Glycosylation of the overlapping sequons in yeast external invertase: effect of amino acid variation on site selectivity in vivo and in vitro

Anthony Reddy, Barbara S. Gibbs, Yun-Li Liu, James K. Coward, Li-Ming Changchien and Frank Maley

Wadsworth Center, New York State Department of Health, Albany, NY 12201–0509, USA and the Interdepartmental Program in Medicinal Chemistry, College of Pharmacy, and Department of Chemistry, University of Michigan, Ann Arbor, MI 48109–1053, USA

Received on August 14, 1998; revised on September 28, 1998; accepted on September 28, 1998

Yeast invertase contains 14 sequons, all of which are glycosylated to varying degrees except for sequon 5 which is marginally glycosylated, if at all. This sequon overlaps with sequon 4 in a sequence consisting of Asn92–Asn93–Thr94–Ser95 (Reddy et al., 1988, J. Biol. Chem., 263, 6978–6985). To determine whether glycosylation at Asn93 is sterically hindered by the oligosaccharide on Asn92, the latter amino acid was converted to a glutamine residue by site-directed mutagenesis of the SUC2 gene in a plasmid vector which was expressed in Saccharomyces cerevisiae. A glycopeptide encompassing sequons 3 through 6 was purified from a tryptic digest of the mutagenized invertase and sequenced by Edman degradation, which revealed that Asn93 of sequon 5 contained very little, if any, carbohydrate, despite the elimination of sequon 4. When Ser and Thr were inverted to yield Asn–Asn–Ser–Thr carbohydrate was associated primarily with the second sequon, in agreement with numerous studies indicating that Asn-X-Thr is preferred to Asn-X-Ser as an oligosaccharide acceptor. However, when the invertase overlapping sequons were converted to Asn–Asn–Ser–Ser, both sequons were clearly glycosylated, with the latter sequon predominating. These findings rule out steric hindrance as a factor involved in preventing the glycosylation of sequon 5 in invertase. Comparable results were obtained using an in vitro system with sequon-containing tri- and tetrapeptides acceptors, in addition to larger oligosaccharide acceptors.

Key words: invertase glycosylation/overlapping sequons/glycosylation of peptides in vitro/sequon mutagenesis/oligosaccharyltransferase

Introduction

The use of small sequon (Asn-X-Ser/Thr)-containing peptides as oligosaccharide acceptors in an in vitro glycosylation system has been instructive in revealing those amino acids (X) that support glycosylation better than others, as well as demonstrating that threonine plays a more positive role in this process than serine (Ronin et al., 1981; Bause, 1984; Kornfeld and Kornfeld, 1985; Kaplan et al., 1987; Kakuruzinska et al., 1987; Gavel and von Heinje, 1990). However, when sequon glycosylation is considered in the context of an intact protein the picture becomes more complex, and is even more so on comparing the extent of glycosylation of each sequon in a protein containing several sequons. This was clearly shown in the case of yeast external invertase (EC3.2.1.26), an enzyme that contains 14 sequons per subunit (Reddy et al., 1988; Ziegler et al., 1988). As shown in these studies, each sequon is glycosylated to a variable extent, except for the overlapping sequons 4 and 5 (Asn–Asn–Thr–Ser) where sequon 4 (overlaid) is fully glycosylated, while sequon 5 (underlined) is glycosylated to a minimal degree. Basically, the degree of glycosylation of each sequon in yeast external invertase is consistent with a statistical compilation of over 600 glycosylation sites in proteins (Gavel and von Heinje, 1990), with some slight exceptions. The fact that sequon 5 of invertase is marginally glycosylated might be due in part to steric factors, as proposed for the case of closely spaced sequons (Gavel and von Heinje, 1990), or to the presence of Ser in the third position of sequons in relatively short peptides (Ronin et al., 1981; Bause, 1984; Kaplan et al., 1987). Another factor influencing sequon glycosylation is the presence of specific amino acids in the (X) position of sequons in such glycoproteins as the rabies virus glycoprotein (Shakin-Eshelman et al., 1996; Kasturi et al., 1997), where Asn-(X)-Ser sequences with Trp, Asp, Glu, and Leu in (X) were poorly glycosylated in vitro, relative to a similar series of sequons containing Thr in the third position. However, it may not be possible to extrapolate from the in vitro system to what occurs in vivo, since glycosylation occurs at the Asn of a sequon as the newly synthesized peptide passes into the lumen of the endoplasmic reticulum some 12–14 amino acids from the membrane anchor (Nilsson and von Heinje, 1993). At this distance some protein folding can occur, depending on the amino acids associated with this peptide, and might influence the accessibility of the sequon to oligosaccharyl transferase. In fact, this has been shown to be the case for carboxypeptidase Y where peptide folding and glycosylation compete, with the net result being that the more folded structures are glycosylated to a lesser extent (Holst et al., 1996).

So, while in vitro studies measuring the rate and extent of glycosylation of peptides and proteins provide useful information, they may not truly reflect what occurs in vivo. As pointed out by Sheltikoff et al. (1996), the amount of carbohydrate associated with a sequon in a protein depends on the rate of mRNA translation, the rate of protein folding, and the availability of dolichyl pyrophosphate-oligosaccharides. In addition, when sequons are in proximity to one another the first to be glycosylated might interfere sterically with the glycosylation of the second (Gavel and von Heinje, 1990). This situation appeared to be the case for the overlapping sequons 4 and 5, of yeast invertase (Asn92–Asn93–Thr94–Ser95), where the first sequon was almost completely glycosylated, but the second was barely glycosylated, if at all (Reddy et al., 1988). Although the presence of overlapping sequons in glycoproteins is a rare event, the case of invertase provides an opportunity to understand how each sequon influences the glycosylation of the other in vivo and what
Fig. 1. Tryptic peptide map of the mutagenized invertases. Tryptic digests of the S-carboxymethylated invertase mutants (N92Q; T94S; T94S, S95T) (2 mg each) were resolved separately by HPLC on an Aquapore C8 reverse phase column (see Materials and methods). The peptides were eluted using 0–60% gradient of solvent B over 90 min at a flow rate of 0.6 ml/min. Solvent A, 0.1% CF3COOH; Solvent B, 90% CH3CN 0.1% CF3COOH. Aliquot of fractions (0.15 ml) was assayed for carbohydrate by the phenol-sulfuric acid method (Dubois et al., 1956). The pooled fractions (represented by the bars) were concentrated in a Speed Vac, and sequenced. The numbers over the bars represent the following pools and their included sequons: pool 1, sequon 11; pool 2, sequon 7; pool 3, sequons 3 to 6; pool 4, sequon 10; pool 5, sequon 12; pool 6, sequon 2 (see Reddy et al., 1998, for location of these sequons in invertase).

occurs when Thr94 and Ser95 are altered. We have quantitated the extent to which sequons 4 and 5 of invertase are glycosylated before and after mutation, something that is rarely done in studies measuring sequon glycosylation. In parallel with the in vivo studies, we have undertaken in vitro studies with chemically synthesized peptides containing sequons 4 and 5, and variants thereof, to determine whether these sequons are glycosylated in a manner similar to what has been found in vivo.

Results
Site-directed mutagenesis of sequon 4

The overlapping sequons 4 and 5, with putative glycosylation sites at Asn92 and Asn93, were shown earlier by both tryptic peptide sequencing and fast atom bombardment mass spectroscopy to contain the sequence Asn-Asn-Thr-Ser (Reddy et al., 1988), but only sequon 4 was glycosylated. Since a possible reason for the apparent absence of glycosylation at sequon 5 was that the oligosaccharide unit on sequon 4 interfered with the glycosylation of sequon 5, Asn92 was mutated to Gln, an amino acid related to Asn but one that cannot be N-glycosylated. The strategy used was to subclone the SUC2 structural gene from the shuttle vector pEI13 (Reddy and Maley, 1996) into M13mp19 and after mutagenesis, to reinsert the mutated gene into pEI13.

The original pEI13 vector contains the SUC2 gene from the parent vector pRB58, but trimmed to remove the noncoding 3′-sequence and modified to remove a HindIII site in the signal sequence of SUC2 (Reddy and Maley, 1990, 1996) so that only the external or glycosylated form of invertase could be expressed. This vector was then transfected into an invertase-deficient yeast strain using a Ca2+-lithium acetate-dithiothreitol procedure (Reddy and Maley, 1993).

Effect of Asn92Gln mutation on the glycosylation of sequon 5

The isolated invertase mutant was digested with TPCK-trypsin for 18 h, and the resulting peptides and glycopeptides were resolved on a C8 reverse-phase HPLC column (Figure 1). The glycopeptides were located in six pools by testing the elution profile for carbohydrate (Dubois et al., 1956). The fraction eluting at 48 min (pool 3) was found to encompass four putative glycosylation sites in sequons 3–6 (Reddy et al., 1988). Table I shows the results obtained on sequence analysis of untreated and PNGase-treated peptide. The presence of Gln at cycle 16 confirmed the site-directed mutation of Asn92 to Gln (N92Q). PTH-amino acids were not detected at cycles 2 (Asn78) and 23 (Asn99), indicating that these residues in sequons 3 and 6, respectively, were glycosylated. These results are in agreement with earlier studies which showed that glycosylated PTH-Asn is not extracted on Edman sequencing (Paxton et al., 1987). By
contrast, cycle 17 (Asn93) of the mutant yielded mainly Asn, signifying that the Asn of sequon 5 was not glycosylated, at least to a detectable degree. To rule out the possibility that Asn92 was glycosylated to some small extent, the peptide was digested with PNGase F and sequenced. Since PNGase F digestion results in the conversion of an oligosaccharide-bearing Asn to Asp (Plummer et al., 1984), one could expect an Asp at cycles 2 and 23, and indeed this was the case. However, cycle 17 (Asn93) yielded no detectable Asp, only Asn, indicating that despite the fact that sequon 4 had been abolished, Asn93 of sequon 5 was still not glycosylated. Thus, the presence of an oligosaccharide on sequon 4 was not apparently an impediment to the glycosylation of sequon 5.

Effect of serine and threonine alteration in sequons 4 and 5 on their extent of glycosylation

Since impairing the glycosylation of sequon 4 by converting Asn92 to a Gln did not enhance the glycosylation of sequon 5, another potential explanation for the absence of carbohydrate was sought based on the fact that serine in the sequon Asn-X-Ser has been shown to be much less effective than threonine in supporting sequon glycosylation (Bause and Legler, 1981; Bause, 1984; Kaplan et al., 1987; Kasturi et al., 1995; Picard et al., 1995). To investigate this possibility, the positions of serine and threonine in sequons 4 and 5 were interchanged by site-directed mutagenesis and the isolated invertase was digested with trypsin to yield a peptide containing sequons 4 and 5 (Figure 1, pool 3) in addition to sequon 3. To obtain higher yields and more quantitative results sequon 3 was removed from this glycopeptide by chymotrypsin treatment and the resulting peptide was isolated by HPLC (Figure 2, S4.5). On sequencing the isolated chymotryptic peptide, it was found to be seven residues shorter than prior to treatment (Table II). In those cases where 100% N-glycosylation occurred, no PTH amino acid would be evident (Paxton et al., 1987), whereas partial glycosylation could be determined by the amount of PTH-Asn observed. An even more accurate estimate of the extent of sequon glycosylation in the tryptic and chymotryptic peptides was obtained by treating them with PNGase F as described above for analysis of the glycopeptide from the N92Q mutant (Table I). As shown in Table II and III, Asn93 of sequon 5 in -NN93TS- and -QN93TS- was only marginally glycosylated, which is consistent with the more qualitative results in Table I. However, on converting NNTS to NNST the results were reversed in that Asn92 of sequon 4 was poorly glycosylated while Asn93 of sequon 5 was heavily glycosylated (Table II, III), a finding consistent with what was found for the overlapping NNST sequons of human serum cholinesterase (Lockridge et al., 1987). These findings emphasize again that a sequon ending in Thr is readily glycosylated in most cases, while a sequon ending in Ser is not. It should be emphasized, though, that there are exceptions to this rule, as shown in the case of invertase where sequon 3 (NDS) is glycosylated as well as, if not better than sequon 2 (NDT), and sequon 10 (NIS) is fully glycosylated (Reddy et al., 1988).

Table I. Sequence analysis of a tryptic peptide from external invertase containing sequons 3–6, but with sequon 4 mutated

<table>
<thead>
<tr>
<th>Cycle number</th>
<th>Peptide</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
<th>17</th>
<th>18</th>
<th>19</th>
<th>20</th>
<th>21</th>
<th>22</th>
<th>23</th>
<th>24</th>
</tr>
</thead>
<tbody>
<tr>
<td>Untreated‡</td>
<td>R—D</td>
<td>S</td>
<td>G</td>
<td>A</td>
<td>F</td>
<td>S</td>
<td>G</td>
<td>S</td>
<td>M</td>
<td>V</td>
<td>V</td>
<td>D</td>
<td>Y</td>
<td>Q</td>
<td>N</td>
<td>T</td>
<td>S</td>
<td>G</td>
<td>F</td>
<td>F</td>
<td>—</td>
<td>D</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PNGase treated§</td>
<td>D</td>
<td>D</td>
<td>S</td>
<td>G</td>
<td>A</td>
<td>F</td>
<td>S</td>
<td>G</td>
<td>S</td>
<td>M</td>
<td>V</td>
<td>V</td>
<td>D</td>
<td>Y</td>
<td>Q</td>
<td>N</td>
<td>T</td>
<td>S</td>
<td>G</td>
<td>F</td>
<td>F</td>
<td>D</td>
<td>D</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

‡This peptide represents the mutated invertase sequence (QNTS) encompassing residues 77 to 105 that was isolated from pool 3 (in vivo) in Figure 1. Sequence analysis was performed on 0.5 nmol of the peptide. See Materials and Methods for additional details. The Asn to Gln change at position 92 (cycle 16) is underlined; Asn 93 is double underlined, to indicate that it was glycosylated only slightly, if at all. The absence of amino acids (—) at cycles 2 and 23, (sequons 3 and 6) indicates that the Asn residues at these two positions were glycosylated.

§Glycosylation at positions 2 and 23 was verified by PNGase F treatment, which yields an Asp at N-glycosylated positions.
### Table II. Quantitation of N-glycosylation at sequons 4 and 5 in native and mutant invertase

<table>
<thead>
<tr>
<th>Peptide</th>
<th>Location in invertase</th>
<th>Sequon</th>
<th>Amino acid sequence (pmol)*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tryptic</td>
<td>77–105</td>
<td>NNTS</td>
<td>V(607) D(240) Y(354) N(15) T(30) S(0)</td>
</tr>
<tr>
<td>Tryptic + PNGase F</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chymotryptic + PNGase F</td>
<td></td>
<td>NNST</td>
<td></td>
</tr>
<tr>
<td>Chymotryptic + PNGase F</td>
<td></td>
<td>NNSS</td>
<td></td>
</tr>
</tbody>
</table>

*Only the quantitative sequence data from amino acid residues 88–94 are presented. The numbers in parentheses represent the amount in picomoles, of each amino acid released as its PTH-amino acid.

### Table III. Estimated extent of glycosylation of invertase sequons based on data in Table II

<table>
<thead>
<tr>
<th>Normal and modified sequons 4 and 5</th>
<th>Sequon</th>
<th>Amount glycosylated (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>NNTS</td>
<td>(4) NNT</td>
<td>96</td>
</tr>
<tr>
<td>QNTS</td>
<td>(5) NTS</td>
<td>&lt;5</td>
</tr>
<tr>
<td>QNTS + PNGase F</td>
<td>(5) NTS</td>
<td>0</td>
</tr>
<tr>
<td>NNST</td>
<td>(4) NNS</td>
<td>&lt;7</td>
</tr>
<tr>
<td>NNST + PNGase F</td>
<td>(5) NST</td>
<td>81</td>
</tr>
<tr>
<td>NNSS</td>
<td>(4) NNS</td>
<td>6</td>
</tr>
<tr>
<td>NNSS + PNGase F</td>
<td>(5) NSS</td>
<td>69</td>
</tr>
<tr>
<td>NNSS</td>
<td>(4) NNS</td>
<td>44</td>
</tr>
<tr>
<td>NNSS + PNGase F</td>
<td>(5) NSS</td>
<td>72</td>
</tr>
<tr>
<td>NNSS</td>
<td>(4) NNS</td>
<td>33</td>
</tr>
<tr>
<td>NNSS + PNGase F</td>
<td>(5) NSS</td>
<td>69</td>
</tr>
</tbody>
</table>

*Data from (Reddy et al., 1988).

The yields of the amino acids in sequons (4) and (5) were estimated by comparison to the yields of the PTH-amino acids on sequencing a chemically synthesized peptide corresponding to residues 82–98 of yeast invertase. While the estimated percentage of each sequon that is glycosylated is only an approximation, these values compare favorably with the more accurate values obtained on PNGase F treatment in the line below.

Based on these results and the findings of others (Bause and Legler, 1981; Welplhy et al., 1983; Kasturi et al., 1995; Picard et al., 1995) in both in vitro and in vivo studies, it was expected that both sequons in NNSS would be poorly glycosylated relative to NNTS and NNST. However, as shown in Table II and III the first sequon is glycosylated fairly well (~30%) while the second is even better (~70%). It should be noted however, that these data address only the extent of intracellular glycosylation of invertase at these sequons and reveal nothing about their rates of glycosylation.

### Isolation and characterization of glycopeptides labeled in vitro

To determine whether the glycosylation products observed in vivo could be reproduced in vitro, OST and [3H]LOS or its truncated version [3H]LDS, were incubated with the overlapping sequon-containing heptadecapeptide 1, derived from invertase residues 82–98, or the mutant heptadecapeptides 2 and 3, in an attempt to obtain [3H]-labeled glycopeptide. The sequons employed are underlined, with the sequons in peptides 2 and 3 being variants of the normal sequon in peptide 1.

1. AFSGSMVVVDYNNTSGFF
2. AFSGSMVVVDYQNTSGFF
3. AFSGSMVVVDYNQTSGFF

These peptides were compared with Bz-Asn-Leu-Thr-NH₂ as acceptors of [3H]LOS. The strategy for isolating the oligosaccharide-containing glycopeptide in a form suitable for amino acid sequencing involved purifying it by binding and elution from ConA-Sepharose, followed by cleavage with thermolysin. However, preliminary results suggested that the use of larger amounts of unlabeled LOS would be necessary to identify the site of glycosylation in peptide 1, a problem that was resolved by the development of a procedure for isolating LOS from bovine pancreas in reasonable quantities (about 0.2–1.0 μmol per preparation; Spiro et al., 1976; Badet and Jeanloz, 1988; Gibbs and Coward, 1999). As described in Materials and Methods, OST-catalyzed glycosylation of peptide 1 followed by HPLC purification of the crude reaction mixture using Con A-Sepharose chromatography and thermolysin treatment, gave a nearly homogeneous product of IVT-Asn 88–96 as judged by HPLC (data not presented). Sequence analysis of this peptide revealed that, as in the case of the invertase tryptic peptide (residues 77–105, Table II), only the first Asn of the overlapping sequons was glycosylated. Studies with heptadecapeptide 2, in which the first Asn was replaced by Gln (QNTS) demonstrated this peptide to be a poor substrate for OST (16% the activity of Bz-NLT-NH₂), similar to what was observed for intact invertase containing this mutation (Table II and III). However, peptide 3, in which Gln and Asn were
reversed (NQTS) was as good an acceptor as peptide 1, that is, 70% as effective as Bz-NLT-NH₂. Thus, OST-catalyzed glycosylation in vitro mirrors the in vivo reaction as demonstrated by site-directed mutagenesis results described earlier in this article (Table II, III). Similar, but less quantitative in vitro results were reported earlier (Liu et al., 1994).

Comparison of small molecular weight sequons as glycosylation acceptors

The extent of glycosylation at various sequons appears to be influenced not only by the location of the sequon within the secondary structure of the folding protein, but also by the amino acids adjacent to the sequon. Therefore, we undertook to determine how glycosylation is affected by the latter factor by employing peptides containing sequons with the four amino acids (NNTS) found normally in sequons 4 and 5 of invertase, as well as several variant sequences including those previously examined (NNTS) found normally in sequons 4 and 5 of invertase, as well as several variant sequences including those previously examined (Table II). The relative ability of these sequons to accept oligosaccharides in vitro is presented in Table IV. While we did not measure the distribution of carbohydrate associated with each Asn in the overlapping sequons, as was done for the thermolysin glycopeptide derived from peptide 1 (vide supra), there was good agreement with the in vivo results presented in Table II and III. As described above, similar results were obtained in the in vitro study using the heptadecapeptide encompassing residues 82−98 of invertase.

Table IV. Relative glycosylation of modified tri- and tetrapeptides encompassing sequons 4 and 5

<table>
<thead>
<tr>
<th>Peptide acceptors</th>
<th>Relative activity ± 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Variants of wild type invertase</strong></td>
<td></td>
</tr>
<tr>
<td>(1) Bz-Asn-Asn-Thr-Ser-NH₂</td>
<td>1.12 ± 0.18</td>
</tr>
<tr>
<td>(2) Bz-Gln-Asn-Thr-Ser-NH₂</td>
<td>0.24 ± 0.01</td>
</tr>
<tr>
<td>(3) Bz-Asn-Gln-Thr-Ser-NH₂</td>
<td>1.14 ± 0.10</td>
</tr>
<tr>
<td>(4) Bz-Glu-Asn-Thr-Ser-NH₂</td>
<td>0.06 ± 0.05</td>
</tr>
<tr>
<td>(5) Bz-Asn-Glu-Thr-Ser-NH₂</td>
<td>0.54 ± 0.02</td>
</tr>
<tr>
<td><strong>Overlapping sequons of invertase</strong></td>
<td></td>
</tr>
<tr>
<td>(6) Bz-Asn-Asn-Thr-Ser-NH₂</td>
<td>1.12 ± 0.18</td>
</tr>
<tr>
<td>(7) Bz-Asn-Asn-Thr-NH₂</td>
<td>0.87 ± 0.07</td>
</tr>
<tr>
<td>(8) Bz-Asn-Asn-Thr-Ser-NH₂</td>
<td>0.29 ± 0.12</td>
</tr>
<tr>
<td>(9) Bz-Asn-Asn-Thr-Thr-NH₂</td>
<td>1.06b</td>
</tr>
<tr>
<td><strong>Overlapping sequons of cholinesterase</strong></td>
<td></td>
</tr>
<tr>
<td>(10) Bz-Asn-Asn-Ser-Thr-NH₂</td>
<td>0.92 ± 0.05</td>
</tr>
<tr>
<td>(11) Bz-Asn-Asn-Ser-NH₂</td>
<td>0.32 ± 0.06</td>
</tr>
<tr>
<td>(12) Bz-Asn-Ser-Thr-NH₂</td>
<td>1.05 ± 0.10</td>
</tr>
<tr>
<td>(13) Bz-Asn-Asn-Ser-Ser-NH₂</td>
<td>0.60b</td>
</tr>
</tbody>
</table>

*Each peptide was dissolved in DMSO to give a stock solution of 7.2 mM; 5 µl of this stock solution was added to the assay mixture as described in Materials and methods to give a final concentration of 360 µM. The activity for each peptide is given relative to the tripeptide, Bz-Asn-Leu-Thr-NH₂, an extensively studied OST substrate, which was also present at 360 µM in the assay mixture. All data are the average of at least triplicate determinations except as noted.*

bSingle determination.

cLockridge et al. (1987).

Replacement of Asn₀₂ with Gln (Table IV, 2) or Glu (Table IV, 4) greatly reduced glycosylation of the remaining sequon, Asn-Thr-Ser. These findings are comparable to the in vivo results (Table III) and are consistent with the poor acceptor properties of sequons containing Ser in the third position. Thus it would appear on comparing the relative activity of Bz-Asn-Thr-Ser-NH₂ (Table IV, 8) with Bz-Asn-Thr-Thr-NH₂ (Table IV, 12) as oligosaccharide acceptors that the latter is much better than the former. As anticipated, all of the peptides containing Ser at the carboxy terminus (Table IV, 2, 4, 8, and 11) do not accept carbohydrate nearly as well as those containing Thr on the sequon carboxy terminus (Table IV, 7, 9, 10, and 12), although, as indicated above, there are some exceptions to this rule in intact invertase (Reddy et al., 1988). Even with the tetrapeptides, most of the carbohydrate appears to be directed to the first sequon of the overlapping invertase sequons in the wild-type sequence (Table IV, 1) as evidenced by the reduced capacity of peptides missing the first Asn or modified at that position (Table IV, 2, 4, 8) to accept carbohydrate. As expected the (Thr)₂-containing peptide (Table IV, 9) was a better acceptor than the comparable (Ser)₂-containing peptide (Table IV, 13).

Discussion

It is becoming increasingly clear from attempts to explain the variability in the extent and/or rate of sequon glycosylation that no one explanation is completely adequate with the possible exception that substituting a proline for the amino acid (X) following Asn in the sequon in Asn-X-Ser/Thr or placing a proline after Ser/Thr prevents glycosylation (Gavel and von Heijne, 1990). In some cases glycosylation can occur, although at a greatly reduced level, where Cys is substituted for Ser/Thr in the third position of a sequon (Bause and Legler, 1981; Mileitch and Broze, 1990; Vance et al., 1997). Other potential reasons for the incomplete glycosylation of sequons in vivo are presented below.

Variability in glycosylation due to structural considerations

A possible reason for the absence of glycosylation at sequon 5 in invertase could be its sequestration in a locally folded region of the growing peptide chain making it inaccessible to OST. Our present understanding of the translocation process across the endoplasmic reticulum in yeast does not preclude this possibility. Thus, the secretory protein, during or after translation, is suggested to be targeted by a cytosolic Hsc70 protein to the translocation apparatus on the endoplasmic reticulum membrane (Miletich and Broze, 1990; Brodsky and Schechman, 1993). The absence of a significant number of glycosylated sequons at the carboxy terminal region of glycoproteins led Gavel and von Heijne (1990) to propose that these sites are not accessible for glycosylation because the translated protein acquires a tertiary structure that blocks the glycosylation sites. That folding is a deterrent to glycosylation is seen in the case of tissue plasminogen activator where prevention of disulfide bond formation results in complete sequon glycosylation both in vitro (Nunnari and Walter, 1992) and in vivo (Bulleid et al., 1992). Alternatively, disulfide bond formation at a sequon in human plasma protein C, where Cys replaces Ser/Thr also limits glycosylation; this is not the case at this sequon when Cys is unoxidized (Allen et al., 1995). However, other reasons for the nonglycosylation of sequons must be considered since the first sequon of human Ig, located at Asn₁₇₄, is not glycosylated, while the remaining five sequons are (Shimizu et al., 1971). Similarly, hippopoptamus RNase is not
glycosylated at Asn21 of its first sequon site (Havinga and Beintema, 1980). Other examples are porcine erythrocyte glycophore, which is not glycosylated at Asn16 (Honma et al., 1980), and yeast trehalase with three sequences at Asn6, Asn210, and Asn291, none of which is glycosylated (Kopp et al., 1993). Results from several laboratories (Abbadi et al., 1986, 1991; Pichon-Pesme et al., 1988; Iishi et al., 1989; Imperiali and Shannon, 1991; Imperiali et al., 1992, 1994) have implicated a variety of protein structural motifs (e.g., -turn, Asx-turn) as important recognition elements in protein N-glycosylation. However, recent studies in which solution structures (NMR) of peptide ligands were compared with the activity of each peptide as an OST substrate or inactivator demonstrated that formation of the Asx-turn may be necessary but not sufficient for the formation of a productive enzyme-ligand complex (Xu et al., 1998). In addition, the variable in vitro and in vivo glycosylation of Asn292 and Asn293 as described in the present study, together with earlier work of Bause and Lehle (1979) and Ronin et al. (1981) on peptides derived from ribonuclease or ovalbumin, suggest that primary sequence is also an important determinant of substrate recognition by OST. Although N-linked glycosylation is a cotranslational process (Hubbard and Ivatt, 1981), it is unclear whether significant secondary or tertiary structure develops in the growing peptide, particularly in the region of sequons and overlapping sequons, as the peptide passes through the membrane into the lumen of the endoplasmic reticulum prior to OST-catalyzed glycosylation (Nilsson and von Heijne, 1993). However, as indicated previously (Imperiali and Rickert, 1995), glycosylation can trigger the formation of structural nucleation elements leading to localized conformational changes. Whether such changes can occur rapidly enough to affect the glycosylation of overlapping sequons is not known, although some evidence has been presented indicating that variable rates of translation and translocation could explain the differences in carbohydrate content in various sequons (Rothman and Lodish, 1977; Glabe et al., 1980; Reddy et al., 1988; Shelikoff et al., 1996).

The role of serine and threonine in influencing glycosylation

The data reported in this article show that Gln92-Asn323-Thr324-Ser325 is a poor carbohydrate acceptor both in vitro (Table I) and in vivo (Table IV). It is unlikely that Gln92 is responsible for the impairment in glycosylation of Asn323-Thr-Ser since a peptide that contains the sequence Asn-Gln-Thr-Ser is a good substrate for OST in vitro (Table IV). As shown in Table III, inversion of Ser and Thr in sequons 4 and 5 does not interfere with the glycosylation of mutated sequon 5 (NST), although mutated sequon 4 (NTS) is poorly glycosylated. The most likely explanation for the results reported in this article regarding the extent of glycosylation of the overlapping sequons 4 and 5 of yeast invertase (NNTS), which is supported by an abundance of evidence, is that Ser in the third position of a sequon negatively modulates its glycosylation while Thr in the third position promotes glycosylation. This statement is consistent with the findings in Table I–III, where Asn-Asn-Thr-Ser is glycosylated predominantly on the first Asn, whereas Asn-Asn-Ser-Thr is glycosylated mainly on the second Asn, as was shown earlier for cholinesterase (Lockridge et al., 1987). However, the presence of Thr in the third position of a sequon does not guarantee that it will be fully glycosylated as seen in the cases of sequons 7 and 14 of invertase (Reddy et al., 1988). To complicate matters, the presence of adjacent serines in the overlapping sequon NNSS (Table II, III) enables both sequons to be glycosylated, which does not appear to be the case when one of the serines is replaced by a threonine.

Thus, in addition to the location of Ser or Thr in a sequon, the nature of the amino acids in the X position and adjacent to the sequon, as well as its accessibility to OST in the folded structure of a protein, all appear to affect the extent of glycosylation of a growing protein.

Materials and Methods

Materials

Chromatography media DE-52 and Sephacryl S-500 were from Whatman Co. and LKB-Pharmacia, respectively. The Aquapore C8 reverse phase column was from Rainin Instruments Co. The Aquapore C8 reverse phase column was from Rainin Instruments Co. The following kits were used: Geneclean from Bio 101, Magic Prep from Promega, and an oligonucleotide-directed mutagenesis kit, version 2.1 from Amersham. M13 phage were prepared in Escherichia coli (strain TGI). E. coli strain HB 101 was used to select and propagate the shuttle vector pEHi3. Invertase was expressed in yeast strain Y275 (MATa ura3–52 leu2–3,112 ade2–101 suc2 [Δleu2]). Endo H (EC 3.2.1.96) was prepared in this laboratory as a cloned product (Trumbly et al., 1985). PNGase F (EC 3.5.1.52) was a generous gift of Dr. Anthony Tarentino (Plummer et al., 1984). ConA Sepharose was purchased from Sigma Chemical Co.

Yeast OST was isolated either as a microsomal (Clark et al., 1990), a salt-stripped microsomal (Kelleher et al., 1992), or as a solubilized protein complex (Chalifour and Spiro, 1988). Biosynthetic Glc[1-3H]Man9(GlcNAc)2-PP-Dol (LOS) (Trimble et al., 1980; Clark et al., 1990) and ([3H]GlcNAc)2-PP-Dol (LDS) (Sharma et al., 1981; Lee and Coward, 1992) were prepared as described previously. Unlabeled LOS from bovine pancreas was isolated and partially purified basically as described by Gibbs and Coward (1999) and Spiro et al. (1976). Peptides were synthesized by solid phase procedures using an Applied Biosystems 431A peptide synthesizer and purified by reverse phase HPLC at either the Peptide Synthesis Core Facility of the Wadsworth Center or the Protein and Carbohydrate Structure Laboratory at the University of Michigan.

2-Acetamido-1-([β-L-aspartyl]-2-deoxy-β-D-glucopyranosyl)-amine was purchased from Sigma and its PTH derivative was prepared as described (Reddy and Maley, 1993) for use together with other PTH amino acid standards in HPLC analysis. Amino acid sequencing of the biosynthetic glycopeptides was performed at the University of Michigan Protein and Carbohydrate Structure Laboratory using an ABI 470 amino acid sequencer.

Methods

DNA manipulations. The shuttle vector pEHi3 (Reddy and Maley, 1996), which contains the SUC2 gene derived from pRB58 (Carlson and Botstein, 1982), was used for the site-directed mutagenesis studies. The coding sequence of external invertase was excised as a 2 kb HindIII fragment and subcloned into M13mp19. The reverse oligonucleotides, CCCACCTGG-TGTTCTGTTGAATC, AAACCCACTGGATTTTGTTAAATC, and AAACCCAGTCGATTGTTGTAATC, were used to prepare the invertase single mutants N92Q, and T94S, and the double mutant [T94S,S95T]. The indicated base changes are underlined. The protocols for site-directed mutagenesis and hybridization screening are as described in the Amersham version 2.1 kit. Mutagenesis was confirmed by DNA sequencing. Following mutagenesis, the SUC2 coding sequence was excised from M13mp19 with HindIII and subcloned back into the HindIII site.
of pEI13 (Reddy and Maley, 1996). Colonies containing SUC2 in the correct orientation in pEI13 were selected by hybridization using a HindIII site-overlapping oligonucleotide.

**Expression and purification of Asn2Gln and other mutant invertases.** Transformed yeast colonies were lifted onto Whatman paper circles and screened for invertase activity (Trumbly, 1986). An invertase-expressing colony was purified by replating on leucine selective solid medium. For the purpose of invertase purification, cells were grown in a 2 l Bioflo III fermentor (New Brunswick Scientific), which contained 1.9 l YPD medium supplemented with 10% glucose and 0.02% adenine sulfate to 0.1 mM CaCl₂ and digested with 3% (w/w) TPCK-trypsin at 37°C. To induce external invertase, the cell paste was suspended in a culture was withdrawn for harvesting. A second batch was grown by adding another 1.8 l of fresh medium to the fermentor flask as described above. The yield of the combined batches was 40 g/l. To induce external invertase, the cell paste was suspended in a half-strength YPD medium containing 0.1% sucrose and shaken at 30°C for 10 h.

The cells (130 g) were collected by centrifugation and broken in a solution of 200 ml of 0.05 M Tris–HCl pH 7.4, 5 mM EDTA, and 0.5 mM phenylmethylsulfonyl fluoride (buffer 1), by agitating three times in a bead beater at 0–4°C with 2 volumes of 0.5 mm glass beads, each for one min, with 1 min of cooling between cycles. The glass beads were washed with buffer 1 before use. The slurry was centrifuged at 1000 r.p.m. in a Sorvall SS34 rotor. Streptomycin sulfate (2 g in 20 ml of buffer 1) was added with stirring to the crude supernatant fraction. Following centrifugation, the supernatant fraction was heated at 30°C for 20 min and the precipitate was removed by centrifugation. The enzyme solution was concentrated on a PM 30 ultrafiltration membrane (Millipore) to less than 50 ml, dialyzed in 0.01 M Na phosphate pH 7.1 (buffer 2), and passed through a DE-52 column (2.5 x 25 cm). Invertase was eluted with a linear gradient from 0 to 0.2 M NaCl in buffer 2 and concentrated on a PM30 membrane to 7 ml. Aliquots of concentrated enzyme (1 ml) were placed on a Sephacryl S-500 column (1.6 x 60 cm) connected to an HPLC equipped with a UV detector. The column was developed with 0.02 M Na acetate pH 5.0, containing 0.8 M NaCl at a flow rate of 0.6 ml/min. The enzyme eluted early from the column, since at this concentration of NaCl the enzyme is mostly an octamer (Reddy et al., 1990). The purified enzyme (6.5 mg) had a specific activity of about 4000 IU/mg of protein.

**Purification of glycopeptides from external invertase.** Wild type invertase (6.5 mg) or its mutants were S-carboxymethylated according to the procedure of Crestfield et al. (1963), diaлизized in 2% acetic acid, and lyophilized. Each carboxymethylated protein (2 mg) was dissolved in 1 ml of a solution of 0.1 M NH₄HCO₃ and 0.1 mM CaCl₂ and digested with 5% (w/w) TPCK-trypsin at 37°C for 18 h. The digest was acidified with CH₃COOH and injected onto an Aquapore C8 reverse phase column. The peptides were eluted with a linear gradient of CH₃CN as described in the Figure 1 caption. The carbohydrate-containing peaks were identified and isolated for amino acid sequencing. The tryptic glycopeptides were treated with chymotrypsin under the same conditions used for trypsin digestion, to obtain shorter peptides. To remove the oligosaccharides the glycopeptides were incubated with PNGase F in 10 mM Tris–HCl pH 8.0, 0.02% sodium azide at 30°C for 16 h. When necessary the glycopeptides were purified further on a C-8 reverse phase column using a shallow CH₃CN gradient.

**Large scale glycosylation of peptide 1 with bovine LOS.** For each of five reactions (four with enzyme, one control without LOS), 0.5 mg of a salt-stripped microsomal P40–OST protein complex was preincubated for 15 min with 1 mM TPCK and a protease inhibitor cocktail (100 ng/ml peptatin, chymostatin, leupeptin, antipain, and 1 µg/ml aprotinin) in a buffer containing 50 mM Tris–HCl pH 7.5, 1% Triton X-100, 4 mM MgCl₂, 1 mM MgCl₂, and 2 mM dithiothreitol. The enzyme mixture was next transferred to a culture tube containing 36 nmol of heptadecapeptide 1 in 5 µl of dimethylsulfoxide, and 3.6 nmol of dried bovine LOS containing 35,000 d.p.m. of [³H]LOS. The final assay volume was 100 µl. The assays were incubated at room temperature for 3 h, after which they were quenched with 3 ml of CHCl₃: MeOH (2:1). The glycosylation product (34,000 d.p.m. ~3.6 nmol) was isolated in the H₂O fraction (Clark et al., 1990). The H₂O fractions were concentrated in vacuo in a Speed Vac (Savant) centrifugal evaporator. The residue was dissolved in H₂O, and purified on a C-18 reversed phase HPLC column (Vydac) with a gradient mobile phase of H₂O-CH₃CN (0.1% CF₃COOH), 0–60% CH₃CN over 60 min. H-containing fractions, corresponding to the glycopeptide product in about 30% CH₃CN, were pooled and concentrated.

**Purification of the glycopeptide using Con A affinity column.** Lectin affinity chromatography was conducted as described previously (Clark et al., 1990). A small Con A column (0.75 ml) was equilibrated in a solution containing 25 mM Tris-Cl, pH 7.5, 500 mM NaCl, 0.5 mM MgCl₂, 0.5 mM MnCl₂, 0.5 mM CaCl₂, and 1% NP-40 (Con A buffer). The HPLC purified glycopeptide was dissolved in 0.2 ml of Con A buffer and slowly loaded onto the column at 4°C. The sample was allowed to incubate with the resin for 1 h. The column was then washed with 10 x 0.75 ml of Con A buffer at 4°C. The desired glycopeptide was eluted subsequently with 8% α-methylmannoside in Con A buffer at room temperature and each fraction was analyzed for radioactivity. α-Methylmannoside in each product-containing fraction was removed by preparative HPLC, the glycopeptide was recollected, and the pooled fractions were evaporated to dryness.

**Thermolysin and Endo H treatment of glycopeptide.** The dried glycopeptide (2 nmol) was incubated with 4 µg of thermolysin in 50 µl of 0.1 M NH₄HCO₃, pH 8.0, at 50°C for 1 h, followed by incubation at 37°C overnight (Ziegler et al., 1988). Completion of the digestion was checked by HPLC using a small aliquot of the reaction mixture. The digested glycopeptide mixture was purified by HPLC using the solvent system described above, and eluted at ~20% CH₃CN. The two products of the thermolysin-catalyzed hydrolysis (~1 nmol each) were incubated with 20 µU Endo H in 0.2 ml of 250 mM Na citrate buffer, pH 5.5, overnight (Tarentino et al., 1978; Clark et al., 1990). The progress of the hydrolysis was monitored by HPLC. When each reaction was complete, the GlcNAc-containing glycopeptides were isolated by preparative HPLC. Two additional HPLC purifications of each glycopeptide was required to obtain sufficient pure material (~500 pmol) for sequencing.

**Acknowledgments**

We gratefully acknowledge the contribution of the Wadsworth Center’s Protein Chemistry and Molecular Genetics Core Facilities to these studies. The Protein and Carbohydrate Core Facility at the University of Michigan is supported by grants from the National Institutes of Health; CA46592 and AR20557. In addition our appreciation goes to Judith Valentinio for excellent assistance.
in preparing this manuscript. We thank Andrew Haidle for his assistance in the acquisition of some of the peptide glycosylation data in Table IV. This work was supported by Grant CA44355 (F.M.) from the National Cancer Institute, USPHS/NIH, and by the Vahlteich Research Fund administered by the College of Pharmacy, University of Michigan (J.K.C.). Y.-L.L. was supported by a fellowship from the Program in Protein Structure and Design, University of Michigan.

Abbreviations

OST, oligosaccharyl transferase; (1H)LOS, lipid oligosaccharide labeled in its mannose residues; (1H)LDLS, lipid disaccharide (chitoitiate) labeled in its N-acetyl glucosamine residues; Con A, concanavalin A; WGA, wheat germ agglutinin; Endo H, endo-β-N-acetylglucosaminidase H; PNGase F, peptide-N-acetylglucosaminidase F; IVT-Asn82–98-DS, chitoitinate-containing invertase peptide including amino acid residues 82–98 (encapsoming sequons 3–5); IVT-Asn82–98-OS, same as IVT-Asn82–98-DS peptide but containing a mannooligosaccharide; IVT-Asn88–98-DS, invertase peptide including amino acids 88–96 (sequons 4 and 5); IVT-Asn88–98-OS, same as IVT-Asn88–98-DS peptide but with a mannosyl oligosaccharide; Bz, benzoyl; PTH, phenylthiohydantoin; TPCK, N-tosyl-L-phenylaline chloromethyl ketone.

References


Glycosylation of overlapping sequons


