Available online at www.sciencedirect.com
ScienceDirect
Journal of Molecular Graphics and Modelling 26 (2007) 255–268
Journal of Molecular Graphics and Modelling
www.elsevier.com/locate/JMGM

Fold-recognition and comparative modeling of human β3GalT I, II, IV, V and VI and β3GalNAcT I: Prediction of residues conferring acceptor substrate specificity

Ronak Y. Patel, Petety V. Balaji *

School of Biosciences and Bioengineering, Indian Institute of Technology Bombay, Powai, Mumbai 400076, India

Received 23 May 2006; received in revised form 19 November 2006; accepted 10 December 2006
Available online 15 December 2006

Abstract

β3GalTs are type II transmembrane proteins that transfer galactose from UDP-Gal donor substrate to acceptor GlcNAc, GalNAc or Gal in β1 → 3-linkage. β1 → 3-linked galactose have been found to be a part of many glycans like glycosphingolipids, core tetrasaccharide of proteoglycans, type 1 chains. The 3-D structure of none of the β3GalTs is known to date. In this study, the 3-D structures of human β3GalT I, II, IV, V, VI and β3GalNAcT I have been modeled using fold-recognition and comparative modeling methods. Residues that constitute the UDP-Gal binding site have been predicted. The models are able to qualitatively rationalize data from the site-directed mutagenesis experiments reported in the literature. Residues likely to be involved in conferring differential acceptor substrate specificity have been predicted by a combination of specificity determining positions prediction (SDPs) and subsequent mapping on the generated 3-D models.

Keywords: Homology modeling; Specificity determining positions; Glycosyltransferase; β3-Galactosyltransferase; GT-A fold; Spore coat polysaccharide biosynthesis protein SpsA

1. Introduction

Several types of glycans contain the β1→3-linked galactose [1]. Examples include glycosphingolipids GM1, Gal-Gb_4, histo-blood group A associated Gal-A glycolipids, core tetrasaccharide of proteoglycans (Galβ3Galβ4Xyl), type 1 chains (Galβ3GlcNAcβR) and mucin-type core 1 (Galβ3GalNAcαR). The transfer of galactose from UDP-Gal donor substrate to acceptor GlcNAc, GalNAc or Gal in β1→3-linkage is catalyzed by β3GalTs [1,2]. These enzymes, along with β3GlcNAcTs and β3GalNAcTs, constitute the β3GlyT subfamily. Six β3GalTs viz., β3GalT I, II, IV, V, VI and VII, have been experimentally characterized so far from humans [3–6]. These six members use the same donor substrate but differ in their acceptor substrate specificity (Table 1).

β3GalTs are type II transmembrane proteins with a short N-terminal cytoplasmic tail, an α-helical transmembrane domain, a stem or neck region followed by the C-terminal catalytic domain [6]. β3GalTs share seven conserved sequence motifs (Fig. 1) [2]. The roles of some of the residues in these motifs have been investigated by site-specific mutagenesis and kinetic analyses of the mutants (Table 2) [7]. In general, knowledge of the 3-D structure (a) facilitates the rationalization of site-specific mutation data; (b) help in understanding the origin of the donor and acceptor substrate specificities; and (c) facilitate the identification of residues, which when mutated will display altered donor and/or acceptor substrate specificities. However, the 3-D structures of none of the eukaryotic β3GalTs are known to date. In view of this, the 3-D structures of the catalytic

Abbreviations: α3GalT, α1→3-galactosyltransferase; β3GalT, β1→3-galactosyltransferase; β3GalNAcT, β1→3-*N*-acetylgalactosaminyltransferase; β3GlcAT, β1→3-glucuronyltransferase; β3GlcNAcT, β1→3-*N*-acetylglucosaminyltransferase; β3GlyT, β1→3-glycosyltransferase; α4HexNAcT, α1→4-*N*-acetylhexosaminyltransferase; Cer, ceramide; Gal, galactose; GalNAc, *N*-acetylgalactosamine; GalT, galactosyltransferase; Glc, glucose; GlcNAc, *N*-acetylglucosamine; GlyT, glycosyltransferase; GM1, Galβ3GalNAcβ4(Neu5Acα2-3)Galβ4Glcβ1-Cer; Gal-Gb_4, Galβ3GalNAcβ3Galα4Galβ4Glcβ1-Cer; α2ManT, α1→2-mannosyltransferase; α3ManT, α1→3-mannosyltransferase; NeuAc, *N*-acetylneuraminic acid; SDP, specificity-determining positions; SDPpred, specificity-determining position prediction; SpsA, spore coat polysaccharide biosynthesis protein SpsA from *Bacillus subtilis*; TMD, transmembrane domain; UDP, uridinediphosphate; Xyl, xylose; XylT, xylosyltransferase

* Corresponding author. Tel.: +91 22 2576 7778; fax: +91 22 2572 3480.
E-mail address: balaji@iitb.ac.in (P.V. Balaji).

1093-3263/$ – see front matter © 2006 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmgm.2006.12.003

Table 1
Acceptor substrate specificities of β3GalTs

Enzyme	Preferred acceptor substrate[a]	Reference
β3GalT I	GlcNAc-β1→	[6,67]
β3GalT II	GlcNAc-β1→/egg ovalbumin GlcNAcβ1-3Galβ1-4Glcβ1-Cer (Lc3)	[6,67,68]
β3GalT IV	GalNAcβ1-4(Neu5Acα2-3)Galβ1-4Glcβ1-Cer (GM2) GalNAcβ1-4Galβ1-4Glcβ1-Cer (AsialoGM2) GalNAcβ1-4(NeuAcα2-8NeuAcα2-3)Galβ1-4Glcβ1-Cer(GD2)	[6,69]
β3GalT V	GalNAcβ1-3Galα1-4Galβ1-4Glcβ1-Cer (Gb4) O-linked *core3* GlcNAcβ13GalNAc Lewis A and sialyl-Lewis A antigen in gastrointestinal and pancreatic tumors	[4,70,71]
β3GalT VI	Galβ1-4Xyl (GAG core)	[3]
β3GalT VII	GalNAc-O-Ser/Thr	[5]

[a] The acceptor saccharide moiety is underlined.

domains of human β3GalT I, II, IV, V and VI and β3GalNAcT I have been modeled using fold-recognition and comparative modeling methods.

2. Methods

2.1. Servers

The NCBI server was used for accessing biomedical literature and protein sequences. Secondary structures were predicted using the GeneSilico meta-server [8] which uses JPRED [9], JUFO [10], PROF [11], PROFsec [12] and PSIPRED [13] servers. TMHMM was used to predict transmembrane domains [14]. The *needle* module of the EMBOSS 3.0 [15] server was used for pair-wise sequence alignment. 123D [16], 3D-PSSM [17], GeneSilico meta-server, FFAS03 [18], INUB [19,20] and SP^3 [21] servers were used for fold-recognition. All the servers were used with default values for the parameters, except where mentioned otherwise. *SDPpred* was used for predicting the specificity determining residues [22]. WebLogo was used for generating sequence logos [23]. The VAST server at NCBI and the FSSP server at EMBL were used to determine 3-D structural similarities.

2.2. Software and hardware

BioEdit [24] was used for display and manipulation of sequences. RasMol [25] and PyMol [26] were used for visualization and rendering of the 3-D structures of proteins. TopDraw [27] was used to generate the topology diagram. Modeller8v1, a homology modeling software, was used for modeling the 3-D structures [28,29]. The models were evaluated using ProsaII 3.0 [30]. Side chain optimization of model was carried out using SCWRL3.0 [31]. Multiple sequence alignments were generated locally using ClustalW [32], TCOFFEE [33] and MUSCLE [34] algorithms. All the softwares were run on an Intel Pentium IV desktop personal

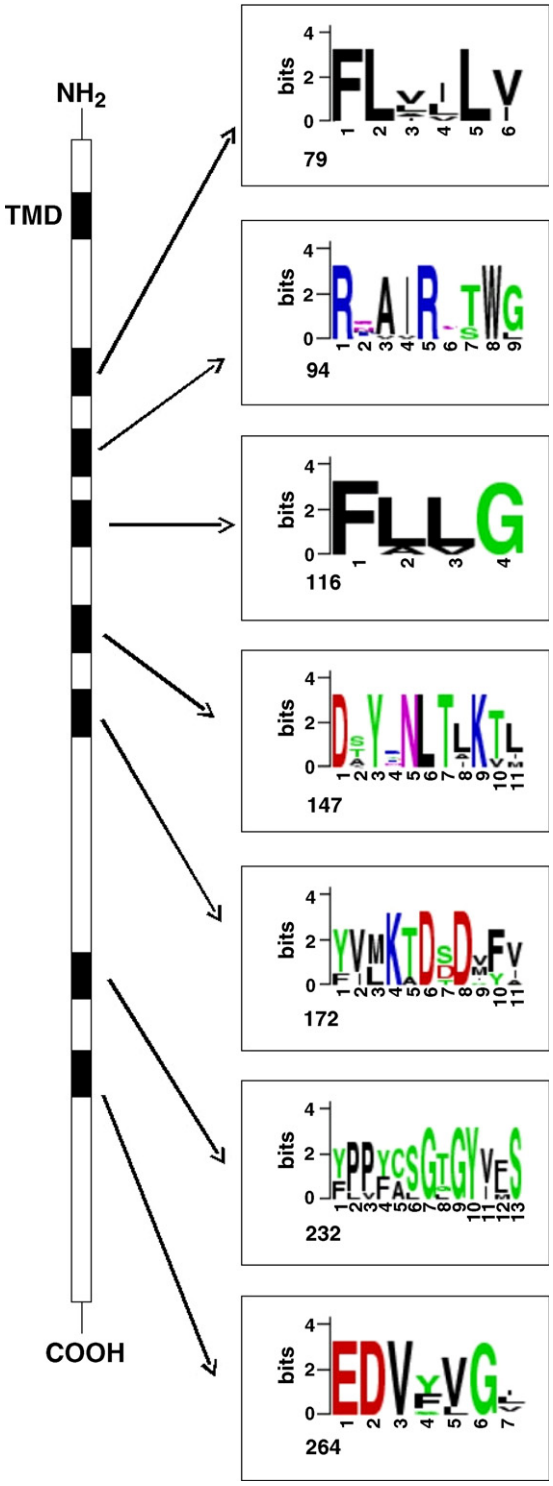

Fig. 1. Sequence logos of the motifs conserved in β3GalT family [2] and their location (not to scale) in the primary sequence. TMD denotes the transmembrane domain. The logos were generated by multiply aligning the sequences of the experimentally characterized β3GalT I, II, IV, V, VI and β3GalNAcT I (Table 3). Within each logo, (a) the numbers along the abscissa indicate the position of residues within the conserved region, (b) the number at the bottom left of each logo indicates the residue number of the first residue in human β3GalT I and (c) the ordinates, in units of bits, are indicative of the information content at each position [64,65]. Throughout the manuscript, the human β3GalT I numbering has been used for all the motifs.

Table 2
Summary of results of mutation studies performed by Malissard et al. [7] on murine β3GalT I

No.	Mutation(s)[a]	Effect	Role assigned	Basis
1 2	F116A-L117A-L118A-G119A I97A-R98A	Activity abolished	Provide β1 → 3-linkage specificity	Residues are conserved in β3GlcNAcTs also
3	D177A-D179A-F181A	Activity abolished	Binding UDP-Gal and Mn^{2+}	(1) Mutation in other GlyTs abolished activity (2) 3D structures of other GlyTs
4	P233A-P234A	Reduced activity	Maintaining the 3-D structure	K_ms for both the donor and acceptor substrates remain unchanged
5 6	C236A C271A			
7 8 9	W101A W162A E264A	Reduced activity	Involved in UDP-Gal binding	Increase in K_m for the donor substrate, UDP-Gal
10 11	E264A W315A	Reduced activity	Involved in acceptor substrate binding	Increase in K_m for the acceptor substrate GlcNAc-β-pNP
12 13 14 15	C73A C167A C295A C326A	Activity abolished	Maintaining the 3-D structure through –S–S– bonds	Flanking residues are not conserved

[a] The primary structures of mouse and human β3GalT Is are identical; hence, the residue numbering is same in human β3GalT I also.

computer. Default values were used for all the parameters, unless specified otherwise.

2.3. Strategy

Even though the objective of the present study is to model the 3-D structures of human β3GalTs, the sequences of other related proteins were also considered for analysis. As mentioned earlier, six β3GalTs have been characterized experimentally so far (Table 1). Of these, the primary structure of β3GalT VII is different from those of others. The protein coding sequence of β3GalT VII is in three exons [5] whereas those of β3GalT I, II, IV and VI are within a single exon [3,6]. Even though β3GalT V gene has four exons, the coding region is found within a single exon [4]. β3GalNAcT I [35] was initially identified as β3GalT III [6]; its sequence is similar to that of β3GalT I, II, IV, V and VI and is also coded by a single exon. In view of these observations, analyses have been performed considering the sequences of β3GalT I, II, IV, V and VI, and β3GalNAcT I (Table 3).

Table 3
Accession numbers and lengths of β3GlyTs considered for the modeling study

Protein	Organism	Accession number	Length
β3GalT I	Human	Q9Y5Z6	326
β3GalT II	Human	O43825	422
β3GalT IV	Human	O96024	378
β3GalT V	Human	Q9Y2C3	310
β3GalT VI	Human	Q96L58	329
β3GalNAcT I	Human	O75752	331

2.4. Datasets

Two datasets were created—*Dataset 1*: *experimentally characterized β3GlyTs* consisting mainly of β3GalTs were identified from literature and their amino acid sequences were retrieved from either UniProt [36] or NCBI protein database. This consists of, besides the enzymes listed in Table 3, mouse β3GalT-I (accession no. AAC53523), -II (AAC53524), -IV (Q9Z0F0), -V (AAF86241), and -VI (Q91Z92), mouse β3GalNAcT I (AAC53525) and rat β3GalT IV (BAA32045). *Dataset 2*: *putative β3GalTs* were identified by querying each protein from Dataset 1 (as above) against the *nr* database using *blastp* [37]. In addition, the human β3GalT sequences were queried against the transcripts generated by ENSEMBL project [38] using *tblastn* [37] locally. The protein sequences of the hits obtained from transcript database were either collected from the corresponding translation given in ENSEMBL or were translated using GENSCAN [39]. The accession numbers, lengths and names of the organisms of the protein sequences are in Table S1.

2.5. Secondary structure prediction

The secondary structures of each of the six target sequences (Table 3) were predicted using the GeneSilico meta-server [8]. Each residue in a protein was assigned one of the three states helix, strand or coil. The consensus secondary structure was derived by comparing the predictions of the different servers. If the secondary structure predicted by the different servers for a residue is not same, then the secondary structure predicted by at least four of the five servers was

taken as the consensus state; in other cases, it was considered as uncertain.

2.6. Fold-recognition and target-template sequence alignment

The most common and highest scoring template suggested by different fold-recognition servers was used as the template for generating the 3-D models of β3GalTs. The initial target-template alignments were derived by taking into consideration the alignments suggested by fold-recognition servers and predicted secondary structure. 3-D models were generated using these alignments and the sequences were re-aligned in an iterative fashion and manually edited based on the ProsaII *Z*-scores along with pair energy score as a function of residue number of generated models and Verify3D [40,41]. The quality of the models was further verified using the protein quality checking tools Anolea [42,43], ProQRes [44], Prove [45] and MetaMQAPII. The Anolea, Prove and Prosa tools were accessed through the Colorado 3D server [46].

2.7. UDP-Gal binding site

The position-orientations of UDP and Mn^{2+} were deduced using the crystal structure of SpsA bound to UDP (1QGQ). However, the coordinates for the saccharide moiety of the donor substrate are not available in this structure. Hence, the coordinates for the atoms of galactose were built based on the crystal structure of the catalytic domain of bovine α3GalT bound to UDP-Gal (1G93) [47]; torsion angles O5(Gal)-C1(Gal)-O-P and C1(Gal)-O-P-O around the diphospho-sugar linkage were set to 93 and −174, respectively. Side chain optimization of the entire protein in the presence of UDP-Gal and Mn^{2+} was carried out using SCWRL3.0 [31].

2.8. Prediction of substrate specificity determining positions

Residues that are subfamily specific were predicted using SDPpred [22]. This tool assumes that such subfamily specific residues correspond to those positions of the multiple sequence alignment where the distribution of residues is closely associated with grouping of proteins by subfamily type. SDPpred searches for positions that are well conserved within specificity groups but differ between them. The input for SDPpred is a multiple sequence alignment of β3GalT and β3GalNAcT I sequences (Datasets 1 and 2, as above) divided in to six groups each containing one subfamily. Alignments generated by three different algorithms were used to prevent erroneous prediction of SDPs due to errors in alignment. Using each multiple sequence alignment algorithm, members within the group were first aligned and subsequently, different groups were aligned. The SDPs identical at both the sequence and group level were considered.

3. Results and discussion

3.1. Comparison of the predicted secondary structures of β3GalTs

The predicted secondary structure for each of the human β3GalTs is given in Fig. 2. Residues constituting the putative N-terminal cytoplasmic tail and the stem region (see legend to Fig. 3) are in random coils. The catalytic domain, presumed to begin from the first conserved motif viz., ^{79}FLxxLx (Fig. 1), has a mixture of α-helices and β-strands. The 24-residue insertion in human β3GalT IV (residues 197–220) is predicted to be a random coil. β3GalT II, IV and VI have 15–25 residue long extensions at the C-terminus relative to other β3GalTs; this region is α-helical in β3GalT IV whereas it is random coil in β3GalT II and VI (Fig. 2).

Secondary structure prediction showed that β3GalTs belong to the α/β-class, as defined in the SCOP database [48]. GlyTs, whose 3-D structures have been determined to date, also belong to the α/β-class. Within this class, GlyTs constitute three-fold types: nucleotide-diphospho-sugar transferases (GT-A fold), UDP-glycosyltransferase/glycogen phosphorylase (GT-B fold) and α2,3/8-sialyltransferase CstII folds. Profile-HMM and clustering analyses showed that family 31 GlyTs in the CAZy database are part of the GT-A fold family [49].

3.2. Template identification by fold-recognition methods

The sequence of each target protein (Table 3) was submitted to fold-recognition servers. Spore coat polysaccharide biosynthesis protein SpsA from *Bacillus subtilis* emerged as the most probable template (Table S2). SpsA has the *nucleotide-diphospho-sugar transferases* (GT-A) fold. Some of the other top two hits suggested by the fold-recognition servers also have the GT-A fold; these are *N. meningitides* GalT LgtC (1GA8), mouse UDP-GalNAc: polypeptide αGalNAcT I (1XHB), the catalytic domain of bovine β4GalT I (2FYD) and human GalNAcT-2 (2FFU). In the case of β3GalT IV, the SP^3, 123D and INUB servers identify SpsA as the template if the sequence is submitted without the 24 residue long insertion that is unique to this protein (Fig. 3). Next to SpsA, mouse UDP-GalNAc:polypeptide αGalNAcT I (1XHB) is the most common template that is identified by the fold-recognition servers. The 3-D structure of this protein is similar to that of SpsA: the root mean square deviation between the two structures is 3.4 Å (*e*-value 10^{-16}; VAST server) and 3.3 Å (*Z*-score 16; FSSP server). This suggests that using UDP-GalNAc:polypeptide αGalNAcT I as the template is likely to yield the same models as those generated using SpsA.

3.3. Target-template sequence alignment and modeling the 3-D structures

The β3GalT I-SpsA alignments suggested by fold-recognition servers are identical to each other for the region from E4 to E5; the alignments are similar to each other for other residues in the region from E1 to H6 (figure showing

this comparison is not shown). The target-template secondary structures also match in this region except that the regions of target proteins which align with E7′, E7″ and H2′ of SpsA are predicted as coils (Fig. 3). The region from E1 to H6 encompasses the two structurally conserved regions of the GT-A fold in SpsA (Fig. 4) [50] and includes all the residues that interact with the donor sugar [51]. The target-template alignments generated by different servers for the region following H6 do not agree with each other; this region has one strand and one helix in all the six target proteins (except β3GalT IV which has one strand and two helices) and the template SpsA (Fig. 2).

3-D models of β3GalT I were generated from the initial alignment obtained from the fold-recognition servers. The sequences were iteratively re-aligned and manually edited based on the ProsaII and Verify3D evaluation of the generated models. The iteratively optimized β3GalT I-SpsA sequence alignment was then used to align the sequences of other target proteins (Table 3) using the information from (1) pair-wise sequence alignments of β3GalT I and target sequence, (2) alignment suggested by fold-recognition servers for the target sequence and SpsA and (3) multiple sequence alignment of the six target sequences generated by three different servers. The alignments derived in this fashion (Fig. 3) were used for

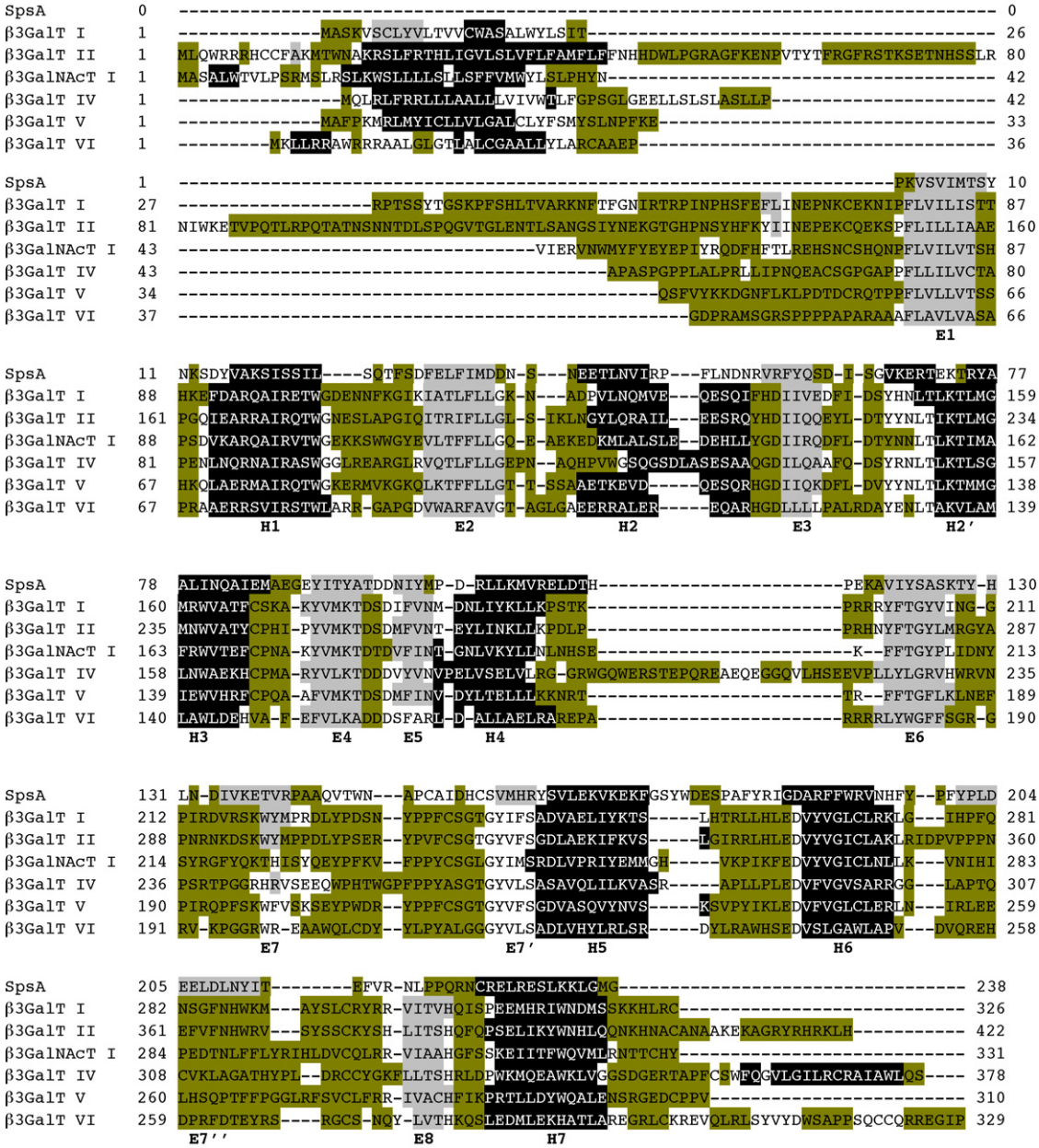

Fig. 2. Consensus secondary structures derived by comparing the predictions of five servers for the target proteins (Table 3). Secondary structure predicted by at least four of the five servers was taken as the consensus secondary structure state; else it was treated as uncertain. The residues with helical, extended and coiled conformation are displayed with black, gray and dark yellow background, respectively. Nomenclature used for identifying the secondary structural elements (H1–H7 and E1–E8) is also shown.

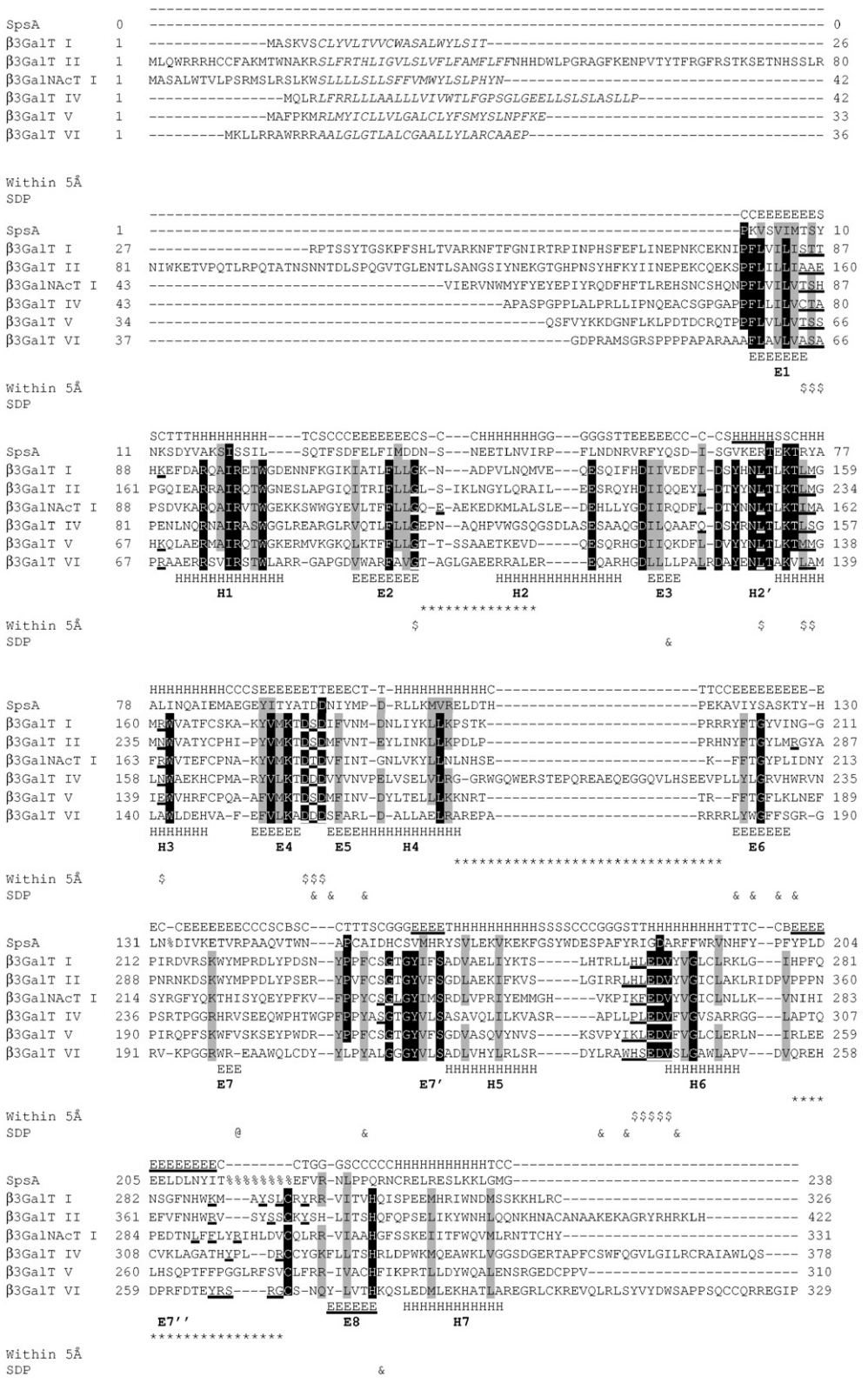

Fig. 3. Multiple sequence alignment of target (β3GalT members) and template (SpsA; PDB id 1QG8) proteins. The numbering on the left of each sequence indicates the position of the first residue in that row. β3GalT I numbering is used to refer to the various alignment positions. A hyphen denotes a single residue gap. Residues in the predicted TMD are shown in italics. Residues preceding the TMD constitute the cytoplasmic domain. The stem/neck region is after the TMD but where the stem region ends and the catalytic domain begins is not yet known. Residues are shaded black/gray if they are completely conserved/have conservative replacements in at least six of the seven aligned sequences. The secondary structure elements (H, helix and E, extended) of SpsA are shown above the alignment and the common secondary structural elements of target sequences is shown below the alignment; the consensus secondary structure for the individual target sequences is in Fig. 2. Nomenclature used for identifying the secondary structural elements (H1–H7 and E1–E8) is also shown. Regions corresponding to E7′, E7″ and H2′ of SpsA (secondary structure state underlined) are predicted as coils in the target proteins by secondary structure prediction servers. In SpsA, the % symbol at position 133

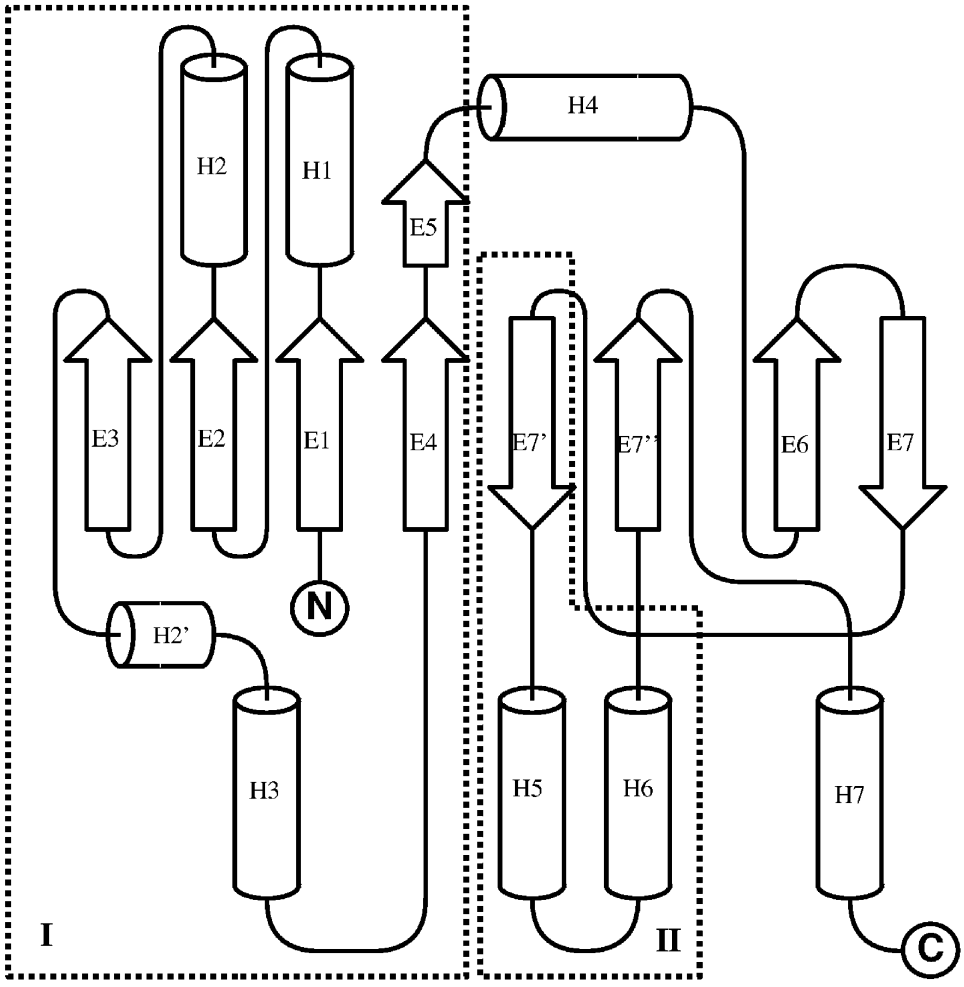

Fig. 4. Topology diagram of SpsA (1QG8) prepared using TopDraw. Arrows and cylinders represent strands and helices, respectively. The secondary structure elements that constitute the two structural motifs conserved in GT-fold family [50] are enclosed within dotted lines. All the secondary structure elements shown in this topology diagram are present in β3GalTs and β3GalNAcT I, even though E7′, E7″ and H2′ are predicted as coils by secondary structure prediction servers.

generating twenty models for each target protein and the model with the lowest ProsaII *Z*-score (Table 4) was selected for further analyses. The region preceding motif 1 (Fig. 1) was not modeled because of the absence of the corresponding region in SpsA whereas the short region after H7 in β3GalT II, IV and VI was modeled ab initio by Modeler8v1. The quality of the final models were assessed by Anolea, MetaMQAPII, ProsaII, Prove and Verify3D (Fig. S1).

3.4. Prediction of the functional residues

The GT-A fold has two conserved structural motifs which contain most of the substrate binding residues (Fig. 4) [50]: a Rossmann-fold-like αβα-sandwich terminated by the short strand E5 after the DxD motif (motif 5; Fig. 1) and a motif comprising one β-strand (E7′) and two α-helices (H5 and H6; Fig. 4). In SpsA, strands E1–E5 and helices H1–H3 cover the first motif whereas E7′, H5 and H6 encompass the second region. All these secondary structural elements are also present in β3GalT family members. ^{177}DSD preceding E5 corresponds to the DxD motif and D265 of motif 7 (at the beginning of H6) corresponds to the residue (D191) that is predicted to be the catalytic base in SpsA [51]. This aspartic acid residue is conserved in β3GlcNAcTs also (Fig. 5).

The modeled 3-D structures were superposed on the 3-D structure of UDP bound to SpsA to deduce the putative UDP-Gal binding site in these models. Residues that are within 5 Å from UDP-Gal were taken to constitute the putative binding site (Table 5). These residues are either part of or flank the motifs 1, 3, 4, 5 and 7 (Fig. 1). The uracil binding pocket is formed by G119 of motif 3 and the residues at end of E1 and beginning of H3. Residues in the loop between E1 and H1 form the ribose

denotes a three-residue stretch and the % symbols at positions 214–222 denote a 14-residue stretch; no electron density was observed for these residues in the crystal structure of SpsA. Asterisks indicate regions that are aligned differently by the three multiple sequence alignment servers ClustalW, TCOFFEE and MUSCLE. Residues marked "$" are within 5 Å from the donor substrate UDP-Gal in at least five of the six target proteins; residues interacting with the ligand are underlined in the respective sequences. Residues marked "&" indicate SDPs. The @ symbol at alignment position 222 indicates that E200 of β3GalT VI occupies this position as suggested by all the three multiple sequence alignment algorithms; however, the gap at this position is dictated by the ProsaII scores of the generated models.

Table 4
ProsaII Q-score[a] of the models

Model	ProsaII Q-score
SpsA (1QG8)	−1.84
β3GalT I	−1.34
β3GalT II	−1.05
β3GalNAcT I	−1.02
β3GalT IV	−0.88
β3GalT V	−0.96
β3GalT VI	−0.80

[a] The ProsaII Z-score is an approximate measure of the difference in free energy of a model under consideration and the mean free energy of the same sequence threaded through unrelated folds, expressed in units of SD. The free energies were calculated with statistical potentials of mean force for single residues and pairs of residues. A more negative Z-score indicates a better structural model. To overcome the fact that the ProsaII Z-score is dependant on the length of the amino acid sequence, the Z-score was normalized using the natural logarithm of the sequence length [72]. The resulting Q-score can be used to discriminate between good and bad 3-D protein models.

binding pocket. Residues of motifs 5 and 7 interact with the phosphate and Gal/GalNAc, respectively (Fig. 6).

3.5. Location of residues whose functional importance has been studied by site-specific mutations

The role of several conserved residues has been investigated by site-directed mutagenesis in mouse β3GalT I (Table 2) [7]. ^{116}FLLG of motif 3 (Fig. 1; mouse and human β3GalT I have the same primary structure) has been attributed to confer linkage specificity and the mutant in which all the four residues of the motif are mutated to alanine was enzymatically inactive. The modeled 3-D structures suggest that the loss of activity of this mutant is because of the G119 which constitutes the uracil binding pocket. In addition, residues L115, F116, L117 and L118 (in E2) form a hydrophobic core along with the residues of motifs 1 (E1) and 2 (H1). This hydrophobic core is conserved in all the members of the GT-A fold family (Fig. 7) suggesting

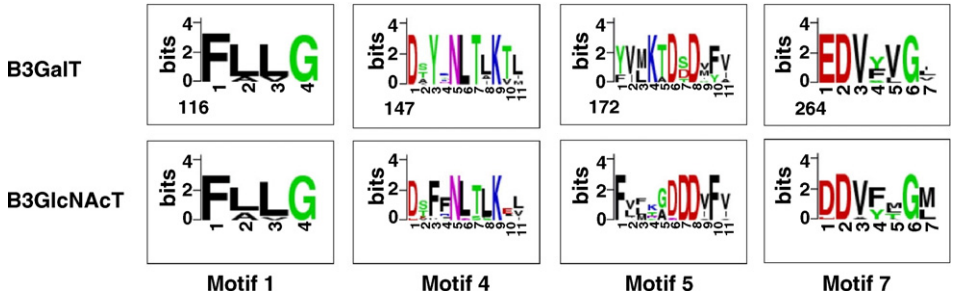

Fig. 5. Comparison of the sequence motifs that are conserved in β3GalTs and β3GlcNAcTs. Only the sequence motifs that contain residues that are within 5 Å from UDP-Gal in its putative binding mode are considered for comparison. The sequence logos were generated using the multiple sequence alignment of experimentally characterized β3GalTs and β3GlcNAcTs. The proteins used for generating the sequence logos for β3GlcNAcT family are as follows: human GlcNAcT II (BAB21530), human GlcNAcT III (BAB21531), human GlcNAcT IV (BAB21532), human GlcNAcT V (BAB40940), human GlcNAcT VI (BAB88882), human GlcNAcT VII (AAM61770), human GlcNAcT VIII (BAD86525), mouse GlcNAcT V (AAK31579), mouse GlcNAcT II (AAD09763), mouse GlcNAcT VII (AAM61769), rat GlcNAcT V (XP_344041) and *Drosophila melanogaster* Brainic (NP_476901).

Table 5
Residues constituting the putative UDP-Gal/UDP-GalNAc binding site

Protein	Residues	
	Uracil	Ribose
β3GalT I	^{85}STT, ^{152}L, ^{157}LM, ^{161}R, ^{294}L, ^{297}Y	^{85}STT, ^{89}K, ^{158}M, ^{161}R, ^{177}DSD
β3GalT II	^{158}AAE, ^{192}G[a], ^{221}L, ^{227}L, ^{232}LM, ^{236}N, ^{376}Y	^{158}AAE, ^{233}M, ^{252}DS
β3GalNAcT I	^{85}TSH, ^{119}G[a], ^{121}E, ^{149}L, ^{155}L, ^{160}IM, ^{164}R	^{85}TSH, ^{161}M, ^{164}R, ^{180}DTD, ^{294}R
β3GalT IV	^{78}CTA, ^{112}G[a], ^{144}Q, ^{150}L, ^{155}LS, ^{159}N	^{78}CTA, ^{156}S, ^{175}DDD, ^{290}E
β3GalT V	^{64}TSS, ^{98}G[a], ^{131}L, ^{136}MM, ^{140}E	^{64}TSS, ^{137}M, ^{156}DS
β3GalT VI	^{64}ASA, ^{97}G[a], ^{125}L, ^{132}L, ^{137}LA, ^{238}W, ^{271}C	^{64}ASA, ^{138}A, ^{156}DD, ^{238}W, ^{241}E

Protein	Residues	
	Phosphate	Galactose/GalNAc
β3GalT I	^{89}K, ^{179}D, ^{289}K, ^{292}Y	^{177}D, ^{238}G, ^{262}HLEDV, ^{289}K
β3GalT II	^{160}E, ^{254}D, ^{338}H, ^{368}R, ^{372}S, ^{374}C	^{233}M, ^{284}R, ^{314}G, ^{337}LHLEDV
β3GalNAcT I	^{87}H, ^{182}D, ^{291}F, ^{294}R	^{180}D, ^{239}SGL, ^{264}KFED,^{289}L, ^{291}F
β3GalT IV	^{176}DD, ^{290}E, ^{317}Y, ^{321}R	^{263}S, ^{288}PLEDV, ^{317}Y
β3GalT V	^{68}K, ^{158}D, ^{240}K	^{216}G, ^{239}IKLEDV
β3GalT VI	^{68}R, ^{157}DD, ^{239}H, ^{266}YRSRGC	^{238}WHSEDV, ^{266}Y, ^{268}S

Residues within 5 Å from UDP-Gal were taken to constitute the putative binding site; the mode of binding of UDP-Gal was deduced by superposing the modeled 3-D structure on the SpsA–UDP complex structure (1QGQ).

[a] This glycine residue, corresponding to G119 of β3GalT I, is part of motif 3 and is conserved in all the β3GalTs. The Cα atom of this residue is 5.2 Å from N3 of uracil in β3GalT I.

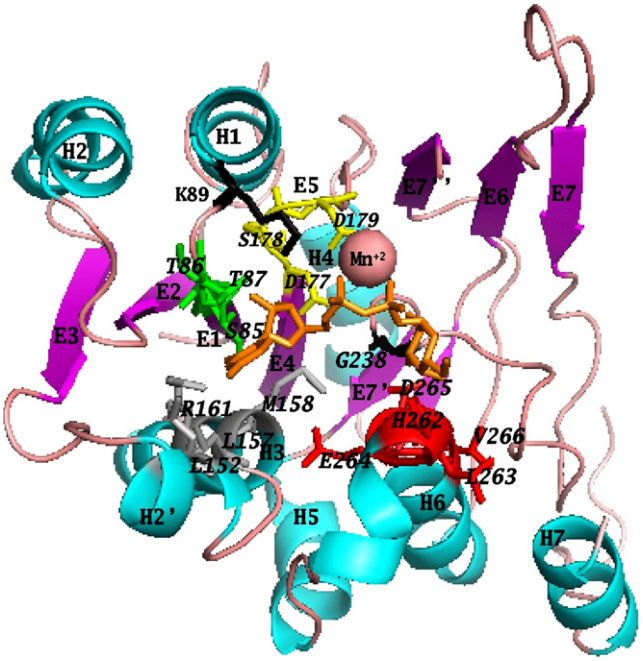

Fig. 6. Putative UDP-Gal binding site in the modeled structure of β3GalT I rendered using PyMol. Helices, sheets and coil regions are colored cyan, magenta and pink, respectively. UDP-Gal is shown as orange sticks. Region of the polypeptide chain containing the residues ^{289}K, ^{292}Y, ^{294}L and ^{297}Y has been omitted for clarity. Residues that lie within 5 Å from UDP-Gal are listed in Table 5.

their role in folding and/or maintenance of the 3-D structure. Based on these observations, it appears that ^{116}FLLG is unlikely to be responsible for linkage specificity.

The mutant D177A-D179A-F181A also loses activity. These residues are part of motif 5 and are within 5 Å of phosphate and galactose (Table 5). ^{177}DSD corresponds to the DxD motif of GlyTs of GT-A fold. Based on experimental studies on α3ManT [52], XylT [53], GM2 synthase [54] and fringe [55], the DxD motif has been implicated in binding phosphate group and the metal ion.

The E264A mutant retains about 20% of wild type enzyme activity and E264 was implicated in binding both acceptor and donor substrates from kinetic analyses (Table 2). This residue is part of motif 7 (Fig. 1). Although it is within 5 Å from galactose, side chain of E264 and galactose are on opposite sides of the backbone. The side chain carboxylate of E264 is 4.7 Å from the side chain amino group of K175 (motif 5; completely conserved in β3GalTs) in the modeled 3-D structure and these residues are surrounded by hydrophobic residues.

3.6. SDPpred and subfamily specific residues

Members of β3GalT family have different acceptor substrate specificity (Table 1) despite sharing primary and secondary structural similarities (Fig. 2). Residues that are likely to be responsible for this differential acceptor substrate specificity were predicted using SDPpred from a multiple sequence alignment of both experimentally characterized and putative β3GalTs and β3GalNAcT Is (Datasets 1 and 2; see Section 2). Fourteen specificity determining positions, spread over three stretches, were predicted (Fig. 3; Table 6). Of these, three are near (in primary structure) motif 5, four are near the conserved G205, three are near motif 7 and the rest are scattered (Fig. 3). The second stretch of specificity determining position is immediately after (in primary structure) the additional residues found only in β3GalT IV (Fig. 3). It is observed in the crystal structures of proteins belonging to the GT-A fold family that the residues involved in immediate acceptor sugar binding are found to be present in the helix containing the catalytic base (H6) and the region in between the two structurally conserved regions (Fig. 3) [50].

Table 6
SDPs and their role as inferred from the modeled 3-D structures

Category	Putative role	Residues at SDPs[a]
I	Acceptor binding site	I180, V207, N209, M222, F235
II	Role in maintaining the architecture of the acceptor binding site	Y202, T204
III	Not involved in binding the acceptor substrate	V142, S178, M184, T258, L261, Y267, Q305

[a] Residue numbering corresponds to β3GalT I.

Predicted SDPs were mapped on to the modeled 3-D structures (Fig. 8) and were classified, by visual inspection, into three groups depending upon their spatial location (Table 6): (1) those that are present at the putative binding site of acceptor (category I), (2) those that are in the vicinity of putative binding site and may thus be critical for maintaining the architecture of the acceptor binding site (category II) and (3) those that are away from the acceptor binding site (category III); the role of these SDPs might be in interacting with specific protein partners.

Wild type bovine β4GalT I exhibits β4GalNAcT activity that is ∼0.1% of β4GalT activity; however, the Y289L mutant can transfer both Gal and GalNAc as efficiently as the wild type transfers Gal [56]. The P234S mutant of α3GalT (GTB) utilizes UDP-GalNAc as the donor substrate and shows only marginal transferase activity with UDP-Gal [57]. Mutating Y299 to cysteine resulted in a 230-fold increase in the GalNAc/GlcNAc converting activity of UDP-galactose 4-epimerase [58]. It is clear from these observations that the changes conferring differential donor substrate specificity (viz., UDP-Gal vis-à-vis UDP-GalNAc) are subtle. In the present study, 3-D structures have been modeled for β3GalT I, II, IV, V and VI as well as for β3GalNAcT I. As mentioned earlier, the latter was initially identified as β3GalT III based on sequence similarity [6]. The subtle changes that confer the differential donor substrate specificity could not be predicted from the present 3-D structure modeling and SDPpred analysis. This is to be expected considering that the models are not of high resolution and the binding mode is putative, derived by simple superposition of the modeled structures on UDP-Gal bound SpsA.

3.7. Comparison of β3GalTs with β3GlcNAcTs

Family GT31 of the CAZy database [59] includes β3GalTs, β3GlcNAcTs and β3GalNAcTs. All these enzymes have

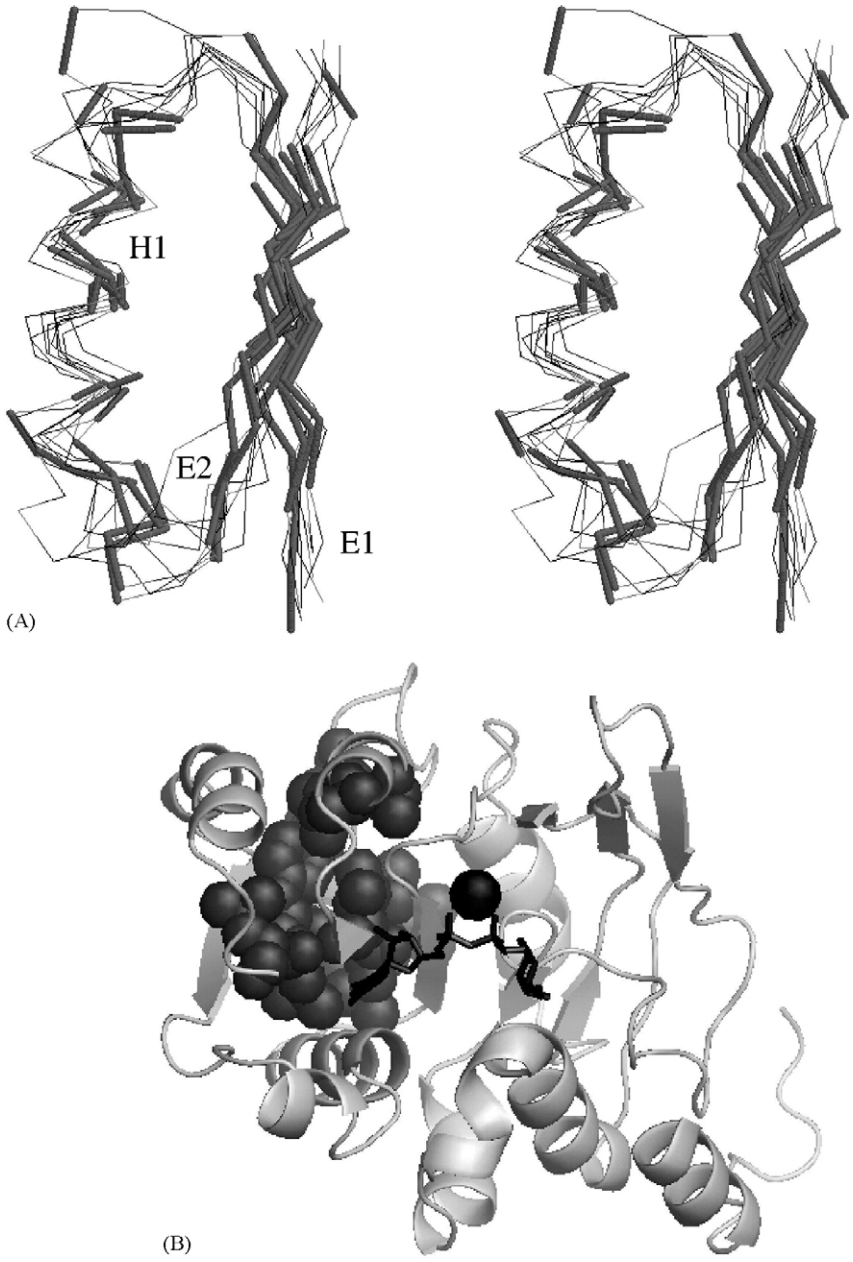

Fig. 7. (A) Stereo view of the conserved hydrophobic core constituted by the secondary structural elements E1, H1 and E2 in the members of the GT-A family. The nomenclature used to identify the secondary structural elements is same as that in Fig. 3. The non-polar amino acids are shown with thick lines and gray color while the polar amino acids are shown with thin lines and black color. The 3-D structure alignment was performed by MAMMOTH-Mult [66]. Names of the proteins are as follows: 1FO8, rabbit β2GlcNAcT; 1GA8, *N. meningitides* galactosyltransferase LgtC; 1LZJ, bovine blood group GlyT B (α3GalT); 1O7Q, bovine α3GalT; 1OMZ, mouse α4HexNAcT; 1PZT, W314A mutant of bovine β4GalT; 1QG8, *B. subtilis* spore coat polysaccharide biosynthesis protein SpsA; 1S4O, *S. cerevisiae* α2ManT; 1V84, human β3GlcAT; 1XHB, mouse UDP-GalNAc: polypeptide αGalNAcT I. Fold classification for 1XHB is from www.cermav.cnrs.fr/glyco3d/index.php; for others, it is from the SCOP database. The 3-D structure-based multiple sequence alignment is in Fig. S2. MAMMOTH-Mult gives only the Cα coordinates after superposition; hence, only these are shown in the figure. (B) The hydrophobic core formed by E1, H1 and E2 in β3GalT I is shown as black spheres. UDP-Gal is shown as black sticks and manganese is shown as black sphere.

similar conserved motifs, catalyze the formation of a β1→3-linkage but utilize different donor–acceptor sugar pairs. None of the β3GalTs can utilize UDP-GlcNAc as a donor sugar [35]. The modeled 3-D structures show that one or more residues from the motifs 3, 4, 5 and 7 (Fig. 1) are within 5 Å of putative UDP-Gal binding sites. ^{177}D of motif 5 is strictly conserved in β3GalTs but is only partially conserved in β3GlcNAcTs (Fig. 5). Incidentally, both the aspartic acid residues of the DxD motif are conserved in α3GalTs and β4GalTs [60–62]. The DxD motif is present as EDD with only the middle aspartic acid residue conserved in family 54 of the CAZy database that has β4GlcNAcTs. This motif has been suggested to contribute to differential donor substrate specificity [63]. Even ^{175}K of motif 5 is conserved in β3GalTs but not in β3GlcNAcTs (Fig. 5).

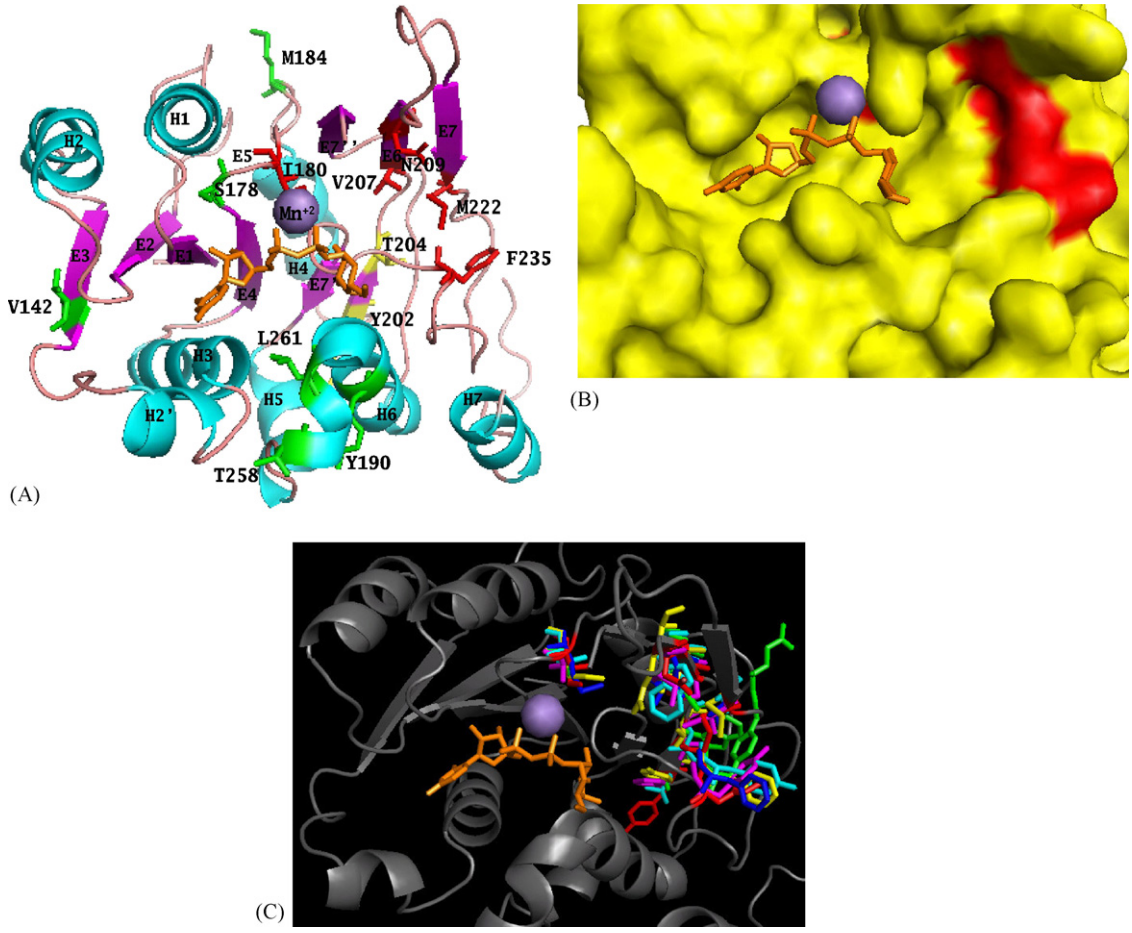

Fig. 8. Positions of SDPs on 3-D model on β3GalT I. UDP-Gal is shown as orange sticks and manganese is rendered as a purple sphere. (A) The categories I, II and III SDPs (Table 5) are shown in red, yellow and green, respectively. The protein is rendered in cartoon representation colored depending upon the secondary structure (cyan: helix; magenta: sheets). (B) Surface rendering of β3GalT I (yellow) and category I SDPs (red). Cartoon representation of β3GalT I and categories I and II SDPs corresponding to β3GalT I, II, IV, V, VI and β3GalNAcT I are shown as sticks (C) and colored red, yellow, cyan, blue, green and magenta, respectively.

Other differences that are evident between β3GalTs and β3GlcNAcTs are in motifs 4 and 7. Residue corresponding to Y149 of motif 4 is phenylalanine in β3GlcNAcTs; L152, present in the active site and strictly conserved among β3GalTs, is only partially conserved in β3GlcNAcTs (Fig. 5). Motif 7 viz., ^{264}EDVxxGx, contains the putative catalytic base in β3GalTs; only the residue corresponding to D265 is conserved in β3GlcNAcTs (Fig. 5). A partially conserved aspartic acid is found in place of E264. It is tempting to suggest that these differences are important for the differential donor substrate specificity.

In conclusion, 3-D structures have been modeled for human β3GalT I, II, IV, V, VI and β3GalNAcT I using fold-recognition and comparative modeling studies. The 3-D structure of SpsA is used as the template. Putative donor substrate binding residues have been predicted and those that are likely to be responsible for differential acceptor substrate specificity have also been predicted. Experimental studies will have to be performed to confirm these predictions. This will assist in the generation of mutants with altered donor/acceptor substrate specificities and these, in turn, are potentially useful for the in vitro manipulation of the glycan chains of glycolipids and glycoproteins and chemoenzymatic synthesis of carbohydrates.

Supplementary information

Supplementary data has two tables and two figures. Table S1 gives the accession numbers, source organisms and lengths of the putative β3GalTs and β3GalNAcTs that formed Dataset 2 (see Section 2). Table S2 contains the top two templates suggested by the various fold-recognition servers. Fig. S1 shows the plots of the scores generated by the protein quality checking servers for the various models. Fig. S2 shows the 3-D structure based sequence alignment of 10 proteins belonging to the GT-A fold family.

Acknowledgements

The authors thank M.S. Sujatha for useful discussion. RYP is grateful to the Indian Institute of Technology Bombay for teaching assistantship. The equipment used to carry out this work was bought from the grant (No. 37(1110)/02/EMR-II) from the Council for Scientific and Industrial Research, India to PVB.

Appendix A. URLs of sites accessed for this study

123D	123d.ncifcrf.gov/123D+.html
3D-PSSM	www.sbg.bio.ic.ac.uk/3dpssm/
BioEdit	www.mbio.ncsu.edu/BioEdit/bioedit.html
BLAST	www.ncbi.nlm.nih.gov/BLAST
CAZy	afmb.cnrs-mrs.fr/CAZY/
ClustalW	www.ebi.ac.uk/clustalw/ ftp://ftp.ebi.ac.uk/pub/software/unix/clustalw/
Colorado3D	http://asia.genesilico.pl/colorado3d/
EMBOSS	emboss.sourceforge.net/
ENSEMBL	www.ensembl.org/index.html
FFAS03	ffas.ljcrf.edu
FSSP	www2.ebi.ac.uk/dali/
GeneSilico	genesilico.pl/meta
GENSCAN	genes.mit.edu/GENSCAN.html
INUB	inub.cse.buffalo.edu
MAMMOTH-Mult	ub.cbm.uam.es/mammoth/mult/index2.php
MetaMQAPII	https://genesilico.pl/toolkit/perform?reset=t;service=metamqap
Modeller	salilab.org/modeller/
MUSCLE	www.drive5.com/muscle/
NCBI	www.ncbi.nlm.nih.gov
PDB	www.rcsb.org
ProQres	http://sbcweb.pdc.kth.se/cgi-bin/bjornw/ProQres.cgi
ProsaII	www.came.sbg.ac.at/Services/prosa.html
PyMol	pymol.sourceforge.net/
RasMol	www.bernstein-plus-sons.com/software/rasmol/
SCOP	scop.mrc-lmb.cam.ac.uk/scop
SCWRL3.0	dunbrack.fccc.edu/SCWRL3.php
SDPpred	math.genebee.msu.ru/∼psn/
SP^3	phyyz4.med.buffalo.edu/hzhou/sp3.html
TCOFFEE	igs-server.cnrs-mrs.fr/∼cnotred/Projects_home_page/t_coffee_home_page.html
TMHMM	www.cbs.dtu.dk/services/TMHMM/
TopDraw	stein.bioch.dundee.ac.uk/∼charlie/software/topdraw/
UniProt	www.ebi.uniprot.org/index.shtml
VAST	www.ncbi.nlm.nih.gov/Structure/VAST/vast.shtml
Verify3D	http://nihserver.mbi.ucla.edu/Verify_3D/
WebLogo	weblogo.berkeley.edu

Appendix B. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jmgm.2006.12.003.

References

[1] M. Amado, R. Almeida, T. Schwientek, H. Clausen, Identification and characterization of large galactosyltransferase gene families: galactosyltransferases for all functions, Biochim. Biophys. Acta 1473 (1999) 35–53.

[2] T. Hennet, The galactosyltransferase family, Cell. Mol. Life Sci. 59 (2002) 1081–1095.

[3] X. Bai, D. Zhou, J.R. Brown, B.E. Crawford, T. Hennet, J.D. Esko, Biosynthesis of the linkage region of glycosaminoglycans: cloning and activity of galactosyltransferase II, the sixth member of the β1,3-galactosyltransferase family (β3GalT6), J. Biol. Chem. 276 (2001) 48189–48195.

[4] S. Isshiki, A. Togayachi, T. Kudo, S. Nishihara, M. Watanabe, T. Kubota, M. Kitajima, N. Shiraishi, K. Sasaki, T. Andoh, H. Narimatsu, Cloning, expression, and characterization of a novel UDP-galactose:β-*N*-acetylglucosamine β1,3-galactosyltransferase (β3Gal-T5) responsible for synthesis of type 1 chain in colorectal and pancreatic epithelia and tumor cells derived there from, J. Biol. Chem. 274 (1999) 12499–12507.

[5] T. Ju, K. Brewer, A. D'Souza, R.D. Cummings, W.M. Canfield, Cloning and expression of human core 1 β1,3-galactosyltransferase, J. Biol. Chem. 277 (2002) 178–186.

[6] M. Amado, R. Almeida, F. Carneiro, S.B. Levery, E.H. Holmes, M. Nomoto, M.A. Hollingsworth, H. Hassan, T. Schwientek, P.A. Nielsen, E.P. Bennett, H. Clausen, A family of human β3-galactosyltransferases. Characterization of four members of a UDP-galactose:β-*N*-acetyl-glucosamine/β-*N*-acetylgalactosamine β-1,3-galactosyltransferase family, J. Biol. Chem. 273 (1998) 12770–12778.

[7] M. Malissard, A. Dinter, E.G. Berger, T. Hennet, Functional assignment of motifs conserved in β1,3-glycosyltransferases, Eur. J. Biochem. 269 (2002) 233–239.

[8] M.A. Kurowski, J.M. Bujnicki, GeneSilico protein structure prediction meta-server, Nucleic Acids Res. 31 (2003) 3305–3307.

[9] J.A. Cuff, G.J. Barton, Application of enhanced multiple sequence alignment profiles to improve protein secondary structure prediction, Proteins 40 (1999) 502–511.

[10] J. Meiler, M. Mueller, A. Zeidler, F. Schmaeschke, JUFO: Secondary Structure Prediction for Proteins, 2002, www.jens-meiler.de.

[11] M. Ouali, R.D. King, Cascaded multiple classifiers for secondary structure prediction, Protein Sci. 9 (2000) 1162–1176.

[12] B. Rost, PROF: predicting one-dimensional protein structure by profile based neural networks, 2000, unpublished.

[13] D.T. Jones, Protein secondary structure prediction based on position-specific scoring matrices, J. Mol. Biol. 292 (1999) 195–202.

[14] A. Krogh, B. Larsson, G. von Heijne, E.L.L. Sonnhammer, Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes, J. Mol. Biol. 305 (2001) 567–580.

[15] P. Rice, I. Longden, A. Bleasby, EMBOSS: The European Molecular Biology Open Software Suite, Trends Genet. 16 (2000) 276–277.

[16] N.N. Alexandrov, R. Nussinov, R.M. Zimmer, Fast protein fold recognition via sequence to structure alignment and contact capacity potentials, in: L. Hunter, T.E. Klein (Eds.), Pacific Symposium on Biocomputing'96, World Scientific Publishing Co., Singapore, 1995, pp. 53–72.

[17] L.A. Kelley, R.M. MacCallum, M.J.E. Sternberg, Enhanced genome annotation using structural profiles in the program 3D-PSSM, J. Mol. Biol. 299 (2000) 499–520.

[18] L. Jaroszewski, L. Rychlewski, Z. Li, W. Li, A. Godzik, FFAS03: a server for profile–profile sequence alignments, Nucleic Acids Res. 33 (2005) W284–W288.

[19] D. Fischer, 3D-SHOTGUN: a novel, cooperative, fold-recognition meta-predictor, Proteins 51 (2003) 434–441.

[20] D. Fischer, D. Eisenberg, Protein fold recognition using sequence-derived predictions, Protein Sci. 5 (1996) 947–955.

[21] H. Zhou, Y. Zhou, Fold recognition by combining sequence profiles derived from evolution and from depth-dependent structural alignment of fragments, Proteins 58 (2005) 321–328.

[22] O.V. Kalinina, P.S. Novichkov, A.A. Mironov, M.S. Gelfand, A.B. Rakhmaninova, SDPpred: a tool for prediction of amino acid residues that determine differences in functional specificity of homologous proteins, Nucleic Acids Res. 32 (2004) W424–W428.

[23] G.E. Crooks, G. Hon, J.M. Chandonia, S.E. Brenner, WebLogo: a sequence logo generator, Genome Res. 14 (2004) 1188–1190.

[24] T.A. Hall, BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT, Nucleic Acids Symp. Ser. 41 (1999) 95–98.

[25] R.A. Sayle, E.J. Milner-White, RASMOL: biomolecular graphics for all, Trends Biochem. Sci. 20 (1995) 374.

[26] W.L. DeLano, The PyMOL Molecular Graphics System, DeLano Scientific, San Carlos, CA, USA, 2002.

[27] C.S. Bond, TopDraw: a sketchpad for protein structure topology cartoons, Bioinformatics 19 (2003) 311–312.

[28] M.A. Marti-Renom, A. Stuart, A. Fiser, R. Sanchez, F. Melo, A. Sali, Comparative protein structure modeling of genes and genomes, Annu. Rev. Biophys. Biomol. Struct. 29 (2000) 291–325.

[29] A. Sali, T.L. Blundell, Comparative protein modeling by satisfaction of spatial restraints, J. Mol. Biol. 234 (1993) 779–815.

[30] M.J. Sippl, Recognition of errors in three-dimensional structures of proteins, Proteins 17 (1993) 355–362.
[31] A.A. Canutescu, A.A. Shelenkov, R.L. Dunbrack Jr., A graph theory algorithm for protein side-chain prediction, Protein Sci. 12 (2003) 2001–2014.
[32] J.D. Thompson, D.G. Higgins, T.J. Gibson, CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice, Nucleic Acids Res. 22 (1994) 4673–4680.
[33] C. Notredame, D. Higgins, J. Heringa, T-Coffee: a novel method for multiple sequence alignments, J. Mol. Biol. 302 (2000) 205–217.
[34] R.C. Edgar, MUSCLE: multiple sequence alignment with high accuracy and high throughput, Nucleic Acids Res. 32 (2004) 1792–1797.
[35] T. Okajima, Y. Nakamura, M. Uchikawa, D.B. Haslam, S.I. Numata, K. Furukawa, T. Urano, K. Furukawa, Expression cloning of human globoside synthase cDNAs. Identification of β3Gal-T3 as UDP-N-acetylgalactosamine:globotriaosylceramide β1,3-*N*-acetylgalactosaminyltransferase, J. Biol. Chem. 275 (2000) 40498–40503.
[36] A. Bairoch, R. Apweiler, C.H. Wu, W.C. Barker, B. Boeckmann, S. Ferro, E. Gasteiger, H. Huang, R. Lopez, M. Magrane, M.J. Martin, D.A. Natale, C. O'Donovan, N. Redaschi, L.S. Yeh, The universal protein resource (UniProt), Nucleic Acids Res. 33 (2005) D154–D159.
[37] S.F. Altschul, T.L. Madden, A.A. Schaeffer, J. Zhang, Z. Zhang, W. Miller, D.J. Lipman, Gapped BLAST and PSI-BLAST: a new generation of protein database search programs, Nucleic Acids Res. 25 (1997) 3389–3402.
[38] T. Hubbard, D. Andrews, M. Caccamo, G. Cameron, Y. Chen, M. Clamp, L. Clarke, G. Coates, T. Cox, F. Cunningham, V. Curwen, T. Cutts, T. Down, R. Durbin, X.M. Fernandez-Suarez, J. Gilbert, M. Hammond, J. Herrero, H. Hotz, K. Howe, V. Iyer, K. Jekosch, A. Kahari, A. Kasprzyk, D. Keefe, S. Keenan, F. Kokocinsci, D. London, I. Longden, G. McVicker, C. Melsopp, P. Meidl, S. Potter, G. Proctor, M. Rae, D. Rios, M. Schuster, S. Searle, J. Severin, G. Slater, D. Smedley, J. Smith, W. Spooner, A. Stabenau, J. Stalker, R. Storey, S. Trevanion, A. Ureta-Vidal, J. Vogel, S. White, C. Woodwark, E. Birney, Ensembl 2005, Nucleic Acids Res. 33 (2005) D447–D453.
[39] C. Burge, S. Karlin, Prediction of complete gene structures in human genomic DNA, J. Mol. Biol. 268 (1997) 78–94.
[40] J.U. Bowie, R. Luthy, D. Eisenberg, A method to identify protein sequences that fold into a known three-dimensional structure, Science 253 (1991) 164–170.
[41] R. Luthy, J.U. Bowie, D. Eisenberg, Assessment of protein models with three-dimensional profiles, Nature 356 (1992) 83–85.
[42] F. Melo, E. Feytmans, Assessing protein structures with a non-local atomic interaction energy, J. Mol. Biol. 277 (1998) 1141–1152.
[43] F. Melo, E. Feytmans, Novel knowledge-based mean force potential at atomic level, J. Mol. Biol. 267 (1997) 207–222.
[44] B. Wallner, A. Elofsson, Identification of correct regions in protein models using structural, alignment, and consensus information, Protein Sci. 15 (2006) 900–913.
[45] J. Pontius, J. Richelle, S.J. Wodak, Deviations from standard atomic volumes as a quality measure for protein crystal structures, J. Mol. Biol. 264 (1996) 121–136.
[46] M. Sasin, Bujnicki JM, COLORADO3D, a web server for the visual analysis of protein structures, Nucleic Acids Res. 32 (2004) W586–W589, Web Server issue.
[47] L.N. Gastinel, C. Bignon, A.K. Misra, O. Hindsgaul, J.H. Shaper, D.H. Joziasse, Bovine α1,3-galactosyltransferase catalytic domain structure and its relationship with ABO histo-blood group and glycosphingolipids glycosyltransferases, EMBO J. 20 (2001) 638–649.
[48] A.G. Murzin, S.E. Brenner, T. Hubbard, C. Chothia, SCOP: a structural classification of proteins database for the investigation of sequences and structures, J. Mol. Biol. 247 (1995) 536–540.
[49] N. Kikuchi, Y.D. Kwon, M. Gotoh, H. Narimatsu, Comparison of glycosyltransferase families using the profile hidden Markov model, Biochem. Biophys. Res. Commun. 310 (2003) 574–579.
[50] C. Breton, L. Snajdrova, C. Jeanneau, J. Koea, A. Imberty, Structures and mechanisms of glycosyltransferases superfamily, Glycobiology 16 (2006) 29R–37R.
[51] S.J. Charnock, G.J. Davies, Structure of the nucleotide-diphospho-sugar transferase, SpsA, from *Bacillus subtilis*, in native and nucleotide-complexed forms, Biochemistry 38 (1999) 6380–6385.
[52] C.A. Wiggins, S. Munro, Activity of the yeast MNN1 α-1,3-mannosyltransferase requires a motif conserved in many other families of glycosyltransferases, Proc. Natl. Acad. Sci. U.S.A. 95 (1998) 7945–7950.
[53] C. Gotting, S. Muller, M. Schottler, S. Schon, C. Prante, T. Brinkmann, J. Kuhn, K. Kleesiek, Analysis of the DXD motifs in human xylosyltransferase I required for enzyme activity, J. Biol. Chem. 279 (2004) 42566–42573.
[54] J. Li, D.M. Rancour, M.L. Allende, C.A. Worth, D.S. Darling, J.B. Gilbert, A.K. Menon, W.W. Young Jr., The DXD motif is required for GM2 synthase activity but is not critical for nucleotide binding, Glycobiology 11 (2001) 217–229.
[55] S. Munro, M. Freeman, The notch signalling regulator fringe acts in the Golgi apparatus and requires the glycosyltransferase signature motif DXD, Curr. Biol. 10 (2000) 813–820.
[56] B. Ramakrishnan, P.K. Qasba, Structure-based design of beta 1,4-galactosyltransferase I (beta 4Gal-T1) with equally efficient *N*-acetylgalactosaminyltransferase activity: point mutation broadens beta 4Gal-T1 donor specificity, J. Biol. Chem. 277 (2002) 20833–20839.
[57] S.L. Marcus, R. Polakowski, N.O. Seto, E. Leinala, S. Borisova, A. Blancher, F. Roubinet, S.V. Evans, M.M. Palcic, A single point mutation reverses the donor specificity of human blood group B-synthesizing galactosyltransferase, J. Biol. Chem. 278 (2003) 12403–12405.
[58] J.B. Thoden, J.M. Henderson, J.L. Fridovich-Keil, H.M. Holden, Structural analysis of the Y299C mutant of *Escherichia coli* UDP-galactose 4-epimerase. Teaching an old dog new tricks, J. Biol. Chem. 277 (2002) 27528–27534.
[59] P.M. Coutinho, E. Deleury, G.J. Davies, B. Henrissat, An evolving hierarchical family classification for glycosyltransferases, J. Mol. Biol. 328 (2003) 307–317.
[60] B. Ramakrishnan, P.V. Balaji, P.K. Qasba, Crystal structure of beta1,4-galactosyltransferase complex with UDP-Gal reveals an oligosaccharide acceptor binding site, J. Mol. Biol. 318 (2002) 491–502.
[61] R. Almeida, S.B. Levery, U. Mandel, H. Kresse, T. Schwientek, E.P. Bennett, H. Clausen, Cloning and expression of a proteoglycan UDP-galactose: β-xylose β1,4-galactosyltransferase I. A seventh member of the human β4-galactosyltransferase gene family, J. Biol. Chem. 274 (1999) 26165–26171.
[62] E. Boix, G.J. Swaminathan, Y. Zhang, R. Natesh, K. Brew, K.R. Acharya, Structure of UDP complex of UDP-galactose:β-galactoside-α-1,3-galactosyltransferase at 1. 53-A resolution reveals a conformational change in the catalytically important C terminus, J. Biol. Chem. 276 (2001) 48608–48614.
[63] U.M. Unligil, J.M. Rini, Glycosyltransferase structure and mechanism, Curr. Opin. Struct. Biol. 10 (2000) 510–517.
[64] T.D. Schneider, R.M. Stephens, Sequence logos: a new way to display consensus sequences, Nucleic Acids Res. 18 (1990) 6097–6100.
[65] D.W. Mount, Bioinformatics. Sequence and Genome Analysis, CBS Publishers and Distributors, New Delhi, 2003, pp. 196–198.
[66] D. Lupyan, A. Leo-Macias, A.R. Ortiz, A new progressive-iterative algorithm for multiple structure alignment, Bioinformatics 21 (2005) 3255–3263.
[67] T. Hennet, A. Dinter, P. Kuhnert, T.S. Mattu, P.M. Rudd, E.G. Berger, Genomic cloning and expression of three murine UDP-galactose: β-*N*-acetylglucosamine β1,3-galactosyltransferase genes, J. Biol. Chem. 273 (1998) 58–65.
[68] F. Kolbinger, M.B. Streiff, A.G. Katopodis, Cloning of a human UDP-galactose:2-acetamido-2-deoxy-D-glucose 3β-galactosyltransferase catalyzing the formation of type 1 chains, J. Biol. Chem. 273 (1998) 433–440.
[69] H. Miyazaki, S. Fukumoto, M. Okada, T. Hasegawa, K. Furukawa, Expression cloning of rat cDNA encoding UDP-galactose:GD2 β1,3-galactosyltransferase that determines the expression of GD1b/GM1/GA1, J. Biol. Chem. 272 (1997) 24794–24799.

[70] D. Zhou, T.R. Henion, F.B. Jungalwala, E.G. Berger, T. Hennet, The β1,3-galactosyltransferase β3GalT-V is a stage-specific embryonic antigen-3 (SSEA-3) synthase, J. Biol. Chem. 275 (2000) 22631–22634.

[71] D. Zhou, E.G. Berger, T. Hennet, Molecular cloning of a human UDP-galactose: GlcNAcβ1,3GalNAc β1,3galactosyltransferase gene encoding an O-linked core3-elongation enzyme, Eur. J. Biochem. 263 (1999) 571–576.

[72] R. Sanchez, A. Sali, Large-scale protein structure modeling of the *Saccharomyces cerevisiae* genome, Proc. Natl. Acad. Sci. U.S.A. 95 (1998) 13597–13602.