Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

The structure of human α-2,6-sialyltransferase reveals the binding mode of complex glycans

Bernd Kuhn,[a] Jörg Benz,[a] Michael Greif,[b] Alfred M. Engel,[b] Harald Sobek[b] and Markus G. Rudolph[a]*

[a]pRED Pharma Research and Early Development, Discovery Technologies, F. Hoffmann-La Roche AG, Grenzacher Strasse 124, 4070 Basel, Switzerland, and [b]Roche Diagnostics GmbH, Nonnenwald 2, 82377 Penzberg, Germany

Correspondence e-mail: markus.rudolph@roche.com

Received 25 April 2013
Accepted 3 June 2013

PDB References: ST6Gal-I, complex with cytidine and phosphate, 4js1; complex with CMP, 4js2

Human β-galactoside α-2,6-sialyltransferase I (ST6Gal-I) establishes the final glycosylation pattern of many glycoproteins by transferring a sialyl moiety to a terminal galactose. Complete sialylation of therapeutic immunoglobulins is essential for their anti-inflammatory activity and protein stability, but is difficult to achieve *in vitro* owing to the limited activity of ST6Gal-I towards some galactose acceptors. No structural information on ST6Gal-I that could help to improve the enzymatic properties of ST6Gal-I for biotechnological purposes is currently available. Here, the crystal structures of human ST6Gal-I in complex with the product cytidine 5′-monophosphate and in complex with cytidine and phosphate are described. These complexes allow the rationalization of the inhibitory activity of cytosine-based nucleotides. ST6Gal-I adopts a variant of the canonical glycosyltransferase A fold and differs from related sialyltransferases by several large insertions and deletions that determine its regiospecificity and substrate specificity. A large glycan from a symmetry mate localizes to the active site of ST6Gal-I in an orientation compatible with catalysis. The glycan binding mode can be generalized to any glycoprotein that is a substrate of ST6Gal-I. Comparison with a bacterial sialyltransferase in complex with a modified sialyl donor lends insight into the Michaelis complex. The results support an S_N2 mechanism with inversion of configuration at the sialyl residue and suggest substrate-assisted catalysis with a charge-relay mechanism that bears a conceptual similarity to serine proteases.

1. Introduction

Naturally occurring oligosaccharides (or glycans) bound to asparagine (N-linked) or serine (O-linked) residues of glycoproteins play important roles in protein folding, protein stability and enzyme activity, and also in medically important processes such as cell growth and development, cell–cell communication, cancer-cell metastasis, anticoagulation, the immune response and host–pathogen interactions (Moremen *et al.*, 2012). In eukaryotes, the biosynthesis of glycans takes place in the lumen of the endoplasmic reticulum and the Golgi apparatus. There, the structural diversity of glycans (possibly >7000 structures) is ensured by complex biosynthetic pathways regulated by around 200 glycosyltransferases (GTs) with different acceptor preferences and donor specificities (Moremen *et al.*, 2012). Glycans are synthesized from a set of ten lipid-activated or nucleotide-activated donor monosaccharides. In the case of N-linked glycans, synthesis is

usually initiated by *en bloc* addition of a separately synthesized lipid-linked oligosaccharide to an asparagine side chain, followed by further diversification of the growing glycan (Moremen *et al.*, 2012). The glycans contain one or more bifurcations and often terminate in galactose or sialic acid (Fig. 1*a*). Upon completion of glycosylation the glycoproteins are either secreted (*e.g.* immunoglobulins) or are integrated into the cell membrane, where they function in protein recognition and modulate protein activity at the extracellular surface of the cell.

In the glycan structure of glycoproteins a sialyl moiety is usually found at the terminal position of the oligosaccharide (Fig. 1*a*). The most common form of sialic acid, which is a collective term for N- or O-substituted derivatives of neuraminic acid, is the derivative *N*-acetylneuramic acid (Neu5Ac or NANA; Fig. 1). The exposed sialic acid can participate in biological recognition phenomena, for instance in virus–cell interactions with some influenza, coxsackie and polyoma viruses (reviewed in Neu *et al.*, 2011). Changes in glycosylation patterns, often sialylation, of cell-surface proteins are a common malignancy phenotype and these changes are associated with invasion and metastasis (reviewed in Dall'Olio & Chiricolo, 2001). Sialylated oligosaccharides are also implicated in inflammation. For instance, mice lacking the gene for β-galactoside α-2,6-sialyltransferase I (ST6Gal-I) exhibit severe immunosuppression (Hennet *et al.*, 1998). Finally, in a mouse model of rheumatoid arthritis it was shown that sialylation of a therapeutic IgG-Fc immunoglobulin was essential for its anti-inflammatory activity (Anthony *et al.*, 2008).

The biosynthesis of sialylconjugates is controlled by sialyltransferases (STs), a subset of GTs that currently includes at least 20 human genes. The sialic acid moiety from the activated donor cytidine 5′-monophospho-*N*-acetylneuraminic acid (CMP-Neu5Ac) is transferred to a glycan with galactose, *N*-acetylgalactosamine (GalNAc) or sialic acid as the acceptor

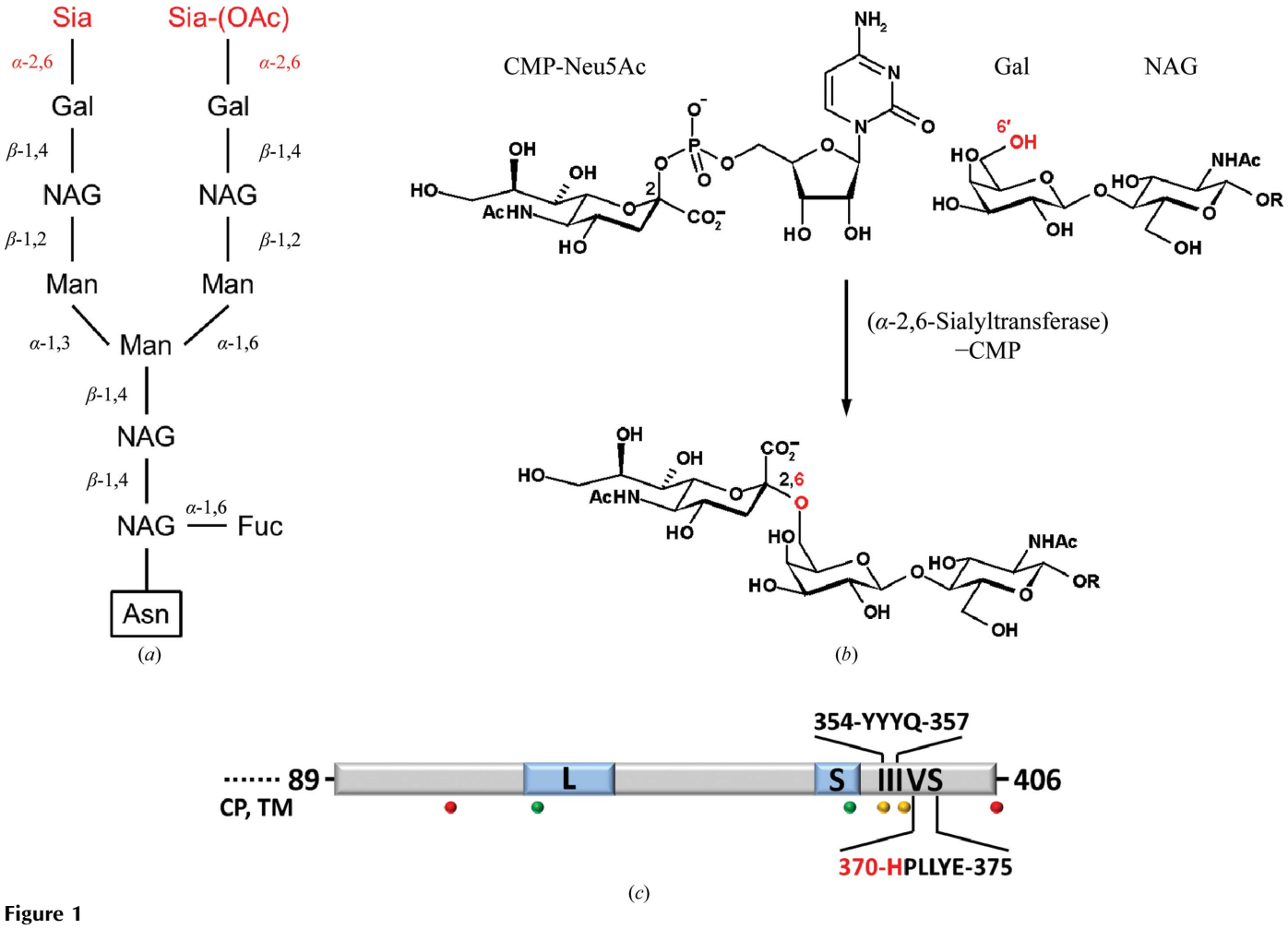

Figure 1
Sialylation of glycoproteins. (*a*) Major glycosylation tree in ST6Gal-I identified by mass spectrometry. The masses are 2350.8 and 2391.9 Da, differing by only an acetyl group at the distal sialyl residue (data not shown). This glycosylation pattern is ubiquitous in nature. (*b*) Scheme of the glycosylation reaction. The 6′-hydroxyl group (marked in red) of a terminal galactose residue acts as a nucleophile for the CMP-activated precursor *N*-acetylneuraminic acid (CMP-Neu5Ac). The reaction proceeds with inversion of configuration of the anomeric C atom. (*c*) Schematic drawing of the ST6Gal-I sequence organization. Residues before position 89 belong to the cytoplasmic part (CP) and transmembrane region (TM). The positions and relative sizes of the four conserved ST sequence motifs are shown. For motifs III and VS the sequence is given using a single-letter code. The catalytic His370 is marked in red. Three disulfide bonds between cysteines 142/406, 184/335 and 353/364 are marked as pairs of red, yellow and green dots, respectively.

(Fig. 1*b*). The transfer is regiospecific and is used to classify STs. Based on the position of the glycosyl acceptor that the sialic acid is transferred to, STs are classified as ST3, ST6 and ST8, forming an α-glycosidic bond between the C2 atom of sialic acid and the 3′-, 6′- or 8′-hydroxyl group of the acceptor, respectively.

The ST6 family consists of two subfamilies, ST6Gal-I and ST6Gal-II, that both catalyse the transfer of Neu5Ac to the 6′-hydroxyl group of a disaccharide containing a terminal galactose (Galβ1–4GlcNAc; Audry *et al.*, 2011; Weijers *et al.*, 2008). The subfamilies share only limited sequence identity (48.9% for the human enzymes) and differ in their detailed acceptor specificities (Takashima *et al.*, 2002). Human ST6Gal-I (EC 2.4.99.1; database entry P15907) belongs to the GT29 family of GTs comprising eukaryotic and viral STs (http://www.cazy.org). Similar to other vertebrate STs, human ST6Gal-I is localized in the membrane of the Golgi apparatus. The sequence exhibits four characteristic regions. The first nine N-terminal residues are cytoplasmic, followed by a single-pass transmembrane sequence that extends into a so-called stem region and a C-terminal catalytic domain facing the luminal side. The transmembrane sequence and stem region contain the Golgi apparatus localization signals (Fenteany & Colley, 2005). Several conserved sequences were identified in the catalytic domains of all eukaryotic STs (Fig. 1*c*) and are referred to as sialyl motifs L (large), S (small), III and VS (very small). The motifs are attributed to different functions: motifs L (Trp181–Gly224; human ST6Gal-I numbering) and III (Tyr354–Gln357) participate in the binding of the sugar donor CMP-Neu5Ac (Datta *et al.*, 2001), motif S (Pro321–Phe343) is involved in donor and acceptor substrate binding and motif VS (His370–Glu375) participates in the catalytic reaction (Datta, 2009). The minimal size of an active ST6Gal-I is somewhat controversial and has been mapped by N-terminal deletions (Legaigneur *et al.*, 2001) and sequence considerations (Datta, 2009). The first 80 residues of human ST6Gal-I including the cytoplasmic part, the transmembrane sequence and the stem region can be deleted without affecting activity. However, deletion of 100 residues renders the enzyme inactive, indicating that the region of residues 80–100 affects substrate binding or catalysis. Another study estimated the start of the catalytic domain of rat liver ST6Gal-I at residue position 63 (Datta, 2009). The catalytic domain of ST6Gal-I contains six cysteine residues, of which two were predicted to form a disulfide bond (Datta *et al.*, 2001).

The three-dimensional structures of nucleotide-sugar-dependent GTs are categorized into two folds termed GT-A and GT-B (Lairson *et al.*, 2008; Audry *et al.*, 2011; Breton *et al.*, 2012). Both categories contain Rossmann folds, a six-stranded parallel β-sheet linked to two pairs of α-helices. Whereas GT-B represents two flexibly linked Rossmann folds, two fused Rossmann folds characterize GT-A. From sequence comparisons ST6Gal-I was predicted to adopt the GT-A variant fold (Audry *et al.*, 2011), a heavily modified single Rossmann-like domain adopting a mixed β-sheet. Here, we describe the crystal structure of an active variant of human ST6Gal-I. The structure was determined in complex with the product cytidine 5′-monophosphate (CMP) and, fortuitously, in complex with cytidine and phosphate. ST6Gal-I indeed adopts the globular GT-A fold, which assigns the boundary of the catalytic domain to Phe93. In contrast to previous findings (Datta *et al.*, 2001), all six cysteines are involved in disulfide bonds. The structures lend insight into the conformational changes of ligands in the nucleotide-binding site and enable a model for the competitive ST6Gal-I inhibitors cytidine 5′-diphosphate (CDP) and cytidine 5′-triphosphate (CTP) to be proposed. Importantly, the structure visualizes a complete glycan at Asn149. In the crystal the glycan localizes to the active site of a symmetry mate, thus delivering valuable information on substrate binding and catalysis. The observed glycan binding mode may serve as a general framework for the binding of any glycoprotein to ST6Gal-I. Interesting mechanistic insight supporting an S_N2 mechanism with inversion of configuration at the sialyl residue is gained by comparison of this ST6Gal-I–substrate complex with other sialyltransferases.

2. Materials and methods

2.1. Cloning, protein production and purification

Based on the full-length sequence of human β-galactoside α-2,6-sialyltransferase I, a truncated version lacking the N-terminal 89 amino acids (Δ89ST6Gal-I) was constructed for transient eukaryotic expression in human embryonic kidney (HEK) cells. In order to ensure the correct processing of ST6Gal-I by the secretion machinery of the host cell line, the natural leader sequence and the N-terminal 89 codons of the ST6Gal-I coding region were replaced by the erythropoietin (Epo) signal sequence plus an Ala-Pro (AP) linker. Codon-optimized cDNA was synthesized for the Epo-AP-Δ89ST6Gal-I gene and cloned into the *Sal*I and *Bam*HI restriction sites of the pM1MT vector (Roche Applied Science). Thus, expression of the ST6Gal-I coding sequence was put under the control of a human cytomegalovirus immediate-early enhancer/promoter region followed by an intron A for regulated expression and a bovine growth hormone polyadenylation signal. A suspension-adapted HEK 293 cell line was used for transient gene expression. Cultures with $\sim 2 \times 10^6$ cells ml^{-1} were transfected with the pM1MT expression plasmid at a concentration of 0.5–1 mg per litre of cell culture complexed by the 293-Free (Merck) transfection reagent according to the manufacturer's guidelines. 3 h post-transfection, the histone deacetylase inhibitor valproic acid was added to 4 m*M* in order to boost protein production (Backliwal *et al.*, 2008). Each day, the culture was supplemented with 6%(*v*/*v*) of a soybean peptone hydrolysate-based feed. 7 d post-transfection, the culture supernatant was collected by centrifugation.

0.1 l culture supernatant was filtrated (0.2 µm) and the filtrate was dialyzed against 20 m*M* potassium phosphate pH 6.5 (buffer *A*). The dialysate was loaded onto an S-Sepharose fast-flow column (1.6 × 2 cm) equilibrated with buffer *A*. After washing with 100 ml buffer *A*, the enzyme was eluted

with a linear gradient of 10 ml buffer *A* and 10 ml of buffer *A* including 200 m*M* NaCl followed by a wash step using 48 ml buffer *A* including 200 m*M* NaCl. Fractions (4 ml) were analyzed by SDS–PAGE, pooled and dialyzed against 50 m*M* MES–NaOH pH 6.0 (buffer *B*). The dialysate was applied onto a Heparin Sepharose fast-flow column (0.5 × 5 cm) equilibrated with buffer *B* and eluted using buffer *B* including 200 m*M* NaCl. Fractions (1 ml) containing the enzyme were pooled and dialyzed against buffer *B*. Protein purity was tested by analytical size-exclusion chromatography on a Superdex 75 column (10/30) equilibrated with buffer *B* including 500 m*M* NaCl. The Epo signal sequence was removed by signal peptidase during protein production.

Sialyltransferase activity was measured using asialofetuin (desialylated fetuin; Roche Applied Science) as the acceptor substrate and CMP-9F-Neu5Ac (CMP-9-fluoresceinyl-NeuAc) as the donor substrate (Brossmer & Gross, 1994), *i.e.* by determining the transfer of 9F-Neu5Ac to asialofetuin. The reaction mixture (34 m*M* MES–NaOH pH 6.0, 0.034% Triton X-100, 0.07% BSA) contained 0.025 µg purified enzyme sample, 20 µl 20 mg ml^{-1} asialofetuin and 8 µl 1.0 mg ml^{-1} CMP-9F-Neu5Ac in a total volume of 50 µl. The sample was incubated at 310 K for 15 min and the reaction was stopped by the addition of 10 µl of a 10 m*M* solution of the inhibitor CTP. The reaction mixture was applied onto a PD10 desalting column equilibrated with 0.1 *M* Tris–HCl pH 8.5. Fetuin was eluted in 0.5 ml fractions from the column using the equilibration buffer. The amount of 9F-fetuin formed was determined by fluorescence spectrophotometry (λ_{ex} = 490 nm; λ_{em} = 520 nm). The specific sialyltransferase activity was determined to be 3.34 nmol µg^{-1} min^{-1}.

Table 1
Data-collection and phasing statistics.

Unless noted otherwise, values in parentheses are for the highest resolution shell.

Data set	Cytidine/phosphate	Iodide	CMP
Wavelength (Å)	1.00	1.77	1.00
Resolution range (Å)	46.4–2.09 (2.16–2.09)	39.1–2.50 (2.59–2.50)	45.6–2.30 (2.38–2.30)
100% criterion† (Å)	2.11	2.52	2.31
Rotation range (°)	180	360	180
Rotation increment (°)	0.25	0.25	0.25
Mosaicity (°)	0.17	0.31	0.21
Space group	$P6_1$	$P6_1$	$P6_1$
Unit-cell parameters (Å)	a = 65.3, c = 162.2	a = 65.8, c = 162.0	a = 64.0, c = 160.6
Unique reflections	22378 (2292)	13321 (1335)	16459 (1632)
Multiplicity	9.9 (10.2)	18.15 (20.0)	10.3 (9.9)
Completeness (%)	96.8 (99.4)	98.0 (98.9)	99.3 (98.7)
R_{meas}‡	0.076 (2.3)	0.149 (3.4)	0.085 (2.5)
$R_{p.i.m.}$‡	0.024 (0.72)	0.034 (0.76)	0.026 (0.79)
$CC_{1/2}$‡	1.00 (0.65)	1.00 (0.56)	1.00 (0.56)
Average $I/\sigma(I)$	13.5 (1.3)	11.2 (1.0)	11.8 (1.0)
Anisotropy, Wilson B (Å^2)	55.2/55.2/32.1	69.7/69.7/41.1	72.7/72.7/55.2
$\langle\lvert E^2 - 1\rvert\rangle$‡§	0.721 (0.736/0.541)	0.761	0.716
Mean $\langle L^2\rangle$‡§	0.312 (0.333/0.200)	0.328	0.296
No. of iodide sites		2	
FOM acentric/centric		0.094/0.119	
PP_{iso}¶		0.582/0.731	
PP_{iso}/drop < 1†† (Å)		4.7/5.8	
PP_{aniso}/drop < 1†† (Å)		0.511/4.1	

† The 100% criterion was calculated using *SFTOOLS* (Winn *et al.*, 2011) and represents the resolution in Å of a 100% completed hypothetical data set with the same number of reflections as the measured data. ‡ *E*- and *L*-values (acentric reflections) and *R* factors were calculated using *PHENIX* (Zwart *et al.*, 2008). *R* values and $CC_{1/2}$ are as defined in Diederichs & Karplus (1997) and Karplus & Diederichs (2012), respectively. § Values in parentheses are the expected values for untwinned and perfectly twinned data, respectively. ¶ PP, phasing power; drop < 1, the resolution at which the PP falls below 1.

2.2. Crystallization and data collection

Δ89ST6Gal-I was concentrated to 15 mg ml^{-1} using a centrifugal filter (Amicon Ultracel-10K), supplemented with 10 m*M* CMP (to stabilize the donor site) and 10 m*M* lactose (to stabilize the acceptor site) and deglycosylated for 3 h at 294 K by the addition of peptide *N*-glycosidase F (1:30 molar ratio), which in principle could remove the entire glycan. This pre-treatment was essential to yield crystals. Crystals of Δ89ST6Gal-I were obtained within 1–2 d in a sitting-drop vapour-diffusion setup at 294 K by mixing 150 nl 15 mg ml^{-1} Δ89ST6Gal-I in 20 m*M* potassium phosphate pH 6.5, 100 m*M* NaCl with precipitant in a 1:1(*v*:*v*) ratio. The cytidine/phosphate complex was obtained from 0.1 *M* MES–NaOH pH 5.8, 20% PEG 2000 MME, 0.01 *M* $CaCl_2$, 0.01 *M* $MgCl_2$. The CMP complex was obtained from 0.1 *M* HEPES–NaOH pH 7.0, 20% PEG 550 MME, 0.01 *M* $CaCl_2$, 0.01 *M* $MgCl_2$. For the iodide derivative (Dauter *et al.*, 2000), a crystal obtained from 0.1 *M* MES–NaOH pH 5.8, 22.1% PEG 2000 MME, 0.04 *M* $CaCl_2$, 0.04 *M* $MgCl_2$ was soaked for ∼10 s in mother liquor supplemented with 0.2 *M* KI. Prior to data collection, crystals were cryoprotected using paraffin oil and flash-cooled by hyperquenching (Warkentin & Thorne, 2007). Data were collected to better than 2.5 Å resolution [$I/\sigma(I) \geq 1.5$ and $CC_{1/2} > 0.5$; Table 1] over 180° (360° for the iodide soak) at 0.25° per image at 100 K on beamline PX-I at the Swiss Light Source using a PILATUS-6M detector. Data were integrated with *XDS* (Kabsch, 2010) and scaled with *SADABS* (Bruker). Indexing and integration was possible in a primitive hexagonal setting of point group 6/*m*. Systematic weak reflections along c^* ($l \neq 6n$) from data processed in space group *P*6 indicated space group $P6_1$ or its enantiomorph. Assuming one Δ89ST6Gal-I molecule in the asymmetric unit, the Matthews coefficient (Matthews, 1968) is 2.9 Å^3 Da^{-1} with a solvent content of 57%.

2.3. Phasing and refinement

Molecular replacement with *Sus scrofa* β-galactoside α-2,3-sialyltransferase, which displays 30% sequence identity (38% similarity) to ST6Gal-I, was unsuccessful; hence, a SIRAS experiment was performed. As a compromise between

Table 2
Refinement statistics.

Structure	Cytidine/phosphate (PDB entry 4js1)	CMP (PDB entry 4js2)
Resolution range (Å)	46.4–2.09 (2.19–2.09)	38.5–2.30 (2.44–2.30)
No. of reflections	22159 (2712)	16396 (2546)
R_{cryst}¶ (%)	19.5 (33.2)	17.5 (34.8)
R_{free}†† (%)	22.8 (37.3)	21.3 (39.3)
No. of residues	318	318
No. of H_2O	41	43
No. of ligands	2	1
No. of sugars†	9	10
Coordinate error‡ (Å)	0.24	0.33
Phase error‡ (°)	34.1	31.8
R.m.s.d. bonds (Å)	0.009	0.011
R.m.s.d. angles (°)	1.25	1.31
Ramachandran plot§ (%)		
Core regions	95.3	96.2
Allowed regions	4.4	3.5
Disallowed regions	0.3	0.3
$\langle B \rangle$ (Å^2)		
Protein	78 ± 25	83 ± 17
H_2O	58 ± 7	72 ± 9
Ligands	53 ± 9	60 ± 3
Sugars	99 ± 43	99 ± 22

†† $R_{cryst} = \sum_{hkl} ||F_{obs}| - |F_{calc}|| / \sum_{hkl} |F_{obs}|$, where F_{obs} and F_{calc} are the structure-factor amplitudes from the data and the model, respectively. R_{free} is R_{cryst} calculated using a 5% test set of structure factors. † The fucose is disordered in the cytidine/phosphate structure. ‡ Coordinate and maximum-likelihood-based phase errors were calculated with *PHENIX* (Zwart *et al.*, 2008). § Calculated using *Coot* (Emsley *et al.*, 2010). The outlier is Ala368, which is well defined by electron density.

maximum anomalous signal and tolerable radiation damage, SAD data for the iodide derivative were collected at a wavelength of 1.77 Å, where the calculated f'' is 8.5 e$^-$. *autoSHARP* (Vonrhein *et al.*, 2007) was used to locate and refine the substructure. Density modification was performed with *SOLOMON* (Abrahams & Leslie, 1996) and an initial model was built with *Buccaneer* (Winn *et al.*, 2011) that could, however, not be refined. Manual addition of secondary-structure elements yielded a model that was initially refined in *BUSTER* (Blanc *et al.*, 2004). The model was rebuilt in *Coot* (Emsley *et al.*, 2010) and during later stages was refined with *PHENIX* (Zwart *et al.*, 2008). Refinement statistics are collected in Table 2. Analysis of an electron-density map calculated with anomalous differences and refined phases as Fourier coefficients revealed the two expected peaks of >10 r.m.s.d. at sensible chemical environments for the iodides.

3. Results and discussion

3.1. Purification and analysis of recombinant human Δ89ST6Gal-I

The truncated recombinant human Δ89ST6Gal-I produced in HEK cells was purified from the culture supernatant by an ultrafiltration step followed by two chromatographic steps. The final enzyme preparation showed a purity of 98% on analytical size-exclusion chromatography. The enzyme migrated as a single band on SDS–PAGE with an apparent molecular weight of 38 kDa. Mass-spectrometric analysis of the enzyme preparation showed that the majority (∼90%) of the Epo-AP-Δ89ST6Gal-I was produced without the N-terminal amino acids AP, while a small fraction retained the N-terminal Pro89, which is present in the crystal structure. The recombinant human Δ89ST6Gal-I exhibits a specific activity of 3.34 nmol μg^{-1} min^{-1}, which is reduced to 5% in the presence of the inhibitor CTP (0.47 m*M*; data not shown). Mass spectrometry of the purified enzyme identified carbohydrate arrays of masses 2350.8 and 2391.9 Da that differ by an acetyl group at the distal sialyl residue and are compatible with the structures shown in Fig. 1(*a*). ∼50% of the protein preparation carries glycans at both predicted N-glycosylation sites, Asn149 and Asn161, while the rest contains only a single glycan (data not shown).

3.2. Structure of ST6Gal-I

The diffraction data of ST6Gal-I indexed in a primitive hexagonal setting with some crystals following 6/*m* and others 6/*mmm* Laue group symmetry. A self-rotation function calculated from data reduced in space group *P*6 (Fig. 2*a*) showed additional twofold axes perpendicular to the crystallographic sixfold and spaced 30° apart, indicating slight hemihedral twinning towards 622 symmetry. This higher metric symmetry is impossible for ST6Gal-I since it would lead to impossibly dense crystal packing: the $P6_1$ asymmetric unit contains a single copy of ST6Gal-I with a Matthews coefficient of 2.9 Å^3 Da^{-1}. It has been shown previously that MAD structure determination using highly twinned data is feasible (Yang *et al.*, 2000; Rudolph *et al.*, 2003) and that even SAD phasing of perfectly twinned data is possible in favourable cases (Yang *et al.*, 2000). The twin fractions of the data sets used for structure determination and refinement were estimated at <0.2 and twinning could be ignored in all subsequent steps without detrimental effects on electron density and structure quality. A substructure of two iodide sites was found using a SIRAS experiment in *RANTAN* (Winn *et al.*, 2011). Refinement of the substructure in *autoSHARP* followed by density modification using *SOLOMON* (Winn *et al.*, 2011) yielded maps with a clear solvent boundary (Fig. 2*b*). SAD phasing was unsuccessful, underlining the isomorphous contribution of the heavy atoms to SIRAS phasing. The phasing power is less than spectacular and falls below 1 at a resolution of ∼5 Å (Table 1). However, the SIRAS electron-density map was of sufficient quality to place secondary-structure elements. During the later stages of refinement, a large glycan was identified in $2mF_o - DF_c$ electron-density maps (Fig. 2*c*). The models for the CMP and cytidine/phosphate complexes were refined to R_{free} values of <23% and displayed excellent stereochemistry (Table 2).

ST6Gal-I adopts the GT-A variant 2 fold, a seven-stranded central β-sheet flanked by α-helices (Figs. 2*c* and 3). The closest member of known structure is the porcine CMP-Neu5Ac β-galactosamide α-2,3-sialyltransferase (ST3Gal-I), a transferase specific for the 3-position. Domain alignment (Holm *et al.*, 2008) shows conserved cores of ∼200 residues (*DALI* score of 17) that superpose with an r.m.s.d. of 2.9 Å (Figs. 3*a* and 3*b*). A notable difference in the core regions is an inverted C-terminal β-strand in human ST6Gal-I relative to

porcine ST3Gal-I. The inversion is owing to the insertion of a small α-helix in front of the terminal β-strand, which reverses the entry point of the polypeptide chain into the β-sheet (arrow in Fig. 3*b*). As a result, the C-termini are placed at opposite ends of the molecules. Further extensive differences between ST6Gal-I and ST3Gal-I are present at the perimeters of the molecules: ST6Gal-I has a large N-terminal appendage spanning residues 89–136 (top part of Fig. 3*a*), which is involved in substrate binding (see below). The first four N-terminal residues (89–92) are stabilized by crystal contacts, and the catalytic domain of ST6Gal-I begins at Phe93. Earlier domain-boundary assignments for ST6Gal-I were tentative (Datta, 2009). A loop region of 11 residues was not visible in ST3Gal-I (cyan spheres in Fig. 3) but is well defined in ST6Gal-I, where it covers the active site. An insertion region in ST3Gal-I ensures substrate orientation to place the 3′-hydroxyl group close to the sialyl donor (see below). This loop region is absent in ST6Gal-I, where substrate orientation is brought about by the N-terminal appendage.

3.3. Nucleotide and nucleoside binding of ST6Gal-I and related STs

CMP was present in a large excess (10 m*M*) in all crystallization trials, and two interesting complexes were determined: one with CMP and another with cytidine and phosphate bound at the active site. The origin of cytidine is owing to impurities in the CMP preparation (data not shown), while the phosphate stems from the protein buffer. In both complexes

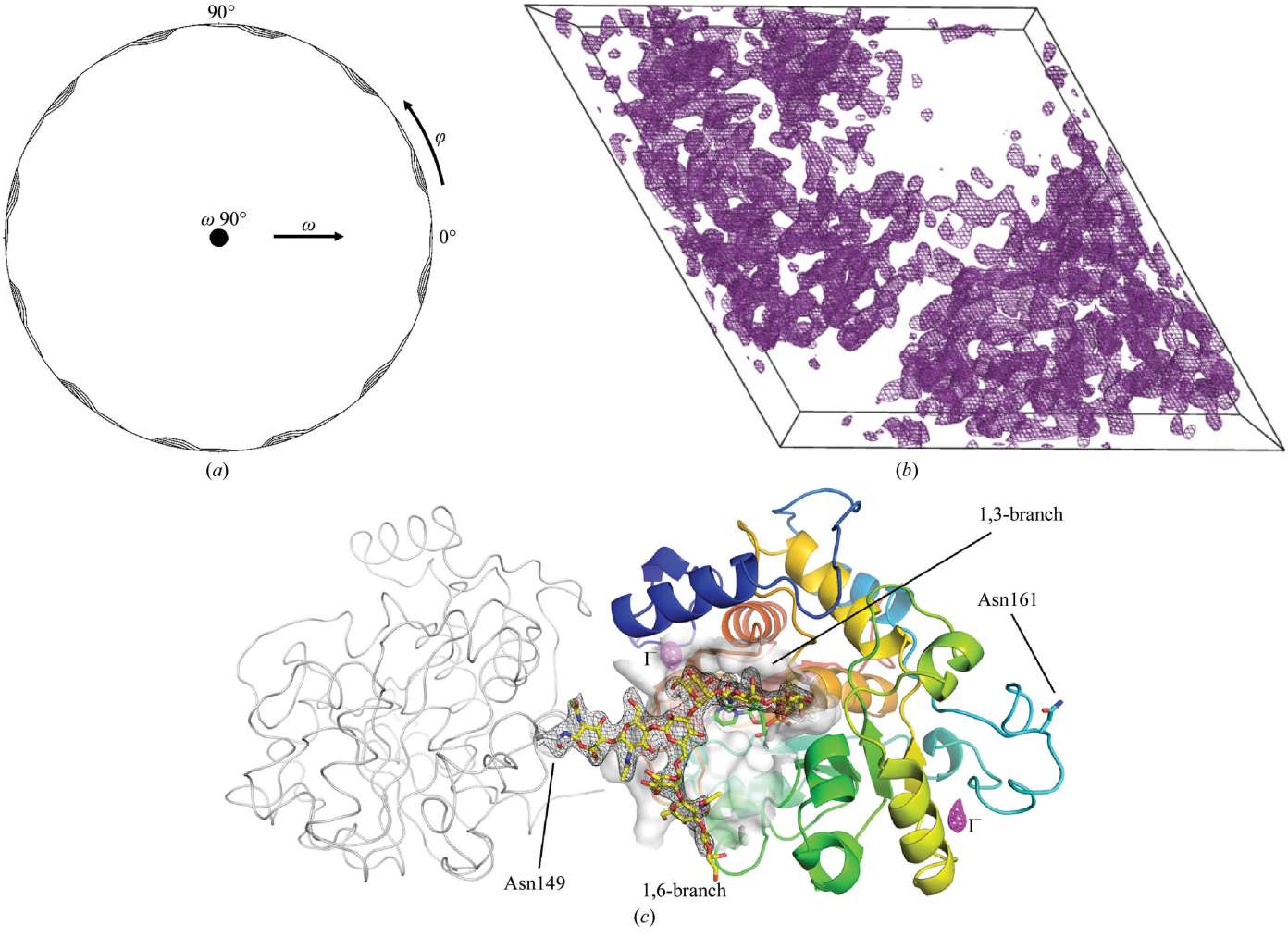

Figure 2
Data analysis and structure determination. (*a*) $\kappa = 180^\circ$ section of a self-rotation function calculated at 3 Å resolution from data processed in space group *P*6. The pattern of twofold axes perpendicular to the crystallographic sixfold with a 30° spacing (perimeter of the circle) indicates slight twinning. (*b*) The SIRAS-phased electron-density map contoured at 1.5 r.m.s.d. after solvent flipping and histogram matching with *SOLOMON* (Abrahams & Leslie, 1996) shows a clear solvent boundary. (*c*) The final model is rainbow-coloured from the N-terminus to the C-terminus. The positions of the two iodide sites (I^-) are indicated by an anomalous map (magenta mesh, contoured at 5 r.m.s.d.) calculated using the anomalous differences of the structure-factor amplitudes and the refined phases as Fourier coefficients. A large glycan (drawn as sticks and grey binding surface) attached to Asn149 is contributed to the active site by a crystallographic neighbour (grey). The σ_A-weighted electron density for the glycan after refinement is contoured at 1 r.m.s.d. and shows that the 1,6-branch of the glycan is less well ordered than the 1,3-branch that reaches into the active site. From the major glycosylation trees shown in Fig. 1(*a*), the two terminal sialic acids (and the fucose in the cytidine/phosphate complex) are not visible in the electron density. The second glycosylation site at Asn161 had its glycan removed by peptide *N*-glycosidase F.

the ligand is deeply buried in the protein and is bound by numerous hydrogen bonds and hydrophobic interactions (Figs. 4*a* and 4*b*). Both complexes share the same hydrogen-bonding patterns for the cytosine moiety but differ at the phosphate-binding site. Cytosine specificity is established by a hydrogen bond from the exocyclic amino group to the main-chain carbonyl group of Cys353. Cys353 also forms a unique disulfide bond to Cys364 that packs on top of the nucleobase (Fig. 3*a*). The ribose is bound in a 3′-*endo* conformation by the first and last residues of an SSG sequence that belongs to sialyl motif S.

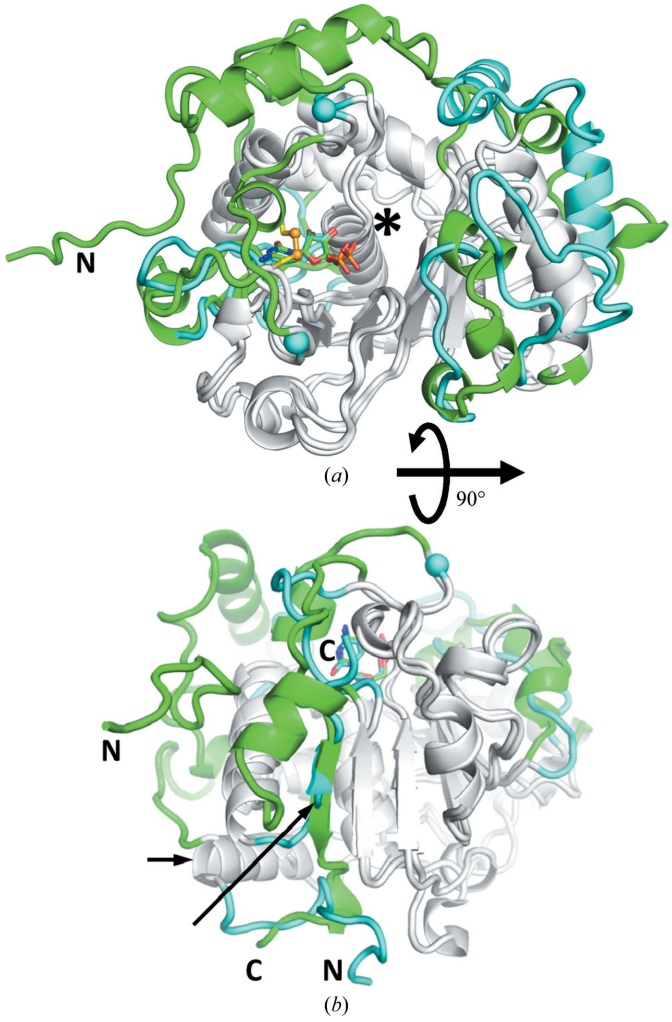

Figure 3
ST6Gal-I adopts a modified GT-A variant 2 fold. (*a*) Superposition of ST6Gal-I (green) with porcine ST3Gal-I (PDB entry 2wnb; cyan; Rao *et al.*, 2009). Conserved parts are drawn in grey to highlight the differences between these structures that share the same GT-A fold. The view is into the active site, showing the close correspondence of the bound CMP in both structures. Cyan spheres mark the boundaries of a missing loop region in porcine ST3Gal-I. The corresponding loop is well defined in human ST6Gal-I and contains the Cys353–Cys364 disulfide bond (yellow) covering the cytidine. The central α-helix orienting its dipole towards the CMP phosphate is marked by an asterisk. (*b*) View rotated 90° counterclockwise about the horizontal axis to highlight the inverted C-terminal β-strand in both structures (long arrow). The termini are labelled and the small horizontal arrow shows the first common secondary-structure element of the two structures.

Comparison of ST6Gal-I with the bifunctional ST3/8 from *Campylobacter jejuni* (Chiu *et al.*, 2004; Fig. 4*c*) and porcine ST3Gal-I (Rao *et al.*, 2009; Fig. 4*d*) reveals that CMP binds in the same conformation but that the chemical environment is quite different. Common features are a hydrogen bond from the α-phosphate to a donor residue, either tyrosine or histidine. The ribose always adopts the 3′-*endo* conformation and the ring O4 atom binds to a protein NH group (not shown in the figures). The 2′- and 3′-hydroxyl groups contact the amide NH groups of an (S/T)(S/T)G sequence motif at the beginning of sialyl motif S, although the specific hydrogen-bonding pattern can vary. For instance, in the bacterial ST3/8 (Fig. 4*c*) the 3′-hydroxyl group accepts a hydrogen bond from the second residue (Ser132) of the motif, while in human ST6Gal-I this hydrogen bond is donated by the third residue (Gly324). The (S/T)(S/T)G sequence motif represents the first turn of an α-helix in the GT-A variant 2 fold, the positively polarized dipole of which is oriented towards the phosphate moiety of the CMP derivative, possibly stabilizing its negative charge (marked with an asterisk in Fig. 3*a*). In porcine ST3Gal-I the 2′-hydroxyl group hydrogen bonds to a water molecule, not the protein (Fig. 4*d*). Of note, the exocyclic amino group of cytosine does not engage in specific contacts with the protein, raising the possibility that porcine ST3Gal-I could also accept the non-natural donor substrate UMP-Neu5Ac, which is impossible for human ST6Gal-I.

In the ST6Gal-I–cytidine–phosphate complex, the inorganic phosphate has moved by 3.5 Å relative to the α-phosphate in CMP owing to bond breakage between the cytosine and the phosphate moiety and a resulting van der Waals repulsion (Fig. 4*b*). This complex represents an off-pathway situation as STs do not hydrolyse CMP, but lends valuable insight into binding of the competitive inhibitors CDP and CTP (Fig. 4e). The contact between phosphate and Tyr354 is lost but is replaced by an extended hydrogen-bond network involving Asn233, Ser323 and the catalytic His370. Superposition of the CMP and the cytidine/phosphate complexes shows that the protein conformations are quite similar, which suggests that the nucleotide-binding pocket is pre-formed. Interestingly, two oxygen atoms from CMP and inorganic phosphate locate to almost the same position (1.2 Å distance). A model for the weak inhibitor CDP (K_i = 19.0 μM; Klohs *et al.*, 1979) can be constructed by condensing the phosphate groups to an anhydride followed by energy minimization. The β-phosphate of the modelled CDP locates to the exact same position as the carboxylate group in an Neu5Ac-containing crystal structure and interacts with Ser323 (see below). Thus, the charge of the β-phosphate mimics the charge of the donor substrate. CDP would interfere with the binding of the donor substrate CMP-Neu5Ac, but not the sugar acceptor. Based on the CDP model, CTP (K_i = 16.0 μM) can be placed into the active site without steric repulsions. The γ-phosphate group of CTP would clash with the galactose of the acceptor substrate (not shown). In accord, cocrystals of ST6Gal-I in complex with CTP could not be obtained. Taken together, the competitive binding modes of CDP and CTP are similar but the detailed modes of inhibition by the nucleotides seem to be different.

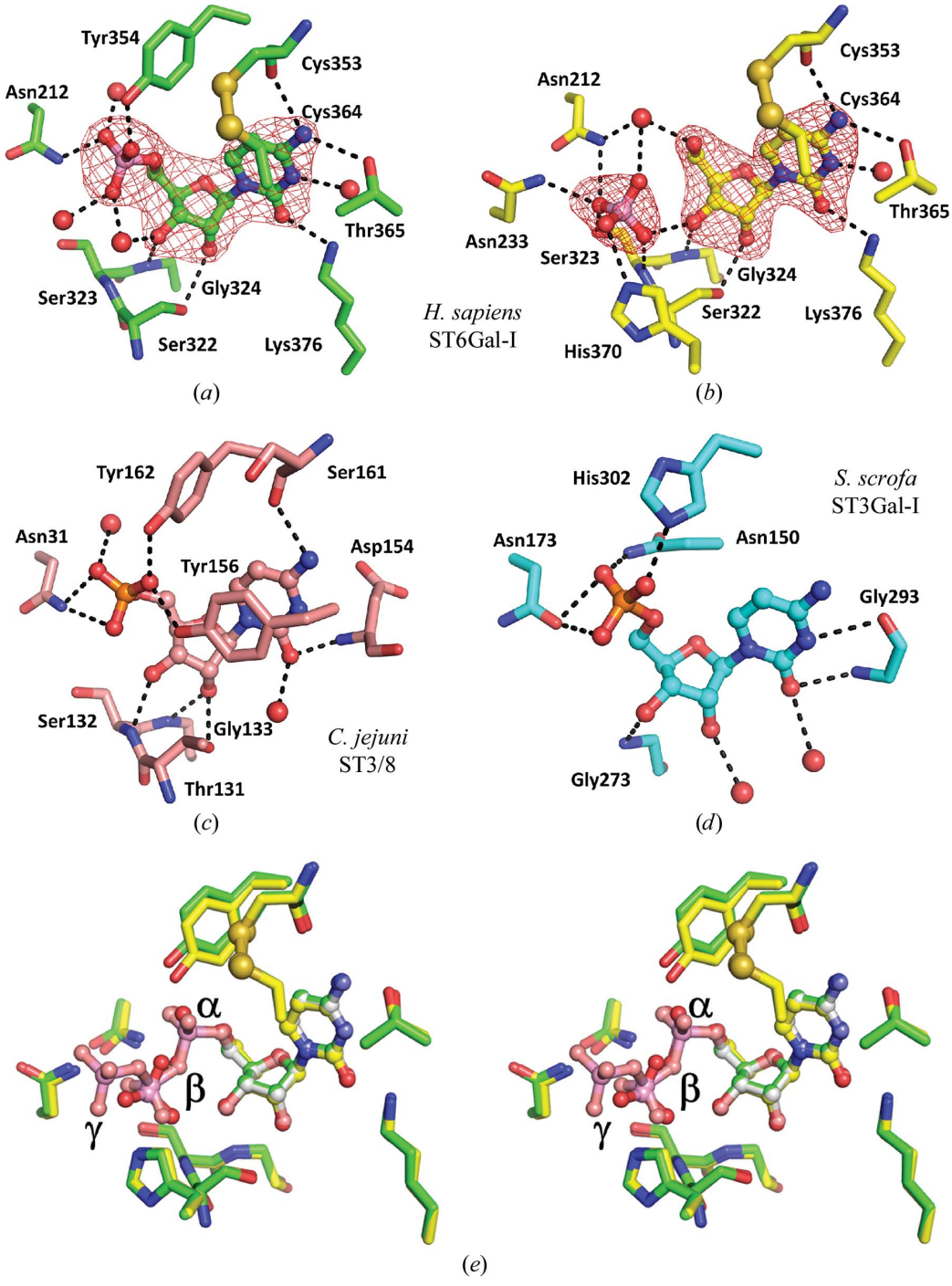

Figure 4
Nucleotide binding to ST6Gal-I. CMP is shown as a ball-and-stick model and protein residues contacting CMP are drawn as sticks. Possible hydrogen bonds are indicated as dashed lines and water molecules as red spheres. The orientations of all panels are the same after superimposing the CMP moieties. (*a*) The ST6Gal-I–CMP complex. Almost the full hydrogen-bonding potential of CMP is fulfilled by water or protein contacts. Specificity for cytosine is ensured by a hydrogen bond from its exocyclic amino group to the main-chain carbonyl group of Cys353. The Cys353–Cys364 disulfide bond cordons off a hydrophobic pocket for cytosine. (*b*) The ST6Gal-I–cytidine–phosphate complex. The phosphate moiety has moved compared with its position in CMP. The red mesh in (*a*) and (*b*) shows σ_A-weighted $mF_o - DF_c$ OMIT electron density for the ligands contoured at 3 r.m.s.d. (*c*) CMP binding to *C. jejuni* ST3/8 (PDB entry 1ro8; Chiu *et al.*, 2004). This bifunctional ST retains the tyrosine (Tyr162) for contacting the α-phosphate. Tyr156 acts as a hydrophobic lid on top of the cytosine. Base specificity is ensured by a hydrogen bond to the carbonyl group of Ser161. The ribose binds to the residue range Thr131–Gly133. (*d*) CMP binding to porcine ST3Gal-I (PDB entry 2wnb; Rao *et al.*, 2009). His302 accepts a hydrogen bond from the α-phosphate. No apparent base specificity is established by hydrogen bonds. (*e*) Cross-eyed stereo image of the superposition of the two ST6Gal-I complexes (*a* and *b*) allows modelling of the competitive inhibitors CDP and CTP into the active site (lighter hue and labelled α–γ). The conformation of the cytosine parts is almost identical in the complexes. The position of the inorganic phosphate in the cytidine/phosphate complex is precisely at the location of a β-phosphate in a low-energy CDP conformation. The γ-phosphate for CTP can be modelled without steric clashes and inclusion of a number of hydrogen bonds (not shown). Only the CTP model is shown. Models of the ligands were constructed and energy-minimized with *MOLOC* (http://www.moloc.ch) while keeping the protein rigid.

While many GT-A-type transferases use UDP-activated donor sugars in complex with a metal ion to neutralize the phosphate charges, STs are metal-ion independent and do not possess the required metal-coordinating D*x*D motif (Audry *et al.*, 2011). The absence of neutralizing positive charges may explain the relatively weak inhibitory activity of CDP and CTP, despite the option of forming several additional hydrogen bonds compared with CMP.

3.4. Glycan binding at the active site of STs

A salient feature that distinguishes the ST6Gal-I structure from other ST structures determined to date is the presence of a large glycan bound to Asn149 that is ordered by a crystal contact (Fig. 2*c*) and mimics a substrate complex (see below). The second glycosylation site at Asn161 in ST6Gal-I does not carry a glycan in the present crystal structures. Indeed, a glycan at this position would disrupt the crystal lattice, highlighting the importance of peptide *N*-glycosidase F treatment to obtain crystals. Apparently, only partially deglycosylated ST6Gal-I can crystallize in this setting. The sometimes beneficial effect of adding catalytic amounts of hydrolases either prior to or *in situ* during crystallization in order to trim macromolecules has been noted previously (Dierks *et al.*, 2005; Johnson *et al.*, 2006; Dong *et al.*, 2007; Bai *et al.*, 2007; Wernimont & Edwards, 2009; Gheyi *et al.*, 2010; Bunker *et al.*, 2012; Anthony *et al.*, 2010). The case of human ST6Gal-I underscores that differential deglycosylation adds another parameter to the crystallizability of proteins and that partial *in situ* deglycosylation may be useful in cases of glycoproteins that are recalcitrant to crystallization.

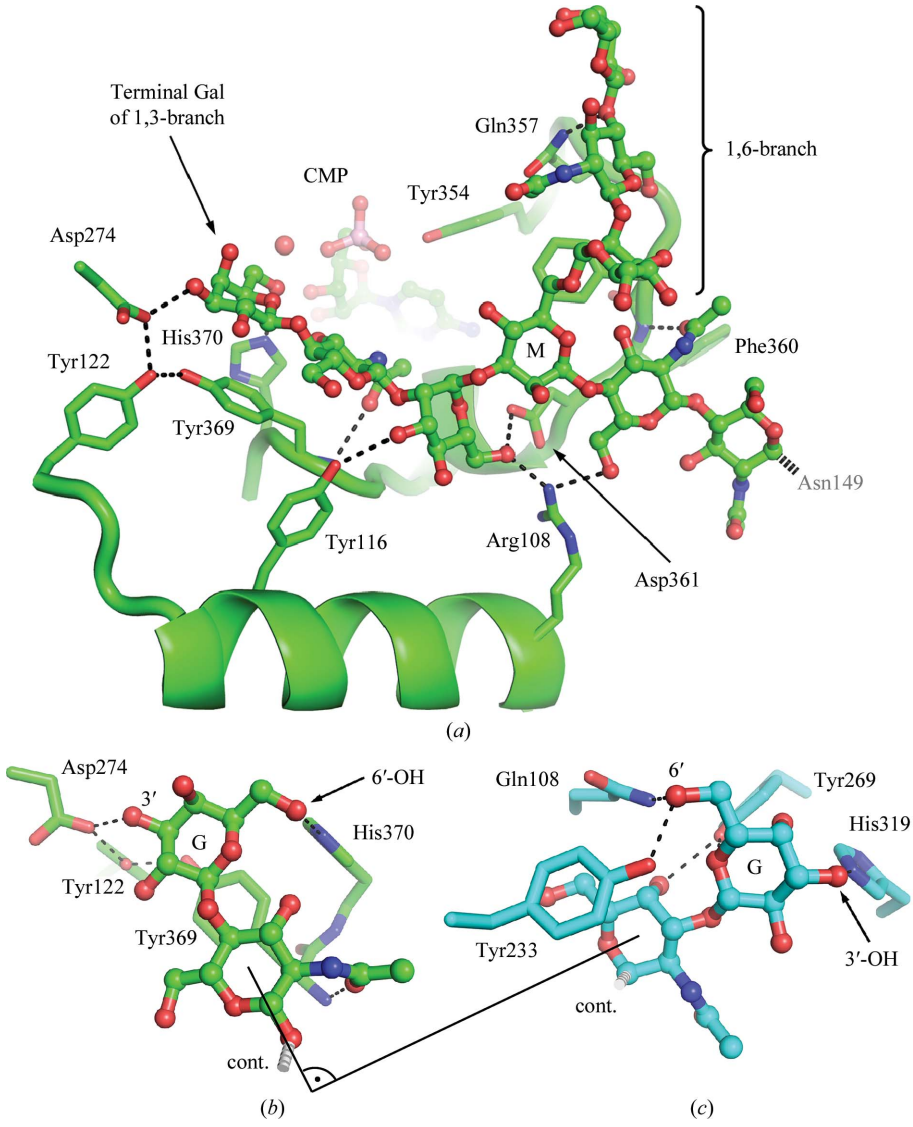

Figure 5
Glycan binding to ST6Gal-I and the origin of regioselectivity. (*a*) Overview of the glycan binding mode to ST6Gal-I. The glycan is drawn as a ball-and-stick model and relevant protein side chains are shown as sticks. The mannose moiety at the glycan bifurcation is labelled M. The 1,6-branch (right side) hydrogen bonds to Gln357 and is involved in fewer interactions than the 1,3-branch. The 1,3-branch entertains five hydrogen bonds and places a galactose into the proximity of CMP and the catalytic His370. (*b*) Close-up of the terminal galactose (G) in ST6Gal-I. A hydrogen bond between the 3′-hydroxyl group and Asp274 has the 6′-hydroxyl group interacting with His370. (*c*) Close-up of the acceptor galactose (G) in ST3Gal-I. As in (*b*) the last two sugar residues are shown. The entry vector into the active site is about 90° different from the situation in ST6Gal-I. The terminal galactose is flipped relative to the situation in ST6Gal-I. Its 6′-hydroxyl group binds to Gln108, leaving the 3′-hydroxyl group to interact with catalytic His319.

The glycan at Asn149 adopts a Y-shape (Figs. 1*a*, 2*c* and 5*a*), with the central mannose bifurcating into a 1,3-branch and a 1,6-branch. No protein–protein interactions are present that could skew the conformation of the glycan. Although detected by mass spectrometry, the sialyl moieties are missing in the structure, pointing to either spontaneous hydrolysis during crystallization or residual sialidase activity of ST6Gal-I. The terminal galactose moiety of the 1,3-branch binds into the active site of ST6Gal-I, although from steric, volume and chemical considerations (both branches end with galactose) either branch could bind. ST6Gal-I exhibits a branch preference, sialidating the 1,3-branch much faster than the 1,6-branch. The 1,3-branch selectivity is independent of the glycan-presenting protein because it is observed both for glycans bound to the Fc part of IgG immunoglobulins and for glycans that are not bound to a protein (Barb *et al.*, 2009). The selectivity of ST6Gal-I for the 1,3-branch is represented in the crystal structure. No protein–protein crystal contacts dictate the conformation of the glycan contributed by the crystallographic neighbour; thus, the glycan adopts its preferred

orientation with the 1,3-branch in the active site. While the 1,6-branch of the glycan forms only a single hydrogen bond to ST6Gal-I (Gln357 in motif III) and the terminal galactose is only weakly defined by electron density, this situation is very different for the 1,3-branch, which is specifically bound by five hydrogen bonds (Fig. 5*a*) and exhibits excellent electron density. Tyr369 packs under the terminal galactose moiety and contributes a number of hydrophobic contacts. Other important 1,3-branch-binding side chains are Tyr116 and Arg108, which face the same side of an α-helix that has not been observed in other ST structures. Deletion of the first 100 residues in ST6Gal-I abolishes activity (Legaigneur *et al.*, 2001). Since Arg108 is the first residue specifically contacting the glycan substrate, it can be envisaged that deletion of the first 100 residues perturbs the structure of this substrate-binding area, leading to a loss of activity. Alternatively, ST6Gal-I could unfold since the deletion of the 100 N-terminal residues would expose hydrophobic surfaces to the solvent. Another aspect of glycan binding to ST6Gal-I relates to substrate specificity at the penultimate saccharide. ST6Gal-I is specific for *N*-acetylglucosamine (NAG) at this position (Wlasichuk *et al.*, 1993), which is explained by a hydrogen bond from the *N*-acetyl group to the main-chain NH group of Tyr369 (Fig. 5*a*). This high substrate selectivity stands in contrast to ST6GalNAc-I, an ST with the same regiospecificity as ST6Gal-I but which uses GalNAc as an acceptor. ST6GalNAc-I displays broad substrate specificity and requires either Ser/Thr (in the case of O-linked GalNAc-O-Ser/Thr) or galactose (in the case of Galβ-1,3-GalNAc-O-Ser/Thr) as the penultimate residue (Kurosawa *et al.*, 1994, 2000; Ikehara *et al.*, 1999).

An important question relates to the regiospecificity of the ST reaction. Sequence alignment of porcine ST3Gal-I and human ST6Gal-I identified a catalytic His370 in the human enzyme which is equivalent to His319 in the porcine enzyme (Rao *et al.*, 2009). The structure of ST3Gal-I was determined in complex with a modified β-1,3-linked *N*-acetyl-2-aminogalactosyl galactoside (Rao *et al.*, 2009). The ST6Gal-I and ST3Gal-I crystal structures reveal different binding modes for the terminal galactose moieties of the glycans that reflect the regiospecificity of the respective STs. Both galactose moieties orient their nucleophilic hydroxyl groups towards the catalytic histidines and hydrogen bond to the imidazole side chain (Figs. 5*b* and 5*c*). When viewed along the glycosidic bond between the penultimate and last sugar moieties, the 3′- and 6′-hydroxyl groups are roughly on opposite sides of the terminal galactose. The ST thus needs to select one of the two faces of the sugar moiety. In human ST6Gal-I this selection is realised by a hydrogen bond between the 3′-hydroxyl group and the Asp274 side chain that itself is fixed in a hydrogen-bond network with Tyr122 and Tyr369 (Fig. 5*b*). In porcine ST3Gal-I, the terminal galactose is rotated by 180° relative to the galactose in ST6Gal-I. Gln108 and Tyr233 interact with the 6′-hydroxyl group and Tyr269 forms two hydrogen bonds to the 4′-hydroxyl (Fig. 5*c*). The latter interaction has been suggested to be crucial for ST3Gal-I specificity (Rao *et al.*, 2009). Indeed, the whole 1,3-branch is flipped when the two STs are compared (Figs. 5*b* and 5*c*). Residues Asp274 in ST6Gal-I and Gln108 and Tyr269 in ST3Gal-I are conserved within their respective subfamilies. Another difference that could help to establish regiospecificity is the direction from which the glycan enters the respective active site, which differs by about 90° between the structures (angle indicated in Figs. 5*b* and 5*c*).

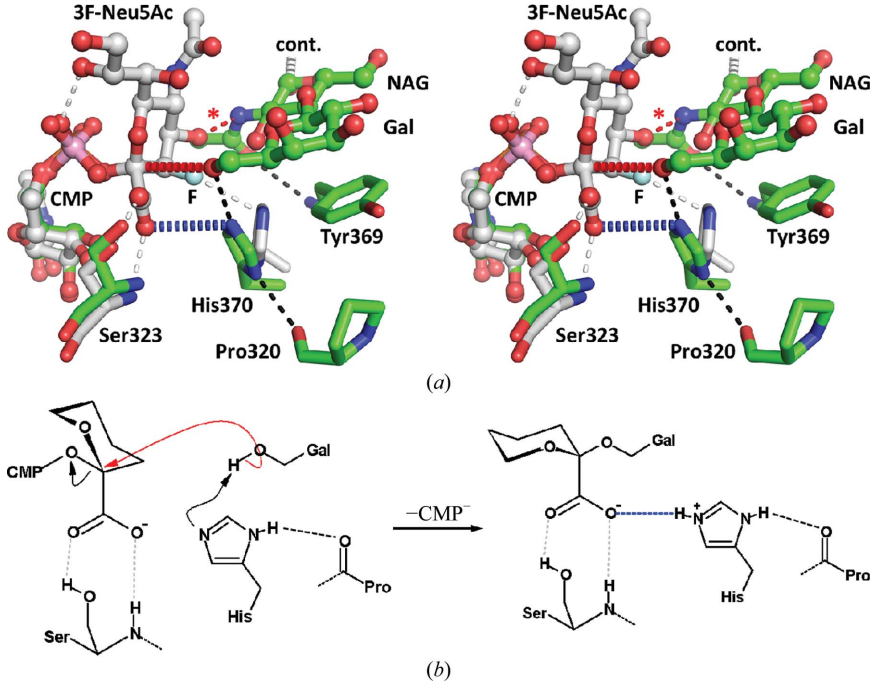

Figure 6
Plausible architecture of the Michaelis complex and catalysis. (*a*) For this stereoview, the crystal structures of ST6Gal-I (green) and *C. jejuni* CstII ST3Gal (PDB entry 2p2v; Chiu *et al.*, 2007) in complex with the inert donor CMP-3F-Neu5Ac (grey; left side) were superimposed on their CMP parts. Hydrogen bonds (thin dashed lines) for CstII are drawn in grey and those for ST6Gal-I are drawn in black. The last two parts of the acceptor glycan (NAG-Gal) in ST6Gal-I are shown on the right side. The F atom in CMP-3F-Neu5Ac is coloured light cyan and labelled F. The presence of the fluorine in CstII is incompatible both with the normal position of the catalytic histidine and also the 6′-hydroxyl group, explaining its inhibition. A favourable interaction between the substrates is a hydrogen bond between sialic acid and NAG (red asterisk). For the model of the Michaelis complex, the fluorine may safely be ignored. The 6′-hydroxyl is in line and at a favourable distance (3.3 Å) for S_N2 attack on the C2 atom of Neu5Ac (thick red dashed line). His370 in ST6Gal-I is pre-oriented by Pro320 and can act as the catalytic base for deprotonation of the 6′-hydroxyl. The generated histidinium ion would then be stabilized by the nearby (3.4 Å) Neu5Ac carboxylate (thick blue dashed line). (*b*) Schematic drawing of the nucleophilic substitution with transferred protons indicated. The nucleophilic attack by the 6′-hydroxyl group is indicated by a red arrow. The substituents of the sialic acid are not shown apart from the carboxylate.

3.5. Mechanistic implications

The substitution reactions of GTs can lead to inversion or retention of the electrophilic centre in the product, irrespective of the enzyme adopting the GT-A or the GT-B fold. In contrast, all STs are inverting enzymes that use an S_N2-like direct displacement mechanism (Audry *et al.*, 2011). The mechanism requires the generation of a nucleophile by a catalytic base and stabilization of the separated charges throughout the reaction, *i.e.* during the transition state and in the leaving group. A key question therefore is which elements in ST6-Gal-I are responsible for nucleophile activation and charge neutralization.

Using the currently available structures of STs, a Michaelis (enzyme–substrate) complex prior to the nucleophilic substitution reaction can be constructed. The crystal structure of *C. jejuni* CstII, an ST6Gal-I with the same specificity as, but from a different subfamily to, porcine ST3Gal-I, has been determined in complex with an inert analogue of the CMP-Neu5Ac donor in which the C3 atom of the neuraminic acid carries an F atom instead of hydrogen (3F-Neu5Ac; Chiu *et al.*, 2004). Comparison of this donor complex with the human ST6Gal-I acceptor structure yields interesting insight into the Michaelis complex. After superposition of the CMP parts of CstII and ST6Gal-I no critical clashes were observed between the donor and acceptor substrates. The fluorine in 3F-Neu5Ac is at a distance of only 1.8 Å from the 6′-hydroxyl of the acceptor. Also as a result of the larger volume of fluorine compared with hydrogen, the catalytic histidine in CstII moves away from the donor substrate. Thus, steric repulsion of the acceptor and catalytic histidine by the fluorine is sufficient to explain the inhibition by 3F-Neu5Ac. No conformational changes of the substrates are necessary to accommodate them in the active site. An equilibrium random-order mechanism of substrate binding is therefore possible for ST6Gal-I, similar to that found previously for ST3Gal-I (Rearick *et al.*, 1979). Indeed, a favourable hydrogen bond between the substrates is possible, connecting a hydroxyl group of the donor and the acetamide group of the acceptor (indicated by a red asterisk in Fig. 6). Intriguingly, in this simple superposition model the 6′-hydroxyl group is within van der Waals distance (3.3 Å) of the electrophilic C2 centre of the donor sugar (indicated by a red dashed line in Fig. 6), ready to initiate the reaction.

The orientation of the His370 imidazole side chain is fixed by a hydrogen bond to the backbone carbonyl group of Pro320 (Fig. 6). The $N^{\delta1}$ atom of His370 must therefore be protonated, leaving the $N^{\varepsilon2}$ atom free to act as a catalytic base. Deprotonation of the 6′-hydroxyl group of the acceptor galactose by His370 would generate a highly active alkoxy nucleophile that attacks the C2 atom of CMP-Neu5Ac. Quantum-mechanical calculations for the inverting GT β-1,2-*N*-acetylglucosaminyltransferase I that possesses an aspartate as the catalytic base suggested a concerted deprotonation and nucleophilic attack (Kozmon & Tvaroska, 2006), which may also apply to STs. A question remains as to what happens to the separated charges. Three GTs involved in the biosynthesis of the antibiotic calicheamicin (CalG1, CalG3 and CalG4) relay the positive charge on the histidine to a nearby aspartate or glutamate in a catalytic dyad. Likewise, serine proteases stabilize the histidinium ion by a charge relay to an aspartate at a hydrogen-bonding distance (a catalytic dyad, with the activated serine completing the classic catalytic triad). In ST6Gal-I the closest negative charge to stabilize the protonated His370 is neuraminic acid, the O atoms of which accept hydrogen bonds from Ser323 (Fig. 6) and which interacts with the positive end of the α-helix dipole at Ser323 (see Fig. 3*a*). This arrangement suggests substrate-assisted catalysis. The distance between the carboxylate group of Neu5Ac and the histidinium ion is 3.4 Å in the model, conveniently close for an electrostatic interaction. A slight movement of either residue could even establish a very strong charged hydrogen bond, similar to the cases of calicheamicin GTs and serine proteases. On the other side of the complex, the developing negative charge on the α-phosphate of the CMP leaving group could be stabilized by Asn212 or entirely neutralized by either proton transfer from Tyr354 or water (Fig. 4*a*). As the α-phosphate of the CMP leaving group is solvent-exposed (Fig. 3*a*), rapid protonation *via* the Grotthus mechanism is a possibility (de Grotthus, 1806). Other strategies to stabilize the negative charge on the leaving group are metal ions in UDP-activated GTs or an oxyanion hole in case of serine proteases. Neither of these options is possible for ST6Gal-I owing to the absence of the metal-coordinating D*x*D motif and of residues that could construct an oxyanion hole. Finally, this model of the Michaelis complex supports the observation that STs invert the configuration at the anomeric centre. The flip at C2 would move Neu5Ac away from the acceptor galactose, avoiding repulsions, severing the putative hydrogen bond between NAG and Neu5Ac, and liberating the sialidated product.

4. Conclusions and outlook

The crystal structure of the first α-2,6-sialyltransferase reveals, by comparison with related enzymes, much insight into substrate binding and catalysis. While there are several solutions for binding of CMP-activated sugars, the conformation of the CMP moiety is conserved. The acceptor galactose in the different STs is bound by very different chemical environments that place the nucleophilic hydroxyl group in hydrogen-bonding contact with the catalytic histidine. Similar to other STs, the hydrogen-bonding pattern of galactose allows rationalization of the regioselectivity of ST6Gal-I. The elongated glycan is visible for the first time to this extent in any ST structure. We describe a stereochemically favourable model for the Michaelis complex that leads to a plausible mechanism for sialyl transfer by ST6Gal-I. The constitution of the Michaelis complex could be further tested by chemical modification of the donor and acceptor substrate, for instance by replacing the carboxylate in Neu5Ac by an uncharged isosteric nitro group or by the removal of the 6′-hydroxyl group of the acceptor galactose. Both variations should trap a non-reactive Michaelis complex. For inverting GTs, including STs, an oxocarbenium-like transition state is proposed that has motivated the development of bisubstrate and transition-state inhibitors with a planar anomeric centre (Schwörer &

Schmidt, 2002). Structural information on such a complex would certainly be of interest for medical and diagnostic applications of STs.

Many human glycoproteins contain complex N-glycans with terminal NAG-Gal sequences that differ in the pattern of sialylation, which is relevant to the function and stability of these proteins. For instance, the anti-inflammatory activity of IgG is entirely dependent on sialylation of the N-linked glycan in the IgG-Fc part, and fully sialylated IgG-Fc exhibits enhanced potency (Anthony *et al.*, 2008). Therapeutic IgG preparations are heterogeneous with respect to their glycosylation, which arises owing to the presence (G2) or lack of one or both (G1 or G0) of the terminal galactose residues in the bi-antennary glycans, which in turn determines whether or not sialic acid can be added. Enzymatic reconstruction of IgG to the fully sialylated glycans would therefore boost activity, potentially allowing lower doses to be used. Complete galactosylation of native IgG-Fc seems facile using UDP-galactose as the substrate and bovine galactosyltransferase as the catalyst (Barb *et al.*, 2009). A bottleneck towards full sialylation is the branch-specificity of ST6Gal-I, which results in monosialylated glycans in the 1,3-branch. After several days of reaction, disialylated glycans are obtained in minute amounts and only after repeated ST6Gal-I treatment using monosialylated IgG-Fc as a substrate (Barb *et al.*, 2009). Indeed, ST6Gal-I also exhibits significant sialidase activity (HS, unpublished observations), thus destroying the sialylated product again in futile catalytic cycles of transferase and hydrolase activity. Sialidase activity of ST6-Gal-I would explain the absence of sialic acid in the glycan of the ST6Gal-I crystal structure. At high concentrations ST6Gal-I could be its own substrate for sialidase activity. Thus, to increase the sialylation yield of IgG using ST6Gal-I, future studies may seek to reduce the substrate specificity and eliminate sialidase activity while maintining the regiospecificity.

We thank the staff at SLS beamline PX II and our colleagues at Expose for support during synchrotron data collection, and M. Thomann and S. Malik for mass-spectrometric analyses of the glycans.

References

Abrahams, J. P. & Leslie, A. G. W. (1996). *Acta Cryst.* D**52**, 30–42.

Anthony, C. S., Corradi, H. R., Schwager, S. L., Redelinghuys, P., Georgiadis, D., Dive, V., Acharya, K. R. & Sturrock, E. D. (2010). *J. Biol. Chem.* **285**, 35685–35693.

Anthony, R. M., Nimmerjahn, F., Ashline, D. J., Reinhold, V. N., Paulson, J. C. & Ravetch, J. V. (2008). *Science*, **320**, 373–376.

Audry, M., Jeanneau, C., Imberty, A., Harduin-Lepers, A., Delannoy, P. & Breton, C. (2011). *Glycobiology*, **21**, 716–726.

Backliwal, G., Hildinger, M., Chenuet, S., Wulhfard, S., De Jesus, M. & Wurm, F. M. (2008). *Nucleic Acids Res.* **36**, e96.

Bai, Y., Auperin, T. C. & Tong, L. (2007). *Acta Cryst.* F**63**, 135–138.

Barb, A. W., Brady, E. K. & Prestegard, J. H. (2009). *Biochemistry*, **48**, 9705–9707.

Blanc, E., Roversi, P., Vonrhein, C., Flensburg, C., Lea, S. M. & Bricogne, G. (2004). *Acta Cryst.* D**60**, 2210–2221.

Breton, C., Fournel-Gigleux, S. & Palcic, M. M. (2012). *Curr. Opin. Struct. Biol.* **22**, 540–549.

Brossmer, R. & Gross, H. J. (1994). *Methods Enzymol.* **247**, 177–193.

Bunker, R. D., Loomes, K. M. & Baker, E. N. (2012). *Acta Cryst.* F**68**, 59–62.

Chiu, C. P. C., Lairson, L. L., Gilbert, M., Wakarchuk, W. W., Withers, S. G. & Strynadka, N. C. J. (2007). *Biochemistry*, **46**, 7196–7204.

Chiu, C. P. C., Watts, A. G., Lairson, L. L., Gilbert, M., Lim, D., Wakarchuk, W. W., Withers, S. G. & Strynadka, N. C. J. (2004). *Nature Struct. Mol. Biol.* **11**, 163–170.

Dall'Olio, F. & Chiricolo, M. (2001). *Glycoconj. J.* **18**, 841–850.

Datta, A. K. (2009). *Curr. Drug Targets*, **10**, 483–498.

Datta, A. K., Chammas, R. & Paulson, J. C. (2001). *J. Biol. Chem.* **276**, 15200–15207.

Dauter, Z., Dauter, M. & Rajashankar, K. R. (2000). *Acta Cryst.* D**56**, 232–237.

Diederichs, K. & Karplus, P. A. (1997). *Nature Struct. Biol.* **4**, 269–275.

Dierks, T., Dickmanns, A., Preusser-Kunze, A., Schmidt, B., Mariappan, M., von Figura, K., Ficner, R. & Rudolph, M. G. (2005). *Cell*, **121**, 541–552.

Dong, A. *et al.* (2007). *Nature Methods*, **4**, 1019–1021.

Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). *Acta Cryst.* D**66**, 486–501.

Fenteany, F. H. & Colley, K. J. (2005). *J. Biol. Chem.* **280**, 5423–5429.

Gheyi, T., Rodgers, L., Romero, R., Sauder, J. M. & Burley, S. K. (2010). *J. Am. Soc. Mass Spectrom.* **21**, 1795–1801.

Grotthus, C. J. de (1806). *Ann. Chim.* **58**, 54–73.

Hennet, T., Chui, D., Paulson, J. C. & Marth, J. D. (1998). *Proc. Natl Acad. Sci. USA*, **95**, 4504–4509.

Holm, L., Kääriäinen, S., Rosenström, P. & Schenkel, A. (2008). *Bioinformatics*, **24**, 2780–2781.

Ikehara, Y., Kojima, N., Kurosawa, N., Kudo, T., Kono, M., Nishihara, S., Issiki, S., Morozumi, K., Itzkowitz, S., Tsuda, T., Nishimura, S. I., Tsuji, S. & Narimatsu, H. (1999). *Glycobiology*, **9**, 1213–1224.

Johnson, S., Roversi, P., Espina, M., Deane, J. E., Birket, S., Picking, W. D., Blocker, A., Picking, W. L. & Lea, S. M. (2006). *Acta Cryst.* F**62**, 865–868.

Kabsch, W. (2010). *Acta Cryst.* D**66**, 125–132.

Karplus, P. A. & Diederichs, K. (2012). *Science*, **336**, 1030–1033.

Klohs, W. D., Bernacki, R. J. & Korytnyk, W. (1979). *Cancer Res.* **39**, 1231–1238.

Kozmon, S. & Tvaroska, I. (2006). *J. Am. Chem. Soc.* **128**, 16921–16927.

Kurosawa, N., Hamamoto, T., Lee, Y. C., Nakaoka, T., Kojima, N. & Tsuji, S. (1994). *J. Biol. Chem.* **269**, 1402–1409.

Kurosawa, N., Takashima, S., Kono, M., Ikehara, Y., Inoue, M., Tachida, Y., Narimatsu, H. & Tsuji, S. (2000). *J. Biochem.* **127**, 845–854.

Lairson, L. L., Henrissat, B., Davies, G. J. & Withers, S. G. (2008). *Annu. Rev. Biochem.* **77**, 521–555.

Legaigneur, P., Breton, C., El Battari, A., Guillemot, J.-C., Auge, C., Malissard, M., Berger, E. G. & Ronin, C. (2001). *J. Biol. Chem.* **276**, 21608–21617.

Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.

Moremen, K. W., Tiemeyer, M. & Nairn, A. V. (2012). *Nature Rev. Mol. Cell Biol.* **13**, 448–462.

Neu, U., Bauer, J. & Stehle, T. (2011). *Curr. Opin. Struct. Biol.* **21**, 610–618.

Rao, F. V., Rich, J. R., Rakić, B., Buddai, S., Schwartz, M. F., Johnson, K., Bowe, C., Wakarchuk, W. W., Defrees, S., Withers, S. G. & Strynadka, N. C. (2009). *Nature Struct. Mol. Biol.* **16**, 1186–1188.

Rearick, J. I., Sadler, J. E., Paulson, J. C. & Hill, R. L. (1979). *J. Biol. Chem.* **254**, 4444–4451.

Rudolph, M. G., Kelker, M. S., Schneider, T. R., Yeates, T. O., Oseroff, V., Heidary, D. K., Jennings, P. A. & Wilson, I. A. (2003). *Acta Cryst.* D**59**, 290–298.

Schwörer, R. & Schmidt, R. R. (2002). *J. Am. Chem. Soc.* **124**, 1632–1637.

Takashima, S., Tsuji, S. & Tsujimoto, M. (2002). *J. Biol. Chem.* **277**, 45719–45728.

Vonrhein, C., Blanc, E., Roversi, P. & Bricogne, G. (2007). *Methods Mol. Biol.* **364**, 215–230.
Warkentin, M. & Thorne, R. E. (2007). *J. Struct. Funct. Genomics*, **8**, 141–144.
Weijers, C. A., Franssen, M. C. & Visser, G. M. (2008). *Biotechnol. Adv.* **26**, 436–456.
Wernimont, A. & Edwards, A. (2009). *PLoS One*, **4**, e5094.
Winn, M. D. *et al.* (2011). *Acta Cryst.* D**67**, 235–242.
Wlasichuk, K. B., Kashem, M. A., Nikrad, P. V., Bird, P., Jiang, C. & Venot, A. P. (1993). *J. Biol. Chem.* **268**, 13971–13977.
Yang, F., Dauter, Z. & Wlodawer, A. (2000). *Acta Cryst.* D**56**, 959–964.
Zwart, P. H., Afonine, P. V., Grosse-Kunstleve, R. W., Hung, L.-W., Ioerger, T. R., McCoy, A. J., McKee, E., Moriarty, N. W., Read, R. J., Sacchettini, J. C., Sauter, N. K., Storoni, L. C., Terwilliger, T. C. & Adams, P. D. (2008). *Methods Mol. Biol.* **426**, 419–435.