Quantitative assessment of the preferences for the amino acid residues flanking archaenal N-linked glycosylation sites

Mayumi Igura and Daisuke Kohda
Division of Structural Biology, Medical Institute of Bioregulation, Kyushu University, Maidashi 3-1-1, Higashi-ku, Fukuoka 812-8582, Japan

Received on July 27, 2010; revised on October 9, 2010; accepted on November 18, 2010

Oligosaccharyltransferase (OST) catalyzes the transfer of an oligosaccharide to an asparagine residue in polypeptide chains. Using positional scanning peptide libraries, we assessed the effects of amino acid variations on the in vitro glycosylation efficiency within and adjacent to an N-glycosylation consensus, Asn-X-Ser/Thr, with an archaeal OST from Pyrococcus furiosus. The amino acid variations at the X_{-2}, X_{-1} and X_{+1} positions in the sequence X_{-2}X_{-1}Asn-X-Ser/Thr-X_{+1} strongly influenced the glycosylation efficiency to a similar extent at position X. The rank orders of the amino acid preferences were unique at each site. We experimentally confirmed that the archaeal OST does not require an acidic residue at the −2 position, unlike the eubacterial OSTs. Pro was disfavored at the −1 and +1 positions, although the exclusion was not as strict as that at X, whereas Pro was the most favored amino acid residue among those studied at the −2 position. The overall amino acid preferences are correlated with a conformational propensity to extend around the sequon. The results of the library experiments revealed that the optimal acceptor sequence was PYNVTK, with a K_m of 10 µM. The heat-stable, single-subunit OST of P. furiosus is a potential candidate enzyme for the production of recombinant glycoproteins in bacterial cells. Quantitative assessment of the amino acid preferences of the OST enzyme will facilitate the proper design of a production system.

Keywords: archaea / enzyme kinetics / N-linked protein glycosylation / oligosaccharyltransferase / peptide library

Introduction

N-linked glycosylation refers to the attachment of an oligosaccharide chain onto an asparagine residue and ubiquitously occurs on secreted and membrane-bound proteins in eukaryotes and archaea, and in some bacteria (Burda and Aebl 1999; Knauer and Lehle 1999; Yan and Lennarz 2005; Abu-Qarn et al. 2008). Oligosaccharyltransferase (OST) catalyzes covalent bond formation between an oligosaccharide chain and an asparagine residue in polypeptide chains. OST is a multi-subunit membrane protein complex in higher eukaryotes (Knauer and Lehle 1999), but a single-subunit membrane protein in lower eukaryotes, archaea and eubacteria (Feldman et al. 2005; Kelleher and Gilmore 2006; Igura et al. 2008). The catalytic subunit of the OST enzymes has a common ancestor, but is referred to differently among the three domains of life, as staurosorpin and temperature sensitive 3 (STT3) in eukaryotes, archaeal glycosylation B (AgIB) in archaea and protein glycosylation B (PglB) in eubacteria. The primary sequences of the STT3/AgIB/PglB proteins consist of 600–1000 residues and share a common architecture (Kim et al. 2005; Li et al. 2010): There is a multispans transmembrane region in the N-terminal half of the primary sequence, and the C-terminal half forms a globular domain.

The sequence alignment of the STT3/AgIB/PglB protein family members revealed two conserved motifs. A five-consecutive-residue motif, WWDYG (Trp-Trp-Asp-Tyr-Gly), resides in the C-terminal globular domain, whereas a diacidic motif, DXD or EXD (X denotes any amino acid residue), is found in the first luminal/extracellular loop of the N-terminal transmembrane region. The WWDYG motif is a characteristic fingerprint of the STT3/AgIB/PglB proteins. The central aspartate in the motif is considered to function as a catalytic base in the oligosaccharide transfer reaction. Mutagenesis studies demonstrated the essential roles of these two motifs in the catalytic functions of the STT3s from yeast and Leishmania major, and PglB from Campylobacter jejuni (Wacker et al. 2002; Yan and Lennarz 2002; Glover et al. 2005; Igura et al. 2008; Hese et al. 2009; Maita et al. 2010). In addition to these two motifs, the structure-based sequence alignment revealed a third conserved motif (Igura et al. 2008). The third motif was classified into three groups, according to the different consensus patterns, DXXXXX(M/I), DXXMXXX(K/I) and MXIIIXXX(U/V/W) (Maita et al. 2010). DK, DM and MI are the abbreviated forms of the three long consensus patterns. The DK and MI motifs occupy the identical position close to the WWYDG motif in the three-dimensional structures (Igura et al. 2008; Maita et al. 2010). We proposed the classification of the catalytic centers of the STT3/AgIB/PglB proteins into three types, according to the DK/DM/MI motifs. Eukaryotic STT3 proteins exclusively...
contain the DK motif (designated as E-type), whereas eubacterial PglB proteins contain only the MI motif (B-type). In contrast, archaeal AglB proteins contain either the DK, DM or MI motif (E-type, A-type or B-type). The *Pyrococcus furiosus* AglB, for which the structure was determined, contains the DK motif, and thus the *Pyrococcus* structure is suitable as a eukaryotic model. In vivo mutagenesis studies of the STT3 proteins from yeast and *L. major* showed the requirement of the DK motif for survival (Igura et al. 2008; Hese et al. 2009). An in vitro mutagenesis study of the *C. jejuni* PglB protein also confirmed the involvement of the MI motif in the activity (Maita et al. 2010).

N-linked glycosylation occurs specifically at the Asn-X-Ser/Thr sequon, where X is any amino acid except for Pro (Gavel and von Heijne 1990). The exclusion of a proline residue at the position C-terminal to Ser/Thr is also evident. The eubacterial *C. jejuni* sequon was extended to Asp/Glu-X-1-Asn-X-Ser/Thr, where both X-1 and X are any amino acid except for Pro (Kowarik, Young, et al. 2006). The requirement of the third hydroxyamino acid residue in the N-glycosylation consensus suggests the formation of a turn-like conformation, in order to bring the hydroxy group of Ser/Thr closer to the carboxamide group of Asn during the reaction to increase the nucleophilicity of the Asn (Bause and Legler 1981; Imperiali et al. 1992), in a substrate-assisted catalytic mechanism. The Cβ stereoisomer of Thr, 1-allo-threonine, cannot replace the function of the Thr residue (Bause and Legler 1995), and Gln cannot serve as a surrogate for the Asn residue (Igura et al. 2008), indicating the requirement for the strict geometry of the turn-like reactive conformation during the reaction. The exclusion of Pro at position X also supports this notion. Recently, the eubacterial *C. lari* sequon was expanded to an atypical sequence, Asp-Ala-Asn-Thr, albeit with low efficiency (Schwarz, Lizak, et al. 2010). Further investigations are necessary to determine the requirement of the turn-like reactive conformation in the catalysis by the OST enzymes.

Comprehensive statistical analyses of the eukaryotic and archaeal glycosylation sites revealed additional, but unremarkable, amino acid preferences at position X or adjacent positions to the sequon (Ben-Dor et al. 2004; Petrescu et al. 2004; Abu-Qarn and Eichler 2007). For example, small deviations from the expected amino acid composition were detected around the occupied sequons. There is an increased occurrence of aromatic residues at the X-1 and X-2 positions in eukaryotic glycoproteins, but there is an opposite trend in archaeal glycoproteins. In contrast, experimental analyses of an individual N-glycosylation sequon of a particular protein/peptide showed the pronounced effects of amino acid variations on the glycosylation efficiency with the eukaryotic canine OST (Shakin-Eshleman et al. 1996; Kasturi et al. 1997; Mellquist et al. 1998), and the eubacterial OST from *C. jejuni* (Chen et al. 2007). Detailed knowledge about the effects of amino acid residues on the glycosylation efficiency will provide clues toward understanding the mechanism that determines whether a sequon is occupied or not. For future applications, the quantitative assessment of the amino acid preferences of OST is crucial for the proper design of a recombinant glycoprotein production system in bacterial cells, similar to the system incorporating the *C. jejuni* OST (Schwarz, Huang, et al. 2010).

## Results and discussion

### Effects of amino acid variations on N-glycosylation

The NXS peptide library consists of eight-residue peptide sequences, sharing a common sequence, Cys-Arg-Gly-Ala-Ala-Arg (Table I). Each peptide differs at the second position of the sequon (underlined), in order for the systematic evaluation of the amino acid preference of the archaeal OST from *P. furiosus*. The effects of the 19 amino acids (other than cysteine) at position X on the N-glycosylation efficiency were investigated. A fluorescent dye was appended to the N-terminal cysteine residue through thiol-maleimide chemistry, for detection and quantification. In parallel, the NXT library, which contains the Asn-X-Thr sequon, was synthesized and evaluated. The results of the OST assays, using a Triton X-100-solubilized membrane fraction from *P. furiosus* cells, are shown in Figure 1A. Since the Triton-solubilized membrane fraction contains the OST enzyme and an oligosaccharide donor, lipid-linked oligosaccharide (LLO), the OST reaction progresses readily in the presence of peptides containing the N-glycosylation sequon.

Regarding the data in Figure 1, we noticed the somewhat nonlinear behavior of the reaction curves for sequons with high acceptor activities. We decided, however, to use these assay conditions to ensure the wide dynamic range (10–1000 fmol) of the assay. Duplicate or triplicate assays for the same peptides provided an evaluated error of ±5–10% (e.g. the X = Ser assay was duplicated in Figure 1A). Thus, small differences in the amino acid preferences should not be overrated, particularly for sequons with low acceptor activities. Nevertheless, the good linear correlation ($R^2 = 0.97$) between the results of the NXS and NXT library experiments ensures the quantitative nature of the peptide library experiments (Figure 1B).

#### Table I. List of peptide libraries

<table>
<thead>
<tr>
<th>Peptide library</th>
<th>Varying Sequence$^b$</th>
<th>X</th>
<th>Position$^a$</th>
</tr>
</thead>
<tbody>
<tr>
<td>XGNSX -2 correlated of TAMRA-XGNSXR-COOH</td>
<td>A, R, D, E, Y, P</td>
<td>C</td>
<td></td>
</tr>
<tr>
<td>XNNSS -1 correlated of TAMRA-RXNNSXR-COOH</td>
<td>A, D, E, G, Y, P</td>
<td>C</td>
<td></td>
</tr>
<tr>
<td>NXS 0 correlated of TAMRA-RGNXSSXR-COOH</td>
<td>19 amino acids (not C)</td>
<td>C</td>
<td></td>
</tr>
<tr>
<td>NXT 0 correlated of TAMRA-RGNXTAR-COOH</td>
<td>19 amino acids (not C)</td>
<td>C</td>
<td></td>
</tr>
<tr>
<td>NSSX +1 TAMRA-RGNSSXR-COOH</td>
<td>A, F, G, L, Y, W, K</td>
<td>C</td>
<td></td>
</tr>
</tbody>
</table>

$^a$The positions of the library are defined as -X-1-X-2-X-3-X-4-X-5-X-6-X-7.

$^b$C-TAMRA denotes the coupling of TAMRA-maleimide to the thiol group of the N-terminal Cys after peptide synthesis. The N-glycosylation sequon is underlined.
The preference for Thr over Ser at the third position of the sequon was repeatedly observed in the statistical analyses of experimentally verified eukaryotic N-glycosylation sites (Ben-Dor et al. 2004; Petrescu et al. 2004). Similar statistical analyses of archaeal and eubacterial (i.e. C. jejuni) glycoproteins revealed no particular bias for Thr or Ser at the third position (Kowarik, Young, et al. 2006; Abu-Qarn and Eichler 2007). However, the preference for Thr was experimentally confirmed for the archaeal OST in the peptide library experiments (Figure 1B). The slope of the regression line indicates that the sequons with Thr at the third position provide about 6-fold better acceptor activity than those with Ser at the same position, for the \textit{P. furiosus} OST. The discrepancy with the statistical analysis may be due to the small number of glycosylation sites used in the statistical analysis (Abu-Qarn and Eichler 2007). Next, a clear preference for specific amino acids was observed at position X, the second position of the sequon (Figure 2). Hydrophobic residues, such as Val, Trp and Ile, are favored, and Val generates the most efficient peptide substrate, but other hydrophobic residues, such as Leu, Tyr and Phe, are only moderately tolerated substrates. Interestingly, peptides containing Gln or Glu were relatively good substrates, but those containing Asn or Asp were poor substrates. As expected, peptides containing Pro at position X were not substrates.

**Fig. 1.** Effects of the amino acid substitutions at the second position of the N-glycosylation sequon in the peptide libraries, NXS and NXT, on the efficiency of the oligosaccharide transfer reaction catalyzed by the archaeal OST from \textit{P. furiosus}. (A) Glycopeptide products were separated from unmodified peptides by SDS–PAGE and were quantified by fluorescence imaging. The N-glycosylation consensus is underlined. (B) Good correlation exists between the results of the two library experiments.
Amino acid preferences at flanking positions

The preferences of the \textit{P. furiosus} OST for the amino acid residues at other positions were evaluated using other libraries (Table I), as summarized in Figure 2. Here, we defined the positions of the sequon as \(-X_{-2}-X_{-1}-\text{Asn}-\text{Ser/Thr}-X_{+1}\). The amino acid variations at the \(X_{-2}, X_{-1}\) and \(X_{+1}\) positions strongly influenced the glycosylation efficiency to a similar extent at position \(X\) (Figure 2). Pro was disfavored at the \(−1\) and \(+1\) positions, although the exclusion was not as strict as that at \(X\). Interestingly, Pro was the most favored amino acid residue among the six amino acids studied at the \(−2\) position. The rank orders of the amino acid preference differ at each site. For example, the presence of an Asp residue at the \(−1\) position generated a better substrate than Glu at that position, which is in contrast to the opposite order at position \(X\). The eubacterial OSTs from \textit{C. jejuni} and \textit{C. lari} require an acidic amino acid residue at the \(−2\) position for efficient glycosylation (Kowarik, Young, et al. 2006; Schwarz, Lizak, et al. 2010), but the present study showed that the \textit{P. furiosus} OST did not require any acidic residues at the corresponding position. In fact, charged residues, including acidic residues, led to the lowest activity among the six amino acids examined. In accordance with this result, acidic residues are rarely found at the \(−2\) position of the occupied N-glycosylation sequons in eukaryotic and archael glycoproteins (Petrescu et al. 2004; Abu-Qarn and Eichler 2007). Finally, the results of the library experiments revealed that the optimal acceptor sequence is \(\text{PYNVTK}\).

Kinetic parameters of peptide substrates

We determined the kinetic parameters of several peptide substrates. We used the immunoaffinity-purified AglB protein from \textit{P. furiosus} cells (Igura et al. 2008). Crude LLOs were

Fig. 2. Summary of amino acid preferences of the \textit{P. furiosus} OST within and adjacent to the N-glycosylation sequon. Open bars indicate the acceptor activity of the original sequence, RGNSSAR or RGNSTAR. The N-glycosylation consensus is underlined.
separately prepared from *P. furiosus* cells and added to the reaction mixture. Good fitting to the theoretical Michaelis–
Menten curve was obtained (Supplementary data, Figure S1). The peptides containing Thr at the third position of the
sequon exhibited a 3-fold higher $V_{\text{max}}/K_m$ in comparison to the corresponding peptides containing Ser at the same
position (Table II). The $K_m$ of the peptides containing Ala, Ile or Val at position X was in the range 10–50 µM. These $K_m$
values are an order of magnitude higher than those (1–20 µM) reported for the *C. jejuni* PgIB (Chen et al. 2007). This
is ascribable to the enhancement of the affinity for the PgIB protein by the additional acidic residue in the eubacterial
sequon. The rank order of $V_{\text{max}}/K_m$ is Val > Ile > Ala and is identical to the preference in the library experiments (Figure 2). The peptide concentration used in the library experiments was 1.3 µM, which was much lower than the peptide $K_m$s. Consequently, the glycopeptide formation is pro-
tional to the $V_{\text{max}}/K_m$ value of the peptide substrates in our library experiments. As expected, the $V_{\text{max}}/K_m$ of the
substrate containing the optimal sequence, PYNVT, was the highest among the peptide substrates studied. We then embedded
the optimal sequence at the center of a longer amino acid sequence (15 residues). The $K_m$ and $V_{\text{max}}/K_m$ values were
almost identical to those of the original eight-residue sub-
strate. This implies that the *P. furiosus* OST can directly recognize the glycosylation sequon embedded in a long,
unstructured polypeptide chain.

Comparison of amino acid preferences with other OSTs
An in vitro enzymatic study with the eubacterial OST from
*C. jejuni* was reported (Chen et al. 2007). The results con-
firm that the sequon containing Thr is a 3-fold better sub-
strate than that containing Ser, and that Asp is required at the
X−1 position. In the library experiments, the effects of amino
acid variations at positions X−1 and X were studied, using the
DX$_{\text{X−1}}$NVS and DFNXS libraries. Strong amino acid prefer-
ences were observed for the two positions investigated. Other
library experiments using canine pancreas microsomes were
reported (Shakin-Eshleman et al. 1996; Kasturi et al. 1997;
Mellquist et al. 1998). The effects of amino acid substitutions
at positions X and X−1 were studied by the glycosylation
analysis of an engineered glycoprotein possessing a single
sequon, NLSE, in a cell-free translation system. Clear amino
acid preferences were also observed for the two positions
investigated.

We compared the amino acid preferences between the
archaeal OST and the eubacterial OST at position X
(Supplementary data, Figure S2). Contrary to our expec-
tions, no meaningful correlation between the two OSTs was
found ($R^2 = 0.11$). Chen et al. (2007) reported that the *C.
jejuni* OST exhibited a trend similar to that of the eukaryotic
canine OST at position X, but the correlation between them is
not high ($R^2 = 0.33$). We also found no correlation between the
*P. furiosus* OST and the canine OST ($R^2 = 0.005$) at the
same position. In summary, the amino acid preferences at pos-
iton X considerably differ from one another, among the
OSTs from the three domains of life (Supplementary data,
Figure S2).

The catalytic centers of the OST enzymes were classified
into the three types, according the DK/DM/MI motifs (Maita
et al. 2010), which raised the interesting possibility that the
amino acid preferences are dependent on the type of catalytic
center. However, this hypothesis does not fit well with the
present result: Although the canine OST and the *P. furiosus*
OST are members of the same class, they have very different
amino acid preferences.

### Relationship between amino acid preferences and amino
acid indices
What factor determines the amino acid preferences within and
adjacent to the N-glycosylation sequon? None of the simple
properties, such as charge or size, appears to explain the pre-
ferences. Various physicochemical properties, such as the ten-
dency for secondary structure formation, can be assigned
numerical values corresponding to the 20 amino acids.
Hereafter, the amino acid index refers to the set of 20 such
numerical values. The correlation coefficient measures the
strength and the direction of a linear relationship between two
sets of numerical values. We calculated the correlation coeffi-
cients between the amino acid preference (also regarded as a
set of numerical values, i.e. the absolute amounts of glycopep-
tides formed, corresponding to the 20 amino acid residues) at
each position, and the amino acid indices (more than 600)
collected in two databases (Golovin and Henrick 2008;
Kawashima et al. 2008). Figure 3 summarizes the results as a
dot plot, to facilitate an intuitive understanding of the global
distribution of the correlation coefficients. The distribution of
all of the correlation coefficients was almost random and
uniform. To gain insights into hidden relationships, we classi-
fied the amino acid indices into seven categories: (1) amino

### Table II. Kinetic parameters of peptide substrates with *P. furiosus* OST

<table>
<thead>
<tr>
<th>Sequence$^a$</th>
<th>$K_m$ (µM)</th>
<th>$V_{\text{max}}$ (fmol/h)</th>
<th>$V_{\text{max}}/K_m$</th>
<th>Remark</th>
</tr>
</thead>
<tbody>
<tr>
<td>NH$_2$-(C-TAMRA)RGNVTHAR-COOH</td>
<td>8.8 ± 0.5</td>
<td>690 ± 17</td>
<td>78</td>
<td></td>
</tr>
<tr>
<td>NH$_2$-(C-TAMRA)RGNNAR-COOH</td>
<td>10.4 ± 1.3</td>
<td>382 ± 21</td>
<td>37</td>
<td></td>
</tr>
<tr>
<td>NH$_2$-(C-TAMRA)RGNSAR-COOH</td>
<td>44 ± 7</td>
<td>1,140 ± 100</td>
<td>26</td>
<td></td>
</tr>
<tr>
<td>NH$_2$-(C-TAMRA)RGNNSAR-COOH</td>
<td>14.8 ± 3.0</td>
<td>348 ± 35</td>
<td>24</td>
<td></td>
</tr>
<tr>
<td>NH$_2$-(C-TAMRA)RGNNSAR-COOH</td>
<td>13.9 ± 3.5</td>
<td>176 ± 22</td>
<td>13</td>
<td></td>
</tr>
<tr>
<td>NH$_2$-(C-TAMRA)RGNNNSAR-COOH</td>
<td>65 ± 9</td>
<td>407 ± 36</td>
<td>6.3</td>
<td></td>
</tr>
<tr>
<td>TAMRA-APYNVTKR-COOH</td>
<td>10.5 ± 2.1</td>
<td>1,940 ± 170</td>
<td>185</td>
<td>Optimized substrate</td>
</tr>
<tr>
<td>TAMRA-GAGGSYVNTKGAGGS-COH$_3$</td>
<td>10.3 ± 3.9</td>
<td>1,730 ± 290</td>
<td>168</td>
<td>Long substrate</td>
</tr>
</tbody>
</table>

$^a$C-TAMRA denotes the coupling of TAMRA-maleimide to the thiol group of the N-terminal Cys. TAMRA- denotes the coupling of TAMRA-N-hydroxy
succinimide ester to the α-amino group of the N-terminal residue. The N-glycosylation sequon is underlined.

---

579
acid composition, (2) hydrophobicity, (3) extended/strand/sheet, (4) helix, (5) the middle positions of turns/bends, (6) the boundary positions of turns/bends and (7) the others. The correlation coefficients were classified according to the categories (Figure 3). We found significant correlations between the amino acid preference at each position and one category related to a secondary structure, as described in the next section.

**Amino acid preference at position X may be correlated to an extended conformation**

For the amino acid preference at position X, there is a cluster of dots, where each dot represents a correlation coefficient with one of the amino acid indices related to extended conformations (Figure 3). The positive association with the majority of the indices related to extended conformations implies that the dominant conformation of the peptide substrate during the oligosaccharide transfer reaction is an extended structure around position X. In accordance with this notion, the negative association with indices related to the middle positions of a turn conformation indicates that the turn-like conformation is inconsistent with the amino acid preference at X. Similarly, the positive association with the extended conformation indices at the X−1 position also suggests that an extended structure is preferred around the −1 position. Finally, the positive association with the helix indices at the X+1 position suggests that the most favorable conformation at the +1 position is helical. In summary, the correlation analyses suggested that the archaeal OST prefers peptide substrates in an extended conformation with a C-terminal helical propensity in the segment of X−1-N-X-S/T-X+1. Another comprehensive N-glycoproteome analysis also suggested an unexpected enrichment of N-glycosylated sequons in predicted β-sheet regions, in comparison to their corresponding non-glycosylated sites (Zielinska et al. 2010). The present study suggests that these preferential distributions of N-glycosylated sequons in extended structures originate from the catalytic properties of the OST enzymes.

What is the biological significance of the tendency of the *P. furiosus* N-glycosylation sequons to extend? Archaeal OSTs probably work in a co-translational mode, although the archaeal N-glycosylation mode has not been studied in detail. We speculate that the conformational tendency of the sequon to extend is relevant to the efficient co-translational scanning of a nascent polypeptide chain for glycosylation sequons in vivo. This implies that the turn-like reactive conformation that brings the side chains of the Asn and hydroxyamino acid residues together is not the major determinant for the amino acid preferences within and flanking the sequons. According to this view, eukaryotic OSTs and other archaeal OSTs that work in a co-translational mode should have amino acid preferences compatible with the tendency of the sequon to extend. The most likely scenario is, however, that the amino acid preferences within and near the N-glycosylation sequons are both influenced by the requirement for the efficient scanning of the polypeptide chains in an extended conformation, and simultaneously by the transient interaction of the sequon with the active site of OST for effective catalysis. The balance between the two factors varies for each OST enzyme. According to this idea, other archaeal OSTs might also have similar, but distinct amino acid preferences. The diversities in the N-glycans and the phosphorylated dolichol species of the other LLO substrates might modulate the amino acid preference of each archaeal OST. Finally, we provide an explanation for the lack of a correlation between the amino acid preference of the canine OST and the tendency to extend. In the canine OST experiment, complicated factors could conceal the amino acid preference, because the OST reaction was not a simple in

---

**Fig. 3.** Dot plots for the detection of positive or negative association between the amino acid preferences of the *P. furiosus* OST and the structural categories of various amino acid indices at the X−2, X and X+1 positions in the N-glycosylation sequon, X−1-Asn-X-Ser-X+1. Notable clusters of dots are enclosed by circles. Note that the correlation analysis at the X−2 position is not presented, because no meaningful correlations were obtained.
vitro assay, but was coupled with in vitro transcription and translation in a cell-free system (Shakin-Eshleman et al. 1996).

The eubacterial _C. jejuni_ OST transfers the oligosaccharide chain to the folded proteins in a post-translational manner (Kowarik, Numao, et al. 2006). Since the scanning of poly-peptide chains is not required, the competency of an N-glycosylation sequon is independent of the tendency to extend. This is in consistent with the finding that no conformational tendency of the peptide substrate was inferred from the correlation analyses (Supplementary data, Figure S3). The requirement of an acidic residue at the −2 position in the eubacterial N-glycosylation sequon may dictate a special geometry of the peptide segment containing the sequons. Thus, the compatibility with this special geometry of the sequon determines the amino acid preferences of the _C. jejuni_ OST.

About one-third of the potential N-glycosylation consensus sequences are not glycosylated in extracellular proteins (Apweiler et al. 1999; Petrescu et al. 2004; Nita-Lazar et al. 2005). This fact indicates that the presence of a consensus is essential, but not sufficient, for N-glycosylation. What determines the actual occurrence of N-glycosylation at a given sequon? Detailed knowledge about the effects of amino acid residues on the glycosylation efficiency, particularly from the viewpoint of the conformation of the sequon, will provide clues to elucidate the sequon selection rules. Certainly, we need more experimental data on the amino acid preferences of OST enzymes from a variety of sources to answer this biologically important question.

**Comparison with previous statistical analyses**

Statistical analyses of the surrounding sequences of experimentally verified glycosylated and unglycosylated sequons in eukaryotic and archaeal glycoproteins were reported (Ben-Dor et al. 2004; Petrescu et al. 2004; Abu-Qarn and Eichler 2007). Statistically significant biases in the probability of finding certain amino acids or groups were detected at particular positions, but no clear patterns were found to distinguish between the glycosylated and unglycosylated sequons. In contrast, experimental analyses of an individual N-glycosylation sequon of a particular protein/peptide revealed the pronounced effects of a single amino acid residue on the glycosylation efficiency of the OSTs in the three domains of life (Shakin-Eshleman et al. 1996; Kasturi et al. 1997; Mellquist et al. 1998; Chen et al. 2007; this study). The apparent discrepancy between the statistical and experimental analyses could be attributable to the independent, comparable impacts of the amino acid variation on the N-glycosylation efficiency at each position. The effect of a particular position on glycosylation is averaged out by the strong influences of the other positions, when many sequences are statistically analyzed. Alternatively, the conformational restraints exerted by the growing polypeptide chain exiting from the translocon channel may have a more decisive role and eliminate the strong dependency on amino acid residues. The amino acid preferences of the _P. furiosus_ OST in this study were obtained from the in vitro assay, in which OST embedded in detergents acted on short-peptide substrates. We would also like to determine whether the same amino acid preference trends are applicable to the in vivo glycosylation in _P. furiosus_ cells, but this remains to be addressed in the future.

**N-glycosylation system for the production of recombinant glycoproteins in bacterial cells**

An innovative glycoprotein production system using engineered _E. coli_ cells was developed (Schwarz, Huang, et al. 2010). The central component of the glycosylation machinery is the _Campylobacter_ OST. The key to success is the extensive research performed in several laboratories over the past decade. Recently, detailed insights into archaeal N-glycosylation, including not only the OST, but also the glycosyltransferases in the dolichol pathway, have emerged from intensive research efforts using model archaean, such as _Halofexx_ and _Methanococcus_ (Abu-Qarn et al. 2008; Magidovich and Eichler 2009). The information on the enzymatic properties of _Pyrococcus_ OST presented here will facilitate the design of a novel in vitro N-glycosylation system using the archaeal N-glycosylation machinery in engineered _E. coli_ cells. A variety of thermophilic archaean are available, as an attractive source of heat-stable, single-subunit OST enzymes for future glyco-engineering applications.

**Materials and methods**

**Peptide library**

Five library sets were synthesized by custom peptide synthesis (PEPscreen, Sigma-Aldrich, St Louis, MO, USA), as summarized in Table 1. A fluorophore, carboxytetramethylrhodamine (TAMRA), was attached to the thiol group of the N-terminal cysteine residue. Each peptide was dissolved at a concentration of 2 mg/mL in 50 mM HEPES buffer, pH 7.4, containing 10 mM NaCl. The pH was neutralized by adding NaOH. One-hundredth volume of 20 mM 5,6-TAMRA-maleimide (Biotium, Hayward, CA, USA) solution in dimethyl sulfoxide was added, and the solution was incubated for 2 h at room temperature. Modified peptides were purified by reversed-phase high-performance liquid chromatography. The peptides resolved into two peaks, corresponding to the 5- and 6-isomers of TAMRA. The slowly eluting peak was collected, dried and re-dissolved in water containing 0.02% Tween 20. The concentration of the peptides was determined by the absorbance at 555 nm, with an extinction coefficient of 90,000 M⁻¹ cm⁻¹.

**OST assay**

The OST assay was performed by the PAGE method (Kohda et al. 2007). The _P. furiosus_ membrane fraction was prepared as previously described (Kohda et al. 2007). The membrane fraction contains the OST enzyme and the LLO donors in the presence of 1% Triton X-100 (v/v). The reaction mixture (total 8 µL) consisted of 6 µL of buffer A (50 mM Tris–HCl, pH 7.5, containing 1 mM DTT, 10 mM MnCl₂, and 0.02% Tween 20), 1 µL of 10 µM peptide and 1 µL of _P. furiosus_ membrane fraction, and was incubated for 1 h at 65°C. The reaction was stopped by the addition of 1.6 µL of 5× sodium dodecyl sulfate (SDS) sample buffer. For kinetic analysis, the appropriate volume of a peptide stock solution and 5 µL of aliquot of _P. furiosus_ LLO (Igura et al. 2008) were combined and dried in a plastic tube, before mixing with 8 µL of buffer A, 1 µL of
immunoaffinity purified *P. furiosus* OST (Igura et al. 2008) and 3 µL of water. This reaction mixture (total 12 µL) was incubated for 1 h at 65°C. The reaction was stopped by the addition of 2.4 µL of 5× SDS sample buffer. The fluorescence image of the sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS–PAGE) gel was recorded with an LAS-3000 multicolor image analyzer (Fuji Photo Film, Tokyo, Japan) and was quantified using the ImageGauge software (Fuji Photo Film). The kinetic parameters for the peptides were determined by a non-linear least-squares fit of the kinetic data to the Michaelis–Menten equation, using KaleidaGraph 4.0 (Synergy Software, Reading, PA, USA).

**Correlation analysis**

Amino acid indices were obtained from the AAindex database, Release 9.1 (Kawashima et al. 2008) and the PDBeMotif database, Release 1.0 (Golovin and Henrick 2008). The indices in the PDBeMotif database were defined as the potentials calculated based on the statistics of small structural motifs. The signs of the numerical values of the amino acid indices connected to the energy terms were inverted to adjust the magnitude of the relation with other amino acid indices. Microsoft Excel was used for the calculation of the correlation coefficients.

**Supplementary data**

Supplementary data for this article is available online at http://glycob.oxfordjournals.org/.

**Funding**

This work was supported by the Ministry of Education, Culture, Sports, Science and Technology (MEXT) of Japan (grant-in-aid for Scientific Research on Innovative Areas and the Targeted Proteins Research Program) and the Mizutani Foundation for Glycosciences (090036).

**Conflict of interest statement**

None declared.

**Abbreviations**

AgIB, archaeal glycosylation B; LLO, lipid-linked oligosaccharide; OST, oligosaccharyltransferase; PglB, protein glycosylation B; SDS–PAGE, sodium dodecyl sulfate–polyacrylamide gel electrophoresis; STT3, staurosporine and temperature sensitive 3; TAMRA, carboxytetramethylrhodamine.

**References**


