Computational design and experimental evaluation of glycosyltransferase mutants: engineering of a blood type B galactosyltransferase with enhanced glucosyltransferase activity

Taku Nakahara¹, Ole Hindsgaul², Monica M. Palcic² and Shin-Ichiro Nishimura¹,³,⁴

¹Laboratory of Advanced Chemical Biology, Graduate School of Advanced Life Science, Frontier Research Center for the Post-Genome Science and Technology, Hokkaido University, Sapporo 001-0021, Japan, ²Carlsberg Laboratory, Gamle Carlsberg Vej 10, DK-2500 Copenhagen Valby, Denmark and ³Drug-Seeds Discovery Research Laboratory, National Institute of Advanced Industrial Science and Technology (AIST), Sapporo 062-8517, Japan.

Glycosyltransferases are an enormous and diverse class of enzyme encompassing 1% of all sequenced genomes. They catalyze the transfer of a monosaccharide from an activated donor such as a sugar-nucleotide to an acceptor molecule. Though the primary sequences of glycosyltransferases have little homology, X-ray structural studies on glycosyltransferases have revealed that there are two main folds and that the orientation of the sugar donors with respect to the folds is highly conserved. It seems that glycosyltransferases have evolved diversified specificities toward donor sugars by changing the amino acids around the monosaccharide moiety without altering the orientation of the nucleotide moiety. In this study, we designed new glycosyltransferases with altered donor specificities by use of a novel empirical model called the Epimer Propensity Index (EPI). The EPI was constructed using 221 carbohydrate–protein complex structures in the Protein Data Bank with either galactose or glucose in the complex. The blood type B synthesizing glycosyltransferase GTB, a galactosyltransferase was our target enzyme. Two GTB mutants designed to exhibit enhanced glucosyltransferase activity were cloned, expressed and characterized experimentally. The predicted GTB mutants, Ser185Asn and Ser185Cys, exhibited 4.3- and 4.8-fold elevations in $k_{cat}/K_m$ for UDP-Glc relative to that of wild-type enzyme.

Keywords: carbohydrate–protein complex/computational design/glycosyltransferase

Introduction

Sugar chains by themselves contain no genetic information. They are synthesized, recognized and degraded by proteins. Various functions of sugar chains have emerged through evolution. Understanding the evolution of the proteins that interact with sugar chains would be very helpful in understanding glycobiological evolution as well as for the engineering of synthetically useful enzymes.

Glycosyltransferases are enzymes that transfer a monosaccharide residue from a sugar donor, such as a sugar-nucleotide, to an acceptor molecule. Glycosyltransferases have been classified into 79 distinct families based on their amino acid sequences (CAZy, January 2006 release; Coutinho et al., 2003). This classification utilizes BLAST (Altschul et al., 1997) so that homologous sequence pairs with a BLAST e-value better than $10^{-5}$ are grouped together. Despite the numerous classes based on amino acid sequences, the three-dimensional crystal structures of all glycosyltransferases solved to date have only two folds termed the GT-A fold and GT-B fold (Bourne and Henrissat, 2001; Davies et al., 2005). In all GT-A fold glycosyltransferases, a DXD motif anchors the pyrophosphate moiety of the sugar-nucleotide donor via a divalent cation such that the location of the sugar donor on the fold is conserved [sialyltransferases are an exception because they utilize CMP-sialic acid, which has monophosphate moiety instead of pyrophosphate moiety (Chiu et al., 2004)]. This suggests that glycosyltransferases evolved to recognize various donors by changing amino acids around the sugar moiety of the donor, without changing the major mode of interaction with the nucleotide moiety. This idea is supported by the human blood type glycosyltransferases. The human genome codes for two types of blood group glycosyltransferase genes A and B. Blood type A glycosyltransferase (GTA) is an N-acetylgalactosaminyltransferase which transfers N-acetylgalactosamine from a UDP-GalNAc donor to the disaccharide H-antigen to synthesize the A-antigen. Blood type B glycosyltransferase (GTB) is a galactosyltransferase that transfers a galactose from UDP-Gal to the same H-antigen to synthesize the B-antigen. Blood type O is a consequence of a nucleotide deletion or substitution yielding inactive enzyme. There are only four amino acid differences between GTA and GTB, and only two of these (Leu/Met-266 and Gly/Ala-268 of GTA/GTB) are critical for the donor specificity (Seto et al., 1997, 1999). There are also several reports of donor specificity engineering of glycosyltransferases accomplished by changing a few amino acids (Ouzzine et al., 2002; Kubo et al., 2004; Ramakrishnan et al., 2005). In these examples of donor specificity engineering, the amino acids targeted for mutagenesis were selected on the basis of at least partial structural information on the enzyme. The ability to quantitatively predict the donor specificity of the mutated enzyme in silico would greatly facilitate this process.

In this study, we aim to establish a protein design method to engineer the donor specificities of glycosyltransferases by mimicking the evolutionary process of glycosyltransferases. The primary focus is on the orientation (configuration) of the hydroxyl groups on the sugar being transferred. For example, OH-4' of D-glucose in the $^{13}$C$_1$ chair is directed in the equatorial orientation. For D-galactose, the $^{13}$C$_1$ chair conformer is same as the D-glucose $^{13}$C$_1$ chair conformer except that OH-4' is in the axial orientation. The purpose of this study was to develop an index to distinguish between a protein + equatorial-OH sugar complex and a protein + axial-OH sugar complex, and to apply this index to the design of novel glycosyltransferases.

There is abundant information about protein–sugar interactions in the Protein Data Bank (PDB; Berman et al., 2000).
where many sugar structures are co-crystallized with proteins. In a previous study, all of the sugar structures in the PDB were extracted by our original program getCARBO in order to compile a database of these structures (Glycoconjugate Data Bank: Structures available at http://glycoconjugate.jp/structures.html and http://daisy2.nagahama-i-bio.ac.jp/structures). There were 23,608 sugar residues in the PDB (September 2005 release, 32,658 entries). Details on the program and the database will be reported elsewhere. In this study, we utilize this database to establish an empirical index to distinguish between protein + equatorial-OH sugar complexes and protein + axial-OH sugar complexes. We name this index the Epimer Propensity Index (EPI). EPIs were constructed using protein–[sugar bearing equatorial-OH group] complexes and protein–[sugar bearing axial-OH group] complexes, separately. Differences between $\Delta$EPIequatorial and $\Delta$EPIaxial can be used as an index to predict whether the OH group of a putative protein–sugar complex would prefer to be equatorial or axial.

Here, we develop a $\Delta$EPI field for the 4'-OH group of pyranoses in the $\text{C}_4 \text{I}$ chair conformer. As a first test case, the $\Delta$EPI field was applied to wild-type $\beta$-1,4-galactosyltransferase I ($\beta$GalT-I) and its Arg228Lys mutant which exhibits enhanced glucosyltransferase activity (Ramakrishnan et al., 2005). We then applied this method to GTB to predict mutants with enhanced glucosyltransferase activity. Based on our prediction, we experimentally examined whether these mutants transferred glucose more effectively than the wild-type enzyme.

Materials and methods

Dataset preparation

Extraction of sugar structures from the PDB. A scheme for dataset preparation is depicted in Figure 1. The structures of sugars were extracted from the PDB using the program, 'getCARBO'. Details on this program will be reported elsewhere (TN, in preparation). The full dataset of sugar structures in the PDB can be seen in the Glycoconjugate Data Bank:structures. The PDB released in September 2005 that contains 32,658 entries was used. The nomenclature of the sugar structures in the PDB followed IUPAC recommendations (IUPAC Nomenclature of carbohydrates, 1996).

Preparation of a non-redundant sugar residue set. All of the retrieved sugar structures were clustered according to protein sequence, sugar residue structure and sugar location in the protein structure. In this study, only PDB entries with resolutions better than 2.0Å were used. Protein sequences of the PDB entries that contain sugar residue(s) were clustered by using blastclust (Altschul et al., 1997). The threshold of this clustering was 60% identity. In each protein sequence cluster, all of the combinations of sequence pairs were aligned by using bl2seq (Altschul et al., 1997). Based on the alignments, pairs of proteins were superimposed on the corresponding Cα atoms. If two sugar residues with the same IUPAC name had the center-of-mass of each sugar residue located within 4Å in the superimposed coordinates, then these two sugars were clustered into the same group. The sugar residue with the best resolution within a sugar residue cluster was selected to represent that cluster. These representative sugar residues were used as the non-redundant sugar residue set.

$\Delta$EPI field construction

Weighting atom types. Protein atom distributions around the OH group of sugar molecules were assigned to grid points in three-dimensional Cartesian coordinate space. In this study, only the 4'-OH group of $\text{C}_4 \text{I}$ chair conformers were examined, however, this methodology can be extended to other OH-groups. The dataset for $\Delta$EPI field construction consisted of representative sugar residues in the $\text{C}_4 \text{I}$ chair conformation with their 4'-OH group free (i.e. non-glycosylated). The grid points were defined every 0.2 Å along the three axes and spread within 4Å from the C4 atom of the sugar ring. A given protein–sugar complex was superimposed on the reference system according to three atoms: C3, C4 and C5 (Figure 2A). The C4 atom of the ring was set to the origin, the C3 atom was set along the x-axis, and the C5 atom lay on the xy-plane. Protein atoms were classified into four types as follows: hydrogen bond acceptor, hydrogen bond donor, both acceptor and
**Fig. 2.** (A) Reference system of ΔEPI field. All the representative protein–sugar complexes were superimposed on this reference system. The C4 atom of the ring was set to the origin, the C3 atom was set along the x-axis, and the C5 atom lay on the xy-plane. The shaded area indicates the range of grid point assignment, within 4Å from the C4 atom. (B) Protein atom type distribution around equatorial 4′-OH and axial-4′OH groups. Atom type distribution of equatorial-4′OH sugar complexes (left) and axial-4′OH sugar complexes (right) are depicted. All the sugar residues used were in the 4C1 conformation. Spheres are protein atoms within 4Å from the C4 atom. Spheres colors for hydrogen bond acceptor, donor, and neutral were red, blue, green and yellow, respectively. This figure was created using PyMOL (DeLano, 2002).

... donor, and neutral (Jiang et al., 2002). Each atom type contains the following protein atoms (atom nomenclature used here follows that used in the PDB); acceptor of hydrogen bond: backbone O, OD of ASN, OE of GLN, carboxyloxigen in ASP and GLU; donor of hydrogen bond: NE of TRP, ND of ASN, NE of GLN, and N of LYS and ARG; both acceptor and donor: hydroxyl oxygen, ND and NE of HIS, carboxyloxigen in C-terminal, and oxygen of water molecule; neutral: carbon and sulfur atoms and N of PRO. The weights of each atom type were assigned on the grid points according to the position of the protein atoms on the reference system. The weight of a protein atom assigned on a grid point follows a Gaussian curve:

\[ w_p(a, r) = e^{-(r/r_{cut})^2} \]

where \( w_p(a, r) \) is the weight of protein atom type \( a \) assigned on a grid point \( p \), and \( r \) is a distance (Å) between \( a \) and \( p \). The term \( r_{cut} \) defines a reach of weighting. We used 2.0Å for \( r_{cut} \). We only assigned exp(-\((r/r_{cut})^2\)) > 0.01 (weighting cut-off).

**Construction of EPI field.** Weights on the grid points were normalized by dividing by the summation of all the weights assigned on all the grid points so that the total weight assigned on the grid space is 1. Here, we define \( W(a,p) \) as a summation of \( a \) on \( p \). Probability of \( a \) on grid point \( p \) is defined as

\[ f(a,p) = \frac{W(a,p)}{\sum_a \sum_p W(a,p)} \]

Then, we define an empirical potential, named EPI:

\[ EPI = \ln(f(a,p)) - C \]

where \( f(a,p) \) is the normalized weight of atom type \( a \) assigned on grid point \( p \), and \( C \) is a term of the partition function. The EPI field of protein–sugar that bears equatorial 4′-OH group complexes and that of protein–sugar which bears axial 4′-OH group complexes were constructed separately (Figure 2B). We prepared the ΔEPI field by subtracting EPI\(_{axial} \) from EPI\(_{equatorial} \):

\[ \Delta EPI = EPI_{equatorial} - EPI_{axial} = \ln \frac{f_{equatorial}(a,p)}{f_{axial}(a,p)} \]

In the above equation, we assume \( C \) of EPI\(_{equatorial} \) and that of EPI\(_{axial} \) are equal. ΔEPI score of a given protein–sugar complex is calculated by

\[ \Delta EPI_{score} = \sum_{i=1}^{N} \sum_{p=1}^{P} \Delta EPI_{i,p} \cdot e^{-\left(\frac{r}{r_{cut}}\right)^2} \]  

where \( N \) is the total number of protein atoms in the given complex, \( P \) is the number of grid point, \( \Delta EPI_{i,p} \) is ΔEPI value of protein atom \( i \) at grid point \( p \), and \( r \) is a distance between \( i \) and \( p \). A positive value of the ΔEPI score indicates a OH group of focused position of a sugar complexed with a protein tends to be equatorial, and a negative value of the ΔEPI score indicates a OH group of that tends to be axial.

**Evaluation of ΔEPI field.** A new dataset was prepared for evaluation of the constructed ΔEPI field. The new dataset contained the representative sugar residues which have more than one protein atom within 4Å of the C4 atom. Separation of protein + equatorial 4′-OH sugar complex and protein + axial 4′-OH sugar complex by ΔEPI was quantified by a square-root correlation ratio (\( \eta \)):

\[ \eta = \sqrt{\frac{N_{equatorial}/N_{all}(\mu_{equatorial}−\mu_{all})^2}{N_{equatorial}/N_{all}(\mu_{equatorial}−\mu_{all})^2 + N_{axial}/N_{all}(\mu_{axial}−\mu_{all})^2}}/\sigma_{all} \]

where \( N_{all} \) is the total number of protein–sugar complexes, \( N_{equatorial} \) is the total number of protein + equatorial 4′-OH sugar complexes, \( N_{axial} \) is protein + axial 4′-OH sugar complexes, \( \mu_{all} \) is average ΔEPI score of all protein–sugar complexes, \( \mu_{equatorial} \) is average ΔEPI score of all protein + equatorial 4′-OH sugar complexes, \( \mu_{axial} \) is average ΔEPI score of all protein + axial 4′-OH sugar complexes, and \( \sigma_{all} \) is standard deviation of all protein–sugar complexes. \( \eta \) ranges between 0 and 1 and higher values of \( \eta \) indicate better discrimination between equatorial/axial by the ΔEPI field. For cross-validation, the dataset for ΔEPI evaluation was randomly divided into two groups (group 1 and group 2). ΔEPI field constructed by using group 1 and ΔEPI field constructed by group 2 were prepared, and cross-validations (using group 1 as samples evaluated by ΔEPI field of group 2 and vice versa) were also done.

**ΔEPI evaluation of β4GalT-I.** The structures used to evaluate ΔEPI values of β4GalT-I were 1O0R (wild-type; Ramakrishnan et al., 2002) and 1YRO (Arg228Lys, glycosyltransferase enhanced mutant; Ramakrishnan et al., 2005). 1O0R is the entry for wild-type β4GalT-I complexed with UDP-Gal. 1YRO is the entry for the Arg228Lys mutant complexed with UDP-Gal and α-lactalbumin. The resolution of 1O0R is 2.30Å and that of 1YRO is 1.90Å. Calculations of the ΔEPI score were done with all atoms in these PDB entries.

**Molecular modeling of blood type B glycosyltransferase and its mutants**

**Modeling of the missing loop and donor sugar of wild-type GTB.** The missing loop of GTB (Glu174-Arg198) and the missing sugar moiety of the donor sugar were modeled using the molecular modeling platform MOE (Chemical Computing Group Inc.) with the MMFF94s force field (Halgren, 1996). The main template for the GTB model was the crystal structure of the blood group-B glycosyltransferase (PDB ID: 1LZJ; Patenaude et al., 2002). The missing loop was complemented by the bovine α1,3-galactosyltransferase complex with UDP-Gal (PDB ID: 1G93; Gastinel et al., 2001). The missing loop on 1LZJ (Glu174-Arg198) was built with the homology
modeling program in MOE using Lys185-Ala212 of 1G93 as the template for the missing loop (Figure 4). The sugar moity of the donor was modeled in accordance to the position of the sugar moiety of the donor in 1G93. The 4'-OH group was inverted to equatorial to build the wild-type GTB + UDP-Glc complex.

**Modeling of mutant GTBs.** Molecular modeling of GTB mutants was carried out using a function implemented in MOE. Only the side chain at position 185 of the wild-type GTB mutants was carried out using a function implemented in MOE. In addition to the first modeled mutant structure, the same mutant UDP-Glc was changed to other amino acids except proline. In AC-3 (pH 7.0), 0.5 M NaCl, 1 mM dithiothreitol, 5 mM MnCl2 (Seto et al., 2000; Marcus et al., 1997). Tritium-labeled donors were used. For kinetic characterization, seven reaction mixtures which contained different concentrations of donor, 500 μM acceptor and the enzyme were incubated at 37 °C for an appropriate time (depending on the activity of the enzyme). The assay buffer contained 50 mM MOPS buffer (pH 7.0), 20 mM MnCl2 and 1 mg/ml bovine serum albumin. The reactions were stopped by dilution with water. Radioactive products of the enzyme reaction were isolated by reverse-phase extraction (Sep-Pak plus C-18 cartridge, Millipore-Waters), and quantified by counting on a Beckman LS 6500 counter after the addition of 10 ml EcoLite+ liquid scintillation cocktail. The kinetic parameters \( k_{cat} \) and \( K_m \) were calculated by non-linear regression analysis of the Michaelis–Menten equation using a PRISM program (GraphPad Software).

**Results**

**Dataset preparation**

To construct an empirical potential, a non-redundant sugar residue set was prepared (Figure 1). This was based on structural information for sugar residues in the PDB that was retrieved using the getCARBO program we have developed. The program getCARBO extracts sugar residue structures from a PDB format file. It detects and annotates sugar residue structures by using coordinates and element type of atoms in a PDB file. Of the 32 658 PDB entries that were processed by getCARBO, 5673 of the entries contained 23 608 sugar residues in 15 868 glycan structures. Only ligand sugar residues were used because we focused on recognition of sugars by proteins; the set contained 16 101 residues in 11 776 glycans.

To construct a reliable empirical potential, only PDB entries with resolutions better than 2Å were used. Only 2457 of 5673 PDB entries had the required resolution. A non-redundant sugar residue set was prepared by using protein sequence similarity, IUPAC nomenclature of sugar structures and sugar residue locations on homologous proteins. Initially, sequences of proteins that form complexes with sugar residue(s) were clustered using blastclust (Altschul et al., 1997). Sequences that showed more than 60% identity were clustered. A total of 3311 protein sequences of 2457 PDB entries were clustered into 843 sequence clusters. All the combinations of protein sequence pairs in a sequence cluster were aligned using bl2seq (Altschul et al., 1997). All the protein sequence pairs were superimposed on the corresponding C\(_2\) atoms in the alignment. In superimposed structures, sugar residues with center-of-masses that were within 4.0 Å were clustered. The sugar residue with the best resolution in a cluster. After these processes, 2629 representative sugar residues were selected.

In this work, we focused on the 4'-OH group of the 4'C\(_2\) chair conformer. There were only 525 representative sugar residues that satisfied the following conditions: (i) 4'C\(_2\) chair conformer, (ii) 4'-OH group was not glycosidically bonded. These included 410 4'-equatorial sugar residues and 115 4'-axial sugar residues. These 525 sugar residues were used to construct the ΔEPI field.
To evaluate the ΔEPI field, only 221 of 525 sugar residues that had more than one protein atom within 4Å from the C4 atom were used (See supplementary material for the list of the 221 structures). We excluded 304 sugar residues because these had only one or no protein atom within 4Å of the C4 atom. ΔEPI scores of these sugar residues were very small. A substantial number of these barely recognized 4'-OH groups were not suitable for evaluation of discrimination ability of ΔEPI (η). These 221 sugar residues contained 154 O4'-equatorial sugar residues and 67 O4'-axial sugar residues.

Construction and evaluation of ΔEPI
An empirical potential field, the ΔEPI field of 4'-OH of C1, chair conformers, was constructed as described in Materials and methods. A total of 525 representative sugar residues and protein atoms around these sugar residues were superimposed on the reference system depicted in Figure 2A. The protein atoms were classified into four groups based on their hydrogen bonding properties; donor, acceptor, both donor and acceptor, and neutral. The weights of protein atoms were assigned on grid points using a Gaussian curve function (Equation 1). All of the weights were normalized and converted to ΔEPI fields by Equations 3 and 4. The ΔEPI score of a given protein–sugar complex was calculated using Equation 5.

The ΔEPI field was evaluated by using a squared-root correlation ratio (η, Equation 6) of protein + equatorial-4'-OH sugar complexes and protein + axial-4'-OH sugar complexes. η is an index that expresses the degree of separation between protein + equatorial 4'-OH sugar complexes and protein + axial 4'-OH sugar complexes.

Self-evaluation and cross-validations were done. For cross-validations, the evaluation dataset (221 sugar complexes) was divided into two groups and methods. A total of 525 representative sugar residues and protein atoms around these sugar residues) was divided into two groups and methods. A total of 525 representative sugar residues and protein atoms around these sugar residues

ΔEPI evaluation of β4GalT-I and its mutant
β4GalT-I is a glycosyltransferase that transfers galactose from UDP-Gal to OH-4 of N-acetylglucosamine. A mutant of this enzyme (Arg228Lys) is known to have enhanced glycosyltransferase activity. Both wild-type (PDB ID: 1O0R) and mutant (1YRO) crystal structures had been solved.

Calculation of the ΔEPI score was done using all atoms in a PDB file including water molecules. The ΔEPI score of the wild-type enzyme was −2163, and that of the Arg228Lys mutant was −536 (Table I). The glycosyltransferase-enhanced mutant had a ΔEPI score 1627 points higher than the wild-type β4GalT-I enzyme.

Molecular modeling and ΔEPI evaluation for GTB and its mutants
In the structure of GTB deposited in the PDB (PDB ID: 1LZJ), there is no electron density for a flexible loop (Glu174–Arg198, Figure 4). We therefore modeled in these residues using homology modeling. The galactose moiety of the sugar donor was also added. The structure of a homologous galactosyltransferase, χ1,3-galactosyltransferase (PDB ID: 1G93) was used as a template structure for the homology modeling. The latter structure was solved with UDP-Gal. Amino acids which interact with the galactose moiety of the donor on the modeled fragment of χ1,3-galactosyltransferase are completely conserved between GTB and χ1,3-galactosyltransferase. The direction of the 4'-OH group was inverted to equatorial and the entire modeled structure was optimized using molecular mechanics with the MMFF94s force field.

Table I. Values of Epimer Propensity Indexes and kinetic parameters for GTB, β4GalT-I and their mutants

<table>
<thead>
<tr>
<th>Kinetic parameters</th>
<th>Fold</th>
<th>ΔEPI</th>
<th>ΔEPI</th>
</tr>
</thead>
<tbody>
<tr>
<td>UDP- galactose</td>
<td>UDP- glucose</td>
<td>GlcT&lt;sup&gt;a&lt;/sup&gt;</td>
<td>GlcT/ GalT&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td>GTB</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wild-type</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>k&lt;sub&gt;cat&lt;/sub&gt; (s&lt;sup&gt;-1&lt;/sup&gt;)</td>
<td>5.2</td>
<td>0.00053</td>
<td>1.0</td>
</tr>
<tr>
<td>K&lt;sub&gt;m&lt;/sub&gt; (µM)</td>
<td>53</td>
<td>188</td>
<td>0.10</td>
</tr>
<tr>
<td>k&lt;sub&gt;cat&lt;/sub&gt;/K&lt;sub&gt;m&lt;/sub&gt; (µM&lt;sup&gt;−1&lt;/sup&gt;)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ser185Asn</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>k&lt;sub&gt;cat&lt;/sub&gt; (s&lt;sup&gt;-1&lt;/sup&gt;)</td>
<td>0.52</td>
<td>0.0014</td>
<td>2.7</td>
</tr>
<tr>
<td>K&lt;sub&gt;m&lt;/sub&gt; (µM)</td>
<td>166</td>
<td>120</td>
<td>0.0031</td>
</tr>
<tr>
<td>k&lt;sub&gt;cat&lt;/sub&gt;/K&lt;sub&gt;m&lt;/sub&gt; (µM&lt;sup&gt;−1&lt;/sup&gt;)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ser185Cys</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>k&lt;sub&gt;cat&lt;/sub&gt; (s&lt;sup&gt;-1&lt;/sup&gt;)</td>
<td>0.692</td>
<td>0.0020</td>
<td>3.8</td>
</tr>
<tr>
<td>K&lt;sub&gt;m&lt;/sub&gt; (µM)</td>
<td>33</td>
<td>152</td>
<td>0.0028</td>
</tr>
<tr>
<td>k&lt;sub&gt;cat&lt;/sub&gt;/K&lt;sub&gt;m&lt;/sub&gt; (µM&lt;sup&gt;−1&lt;/sup&gt;)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>β4GalT-I&lt;sup&gt;c&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wild-type</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>k&lt;sub&gt;cat&lt;/sub&gt; (s&lt;sup&gt;-1&lt;/sup&gt;)</td>
<td>14.0</td>
<td>0.059</td>
<td>1</td>
</tr>
<tr>
<td>K&lt;sub&gt;m&lt;/sub&gt; (µM)</td>
<td>0.15</td>
<td>0.0004</td>
<td>1.0</td>
</tr>
<tr>
<td>k&lt;sub&gt;cat&lt;/sub&gt;/K&lt;sub&gt;m&lt;/sub&gt; (µM&lt;sup&gt;−1&lt;/sup&gt;)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Arg228Lys</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>k&lt;sub&gt;cat&lt;/sub&gt; (s&lt;sup&gt;-1&lt;/sup&gt;)</td>
<td>0.44</td>
<td>0.16</td>
<td>2.8</td>
</tr>
<tr>
<td>K&lt;sub&gt;m&lt;/sub&gt; (µM)</td>
<td>0.0050</td>
<td>0.0018</td>
<td>4.5</td>
</tr>
<tr>
<td>k&lt;sub&gt;cat&lt;/sub&gt;/K&lt;sub&gt;m&lt;/sub&gt; (µM&lt;sup&gt;−1&lt;/sup&gt;)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<sup>a</sup>Fold of k<sub>cat</sub>/K<sub>m</sub> for glucosyltransferase in comparison with the wild-type.
<sup>b</sup>Fold of k<sub>cat</sub>/K<sub>m</sub> for glycosyltransferase to k<sub>cat</sub>/K<sub>m</sub> for galactosyltransferase in comparison with the wild-type.
<sup>c</sup>All the kinetic parameters of β4GalT-I were from Ramakrishnan et al. (2005).
Ser185 was selected as a mutagenesis target because this amino acid is located closest to the 4'-OH position of the donor sugar. We therefore modeled GTB mutants in which Ser185 was transformed to 18 different amino acids. We did not mutate to proline because proline mutation often causes substantial rearrangement of nearby amino acids. Three model structures were prepared for each mutant structure whose χ angle of the mutated side chains were different (0°, +120° and −120°), except alanine and glycine. These initial models were relaxed and the most energetically stable structure was selected as a mutant model.

Mutated GTB is required to (i) exhibit enhanced glucosyltransferase activity, and (ii) be energetically stable. ΔEPI values and potential energies of each mutant model were calculated (supplementary data), and only aspartic acid, glutamic acid and asparagine met these criteria. The ΔEPI score of the wild-type was −260, and that of the Ser185Asn, Ser185Asp, Ser185Glu mutant were 201, 267, 940, respectively.

**Characterization of glucosyltransferase-activity-enhanced GTB**

We prepared expression constructs for the selected mutants (Ser185Asn, Ser185Asp and Ser185Glu) and expressed them in E.coli. The Ser185Cys mutant was also expressed because it had been cloned previously for another purpose. According to the modeled structure, the ΔEPI score of the Ser185Cys mutant was −189, a slightly positive shift from that of wild-type enzyme. The potential energy of this mutant was 17.0 kcal/mol higher than that of wild-type GTB.

Wild-type and mutated GTB were expressed and purified. Enzymatic activity was initially quantified by measuring the transfer of monosaccharide from radiolabeled donor substrate in a standard assay. The Ser185Asp and Ser185Glu mutant had considerably higher specific activities with UDP-Gal, 4.9 mU/mg and 1.9 mU/mg, respectively, which are 0.09 and 0.04% of the specific activity of wild-type GTB (5.2 U/mg). The ratio of glucosyltransferase/galactosyltransferase activity was 27 times greater than wild-type enzyme based on k_cat and 135 times great based on k_cat/K_M.

The Ser185Cys mutant exhibited 3.8 times greater k_cat with UDP-Glc (0.0020 s⁻¹) compared to wild-type enzyme while the k_cat for UDP-Gal was 57 times lower than wild-type enzyme. The ratio of glucosyltransferase/galactosyltransferase activity increased 219-fold in comparison with the wild-type activity based on k_cat and 168-fold based on k_cat/K_M.

**Discussion**

In this study, we introduce a new index, ΔEPI, for discrimination between protein–sugar complexes that bear equatorial-4'OH and axial-4'OH groups, and our experimental results showed it can predict donor specificities for β4GalT-I and GTB.

To construct ΔEPI, we selected a non-redundant sugar residue set from all of the sugar structures in the PDB. A total of 525 of 23608 sugar residues were selected as the representative sugar residue dataset (Figure 1). The atom distributions are shown in Figure 2B. Hydrogen bonding atoms (red/blue/green spheres in Figure 2) distributed around the hydroxyl group of each form, and neutral atoms (yellow spheres) distributed on the opposite side of the hydroxyl group. High density of neutral atoms around the B face of axial-4'-OH sugars is well known (Sujatha et al., 2004), and it was reproduced in this study. Interestingly, a high density of neutral atoms was also found on the A face of equatorial-4'-OH sugars.

The quality of carbohydrate structure information in the PDB is controversial. Lütteke et al. have pointed out the existence of many erroneous carbohydrate structures in the PDB and service a validation program at their website GLYCOSCIENCES.de (Lütteke and von der Lieth, 2004; Lütteke et al., 2004). In this study, we extracted carbohydrate structures with getCARBO, a program that extracts carbohydrate structures in a PDB format file. getCARBO extracts carbohydrate structures based on atom types and the coordinates of a molecule. Because of this, the carbohydrate residues used for EPI field construction do not contain errors from the viewpoint of their atom types and coordinates.

Using this dataset, the ΔEPI field was constructed and this interaction field could distinguish epimer propensity in self-evaluation (Figure 3, η = 0.716). In cross-validations, η was 0.691 when group 1 ΔEPI field evaluates group 2 samples and 0.629 when group 2 ΔEPI field evaluates group 1 samples. Because the η values were reasonably high and cross-validations showed about the same high η points, this ΔEPI field was considered to give a good index to discriminate the epimer propensity for the protein–sugar complexes existing in the PDB.

Calculation of the ΔEPI score was applied to two specific glycosyltransferases, β4GalT-I and GTB. In the case of β4GalT-I, the wild-type structure complexed with UDP-Gal, and the Arg228Lys mutant structure complexed with UDP-Gal, were known. The Arg228Lys mutant had been shown to transfer glucose more efficiently than galactose. By mutating the arginine to lysine, the side chain conformation of Glu317, which is proximal to the lysine, was changed. In
this altered conformation, the glutamic acid became a steric hindrance to OH4’ of the galactose moiety of the donor. Additionally, the mutation created a space which is suitable for binding to equatorial O4 of glucose (Ramakrishnan et al., 2005). The ΔEPI score of the wild-type β4GalT-I was –2163 and that of Arg228Lys mutant was –536 (Table I). The ΔEPI score increased by 1626 points in the Arg228Lys mutant, and the transfer of UDP-Glc was increased 2.8 times based on kcat and 4.5 times based on kcat/Km in the Arg228Lys mutant (Table I). The ΔEPI score therefore correctly reflected the enhanced glucosyltransferase activity. The ΔEPI score for the wild-type reflects the position of guanidine group of Arg288 which is favorable for galactose binding, and the positive shift of ΔEPI score by mutating this residue is a consequence of an absence of the guanidine group.

In the case of GTB, we had to model the GTB + UDP-Gal complex because there were missing residues in the structure deposited in the PDB. The missing internal loop that interacts with the sugar moiety of the donor was modeled using homology modeling. The structure of bovine x1,3-galactosyltransferase was used as the template for the missing loop on GTB since the amino acids which interact with the donor sugar are completely conserved between GTB and x1,3-galactosyltransferase on the modeled fragment (Figure 4). The location of the missing sugar moiety of the donor sugar was modeled in according to the location of the sugar moiety commonly seen in other GT-A fold glycosyltransferase structures. The manner of donor sugar binding in the GT-A fold glycosyltransferases is highly conserved. The nucleotide pyrophosphate moiety is anchored to the conserved DXD motif via a divalent cation and the sugar moiety is bent backwards towards the base moiety.

Homology modeling of the GTB structure complexed with UDP-Gal has been previously carried out by Heissigerova et al. In our model GTB was complexed with UDP-Glc and we retained water molecules in the template structure. Interactions between the sugar moiety of the donor and protein amino acids are slightly different in the models. In Heissigerova et al., Ser185 is hydrogen bonded to O6’ of the donor sugar, while in our model, Ser185 is closer to O4’ than O6’. This difference is attributed to the water molecule in our model that is sandwiched between Arg188 and Asp302 (Figure 5). This water molecule is also seen in the almost same position on the structure of the x1,3-galactosyltransferase (1G93, HOH 84). And Ser199 on 1G93 is closer to O4’ of donor sugar than O6’, as reproduced in our model.

We selected Ser185 as a target of mutagenesis and designed mutant structures in silico. Then, we calculated potential energies and ΔEPI values for them. Ser185Asn, Ser185Asp and Ser185Glu were selected as promising mutants. Enzymatic characterizations of these promising mutants and Ser185Cys mutant, which had already been constructed for another purpose, were carried out. Shifts of ΔEPI value of Ser185Asn, Ser185Cys, Ser185Asp, Ser185Glu mutagenesis were 461, 71, 527, and 1200, respectively.

Experimental characterization of these mutants revealed that Ser185Asp and Ser185Glu had very low activity; therefore, we did not determine their kinetic parameters. However, the ratios of specific activities for UDP-Glc/UDP-Gal enhancement were 39 times and 79 times for Ser185Asp and Ser185Glu, respectively, which correlate to those of ΔEPI. Further characterizations were carried out only for Ser185Asn and Ser185Cys. Experimental determination of the kinetic parameters of the predicted mutants revealed that kcat of the Ser185-Asn mutant with UDP-Glc was 2.7 times higher than that of wild-type enzyme. The positive value of the ΔEPI score for this enzyme correctly predicted this glucosyltransferase enhancement. In the model of this mutant, the Nσ atom of the Asn185 side chain was located favorably for protein–equatorial 4’OH interaction (Figure 6). We speculate that this is the reason for the high positive ΔEPI score and for the high glucosyltransferase activity of this mutant. In the model of the Ser185Cys mutant, the χ angle of the Cys side chain was different from that of the wild-type Ser side chain. This creates space for the equatorial-4’OH group of UDP-Glc and leads to a rearrangement of the orientation between the donor sugar and surrounding amino acids. It results in slight positive-shift of the ΔEPI score and may be the explanation for the 3.8 times higher activity with UDP-Glc.

The aim of this work was to develop a method to design glycosyltransferases that exhibits novel donor specificity. Protein–sugar complexes existing in the PDB and, as a test case, β4GalT-I and its mutant, were favorably evaluated by the ΔEPI field presented in this work. In the case of GTB, two mutants (Ser185Asn and Ser185Cys), that were computationally predicted to have enhanced glucosyltransferase activities, were experimentally revealed to exhibit the predicted activities. However, the degree of positive shift of ΔEPI was not consistent with the experimental results. The enhancement of glucosyltransferase activity was larger in the Ser185Cys mutant despite the fact that the positive shift of the ΔEPI score was larger in the Ser185Asn mutant. Because the missing sugar moiety of the donor sugar and the amino acids around that sugar moiety were built by molecular modeling, the structure used in this work may differ somewhat from the true structure contributing to the poor correlation. A better template structure containing the missing loop structure or more advanced side chain modeling may help to improve the quantitation. Even though there still remains room for improvement.
of the ΔEPI evaluation, the predicted GTB mutants in this work did exhibit improved glucosyltransferase activities. Improvement of glucosyltransferase activity of the GTB mutants compared favorably with that for the β4GalT-I mutant (Table 1).

Several empirical approaches for protein–ligand interaction have been reported (Boer et al., 2001; Shionyu-Mitsuyama et al., 2003). These approaches were used to investigate the interaction sites on a protein or to search for ligands that bind to a protein, but not to design a new protein. In the present work, we specifically sought to design a new glycosyltransferase with altered donor specificity. Because the binding mode of the donor sugar is conserved among GT-A fold glycosyltransferases, we assumed that new donor specificity can be created by changing amino acid(s) around the donor sugar without changing the major binding mode. This would mimic the evolutionary emergence of the polyspecificity of glycosyltransferases. The ΔEPI score was designed to evaluate such an evolutionary model of glycosyltransferases.

In summary, this report presents a new methodology to engineer altered donor specificity into glycosyltransferases. An empirical score, termed ΔEPI, successfully evaluated sugar–OH protein interactions using three-dimensional structures of proteins. By using the ΔEPI score, mutant galactosyltransferases (GTB mutants, Ser185Asn and Ser185Cys) were predicted to transfer glucose more efficiently. The predicted mutants were prepared, and experimentally confirmed to transfer glucose more efficiently.

Acknowledgements

We thank Dr. Hannelore Peters for the gift of the S185C mutant and Ms Sidsel Ehlers for expert technical assistance in the characterization of the GTB mutants.

References


