of the mammalian enzyme [7]. In addition, the novel Nlt1p is significantly similar to the mammalian ribophorin I protein, the final component of the purified mammalian complex. The homology of the 54 kDa Nlt1p is limited to the amino-terminal domain of the mammalian ribophorin I; the mammalian protein has an additional 130 amino acid carboxy-terminal soluble domain not found in the yeast protein (see Fig. 3). The mammalian oligosaccharyl transferase might have evolved to include function on the cytoplasmic side of the endoplasmic reticulum membrane, and perhaps one of the remaining subunits of the yeast enzyme will include a more extensive carboxy-terminal domain than the three characterized polypeptides. Additionally, none of the five potential glycosylation acceptor sites in the yeast Nlt1p is conserved in the mammalian proteins, and the mammalian proteins are glycosylated at two acceptor sites which are not found in Nlt1p (see Fig. 3). Yeast glycoproteins tend to be more extensively glycosylated than higher eukaryotic proteins; hence, it is not surprising that Nlt1p has more glycosylation sites than the mammalian homologues. However, it is surprising that none of the glycosylation acceptor sites are conserved. This suggests that the oligosaccharides contribute more to the overall folding and solubility of glycoproteins and to their intracellular targeting than they do to their biochemical function.

The cloning of the yeast *NLT1* gene, together with the *WBPI* and *SWPI* genes, furthers the genetic understanding of the complex and essential oligosaccharyl transferase enzyme. These three genes and the genes that encode the remaining

subunits of the enzyme will enable the biochemical examination of the N-linked glycosylation of polypeptides in the endoplasmic reticulum of yeast cells.

**Acknowledgements:** We thank Markus Aebi for providing the *WBPI* and *SWPI* genes and useful discussion; the laboratories of Phillips Robbins, John Abelson, Judith Campbell, and Alexander Varshavsky for plasmids and yeast strains; Ronald Davis for the yeast genomic library; and Ebrahim Zandi for useful discussion. Protein sequencing was done by Dirk Krapf, and nucleic acid sequencing was done by Steve Marsh and Enrique Colayco. R.P. was supported by a National Science Foundation predoctoral fellowship (1991–1994), and this research was supported by National Institutes of Health Grant GM39334 to B.I.

## References

- [1] Herscovics, A. and Orlean, P. (1993) *FASEB J.* 7, 540–550.
- [2] Kukurzinska, M.A., Bergh, M.L.E. and Jackson, B.J. (1987) *Annu. Rev. Biochem.* 56, 915–944.
- [3] Tanner, W. and Lehle, L. (1987) *Biochim. Biophys. Acta* 906, 81–99.
- [4] Gavel, Y. and von Heijne, G. (1990) *Protein Eng.* 3, 433–442.
- [5] Pathak, R., Hendrickson, T.L. and Imperiali, B. (1995) *Biochemistry*, in press.
- [6] Knauer, R. and Lehle, L. (1994) *FEBS Lett.* 344, 83–86.
- [7] Kelleher, D.J. and Gilmore, R. (1994) *J. Biol. Chem.* 269, 12908–12917.
- [8] te Heesen, S., Knauer, R., Lehle, L. and Aebi, M. (1993) *EMBO J.* 12, 279–284.
- [9] te Heesen, S., Janetzky, B., Lehle, L. and Aebi, M. (1992) *EMBO J.* 11, 2071–2075.
- [10] Kelleher, D.J., Kreibich, G. and Gilmore, R. (1992) *Cell* 69, 55–65.
- [11] Silberstein, S., Kelleher, D.J. and Gilmore, R. (1992) *J. Biol. Chem.* 267, 23658–23663.
- [12] Crimaudo, C., Hortsch, M., Gausepohl, H. and Meyer, D.I. (1987) *EMBO J.* 6, 75–82.
- [13] Harnik-Ort, V., Prakash, K., Marcantonio, E., Coleman, D.R., Rosenfeld, M.G., Adesnik, M., Sabatini, D.D. and Kreibich, G. (1987) *J. Cell Biol.* 104, 855–863.
- [14] Wiederrecht, G., Seto, D. and Parker, C.S. (1988) *Cell* 54, 841–853.
- [15] Sikorski, R.S. and Hieter, P. (1989) *Genetics* 122, 19–27.
- [16] Guthrie, C. and Fink, G.R. (1991) *Guide to Yeast Genetics and Molecular Biology*, Academic Press, San Diego, CA.
- [17] Wessel, D. and Flügge, U.I. (1984) *Anal. Biochem.* 138, 141–143.
- [18] Laemmli, U.K. (1970) *Nature* 227, 680–685.
- [19] Towbin, H., Staehelin, T. and Gordon, V. (1979) *Proc. Natl. Acad. Sci. USA* 76, 4350–4354.
- [20] Schagger, H. and von Jagow, G. (1987) *Anal. Biochem.* 166, 368–379.
- [21] Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.
- [22] Ramer, S.W., Elledge, S.J. and Davis, R.W. (1992) *Proc. Natl. Acad. Sci. USA* 89, 11589–11593.
- [23] Elledge, S.J., Mulligan, J.T., Ramer, S.W., Spottswood, M. and Davis, R.W. (1991) *Proc. Natl. Acad. Sci. USA* 88, 1731–1735.
- [24] Schiestl, R.H. and Gietz, R.D. (1989) *Curr. Genet.* 16, 339–346.
- [25] Enenkel, C., Lehman, H., Kipper, J., Gückel, R., Hilt, W. and Wolf, D.H. (1994) *FEBS Lett.* 341, 193–196.
- [26] Masson, P., Lenzen, G., Jacquemin, J.M. and Danchin, A. (1986) *Curr. Genet.* 10, 343–352.
- [27] Kyte, J. and Doolittle, R.F. (1982) *J. Mol. Biol.* 157, 105–132.
- [28] von Heijne, G. (1986) *Nucleic Acids Res.* 14, 4683–4690.
- [29] Albright, C.F., Orlean, P.O. and Robbins, P.W. (1989) *Proc. Natl. Acad. Sci. USA* 86, 7366–7369.
- [30] Zimmermann, J.W. and Robbins, P.W. (1993) *J. Biol. Chem.* 268, 16746–16753.