npg

Structural analysis of the sialyltransferase CstII from *Campylobacter jejuni* in complex with a substrate analog

Cecilia P C Chiu[1], Andrew G Watts[2,4], Luke L Lairson[2,4], Michel Gilbert[3], Daniel Lim[1], Warren W Wakarchuk[3], Stephen G Withers[2] & Natalie C J Strynadka[1]

Sialic acid terminates oligosaccharide chains on mammalian and microbial cell surfaces, playing critical roles in recognition and adherence. The enzymes that transfer the sialic acid moiety from cytidine-5′-monophospho-*N*-acetyl-neuraminic acid (CMP-NeuAc) to the terminal positions of these key glycoconjugates are known as sialyltransferases. Despite their important biological roles, little is understood about the mechanism or molecular structure of these membrane-associated enzymes. We report the first structure of a sialyltransferase, that of CstII from *Campylobacter jejuni,* a highly prevalent foodborne pathogen. Our structural, mutagenesis and kinetic data provide support for a novel mode of substrate binding and glycosyl transfer mechanism, including essential roles of a histidine (general base) and two tyrosine residues (coordination of the phosphate leaving group). This work provides a framework for understanding the activity of several sialyltransferases, from bacterial to human, and for the structure-based design of specific inhibitors.

Glycoproteins, glycolipids and polysaccharides are present on the cell surfaces of many organisms and are central molecules in many biological processes. They participate in cell-cell recognition, cell differentiation and various receptor-ligand interactions throughout biology. Many of these biologically active glycans contain an essential nine-carbon sugar that is known as sialic acid, or *N*-acetyl-neuraminic acid (NeuAc). Gangliosides constitute a large group of sialylated glycosphingolipids found in all vertebrate cell types and are the major glycans of the nerve cells as well as the major sialic acid-bearing glycoconjugates in the brain[1]. They act as receptors for microorganisms and bacterial toxins, regulate cell growth and differentiation, and contribute to cell-cell and cell-matrix interactions[2]. The human mucosal pathogen *C. jejuni* has been shown to express variable cell surface carbohydrate mimics of gangliosides that are associated with virulence[3,4]. Terminal oligosaccharides identical to seven different gangliosides have all been found in various *C. jejuni* strains[4]. The molecular mimicry between *C. jejuni* lipo-oligosaccharides (LOS) outer core structures and gangliosides has also been suggested to act as a trigger for autoimmune mechanisms in the development of Guillain-Barré syndrome[5]. The terminal sialic acid residues are the main sources of diversity in these structures.

We have previously identified the genes responsible for the biosynthesis of the ganglioside mimics in *C. jejuni* using a combination of approaches[6,7], thereby identifying a distinct membrane-associated enzyme (CstII from CAZY family GT42; ref. 8) with α-2,3-sialyltransferase activity linking a CMP-NeuAc donor sugar to a β-galactose acceptor sugar (**Fig. 1**). In addition, several versions of CstII from different strains of *C. jejuni* were found. Despite very high sequence identity (97.3%) between two of these enzymes, CstII$_{O:19}$ and CstII$_{OH4384}$, the CstII from strain OH4384 is bifunctional, carrying an additional α-2,8-sialyltransferase activity in which the initially formed α-2,3-linked sialic acid is used as the acceptor substrate in a second transfer step[6].

The essential role of sialylated oligosaccharides in the biology of both pathogen and host suggests that the enzymes involved in their biosynthesis represent potential targets for therapeutic intervention. There has as yet been very little work on fundamental structure-function aspects of the sialyltransferases from any species. This is probably a result of difficulties in expressing these typically membrane-associated proteins in either bacterial or eukaryotic systems. Using designed constructs from which predicted C-terminal membrane association domains have been removed but that retain full activity, we report here the first detailed structure of a sialyltransferase from any source, CstII from *C. jejuni* strain OH4384. The structure of this bifunctional enzyme has been determined in the absence and presence of a sialic acid donor sugar. This work represents a major advance in our understanding of how sialyltransferases catalyze their essential reactions and provides a structural basis for the design of novel antibacterial agents that function through inhibition of their action.

RESULTS
Oligomerization and membrane attachment
A soluble form of CstII was created by deleting a portion of the predicted membrane association domain (a C-terminal 32–amino acid

[1]Department of Biochemistry and Molecular Biology, University of British Columbia, 2146 Health Sciences Mall, Vancouver, British Columbia V6T 1Z3, Canada. [2]Department of Chemistry, University of British Columbia, 2036 Main Mall, Vancouver, British Columbia V6T 1Z1, Canada. [3]Institute for Biological Sciences, National Research Council, Room 3157, 100 Sussex Drive, Ottawa, Ontario K1A OR6, Canada. [4]These authors contributed equally to this work. Correspondence should be addressed to N.S. (natalie@byron.biochem.ubc.ca).
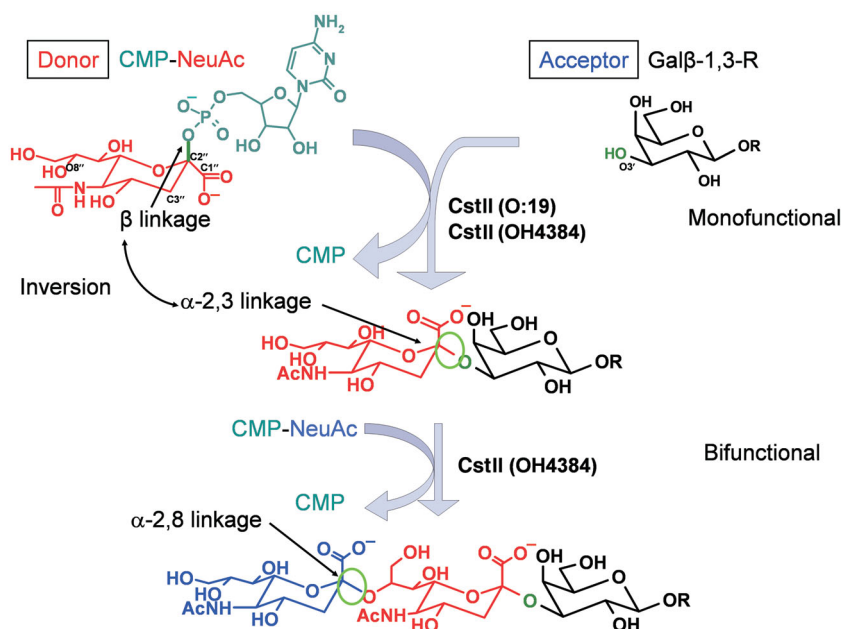
Figure 1 Reaction scheme of CstII. The bond created in the monofunctional and binfunctional reaction is highlighted by a green circle.

there are two molecules in the asymmetric unit, with the four-fold crystallographic axis creating a tetramer similar to that observed in the asymmetric unit of the monoclinic space group (r.m.s. deviation for the 708 Cα atoms is 0.34 Å; **Fig. 2a**). There is approximately 1,400 Å^2 of buried surface area between each pair of monomers in the tetramer, with many stabilizing hydrophobic and hydrogen-bonding interactions. The C termini of the four monomers form an independent folding domain and align on the same side of the tetramer, suggesting a role for the oligomer in effective attachment to the bacterial inner membrane.

Overall architecture

Each monomer of CstIIΔ32 consists of 259 residues organized into two domains (**Fig. 2b**). The first domain (residues 1–154, 189–259) is composed of a mixed α/β fold. The central parallel seven-stranded twisted β-sheet (topology β8, β7, β1, β2, β4, β5, β6) is flanked by five helices on one side (D, E, F, I and J) and four helices (A, B, C and K) on the other. Helix A, the N-terminal region of helix

stretch rich in the basic, hydrophobic and aromatic residues typically observed at membrane-protein interfaces) and by including an I53S mutation that had already been shown to enhance α-2,8-sialyltransferase specificity and stabilize the enzyme[7]. The resulting truncated protein (CstIIΔ32) retains complete activity when compared to the full-length enzyme and was found to form a tetramer in solution by static light scattering analysis (data not shown). Two different crystal forms were produced for complexes with CMP alone (space group P4) and with an inert analog of the donor sugar CMP-3-fluoro-N-acetyl-neuraminic acid (CMP-3FNeuAc; space group P2₁). In the latter, there is a tetramer in the asymmetric unit of the crystal, agreeing with the results from static light scattering in solution. In the P4 space group

B, helix H, helix J and the C-terminal region of helix K adopt a 3₁₀-helical conformation. Helices F, I and part of the β-sheet (β8, β7, β1, β2 and β4) create the nucleotide-binding domain in the form of a Rossmann fold[9]. The binding site of the nucleotide sugar is located in a coil region at the edge of the β-sheet, and is on the same side of the sheet as the C terminus. The active site is not at the interface between monomers of the tetramer, suggesting that the oligomerization of the enzyme has no direct role in catalysis. The second, smaller domain (residues 155–188) is composed of a coil extending from β7, followed by α-helix G and 3₁₀-helix H, then a coil that links back to helix I. This domain forms a lid-like structure that folds over the active site. We observe from our structures that a region of this domain (residues
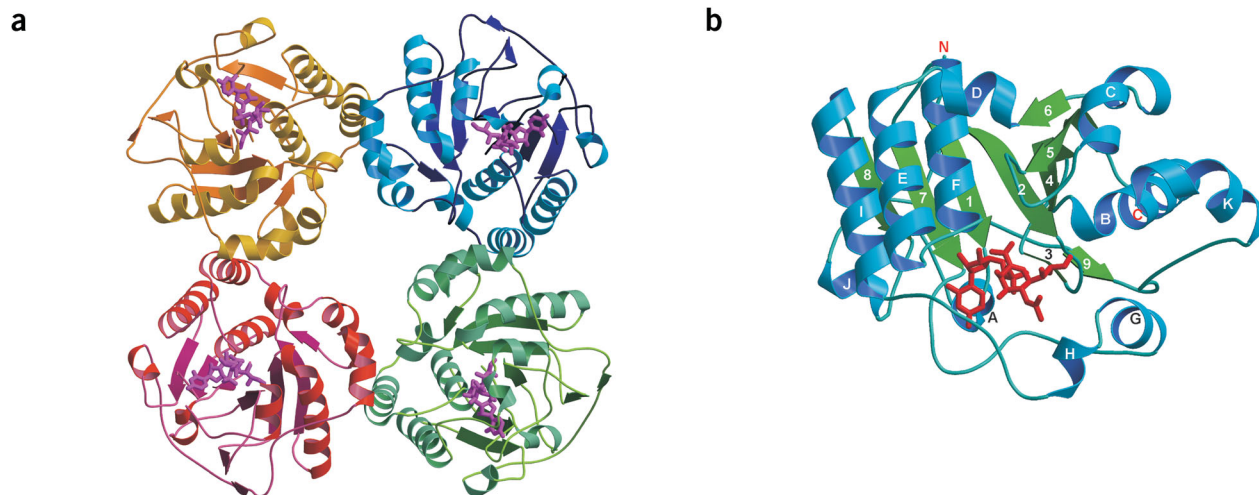
a

b

Figure 2 The overall architecture of CstIIΔ32. (**a**) Arrangement of the CstIIΔ32 tetramer. Each monomer is colored differentially. CMP-3FNeuAc is shown as a stick representation in a magenta color, indicating the location of the catalytic center. (**b**) View of CstIIΔ32 monomer showing the N-terminal domain and the lid domain with bound donor sugar analog. CMP-3FNeuAc is depicted in a red stick representation. The N terminus, C terminus, individual strands and individual helices are labeled.
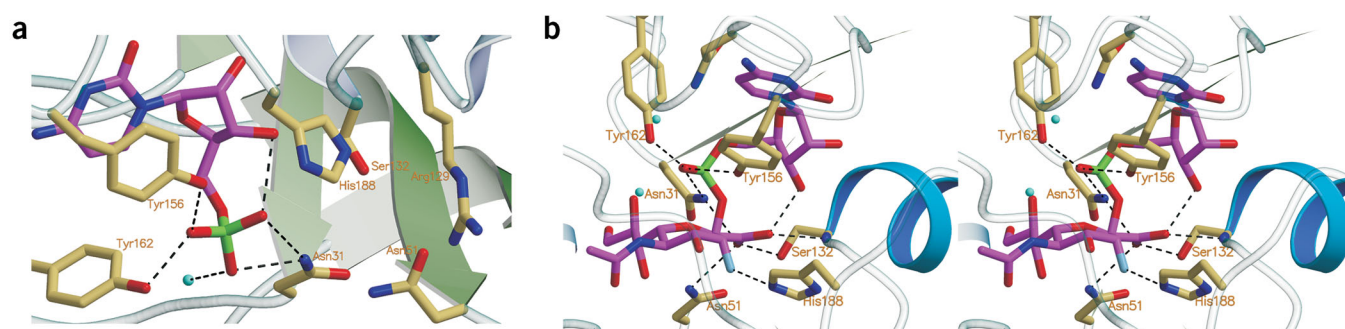
Figure 3 The active site of CstIIΔ32. (**a**) Interactions of CMP and active site residues. CMP is depicted in CPK coloring with carbon atoms in magenta, nitrogen atoms in blue, oxygen atoms in red and phosphorus atom in green. Active site residues involved in CMP binding and catalysis are labeled and shown with carbon atoms in beige, nitrogen in blue and oxygen in red. H_2O molecules are shown as cyan spheres. Dotted lines indicate hydrogen bonding; (**b**) Interactions of CMP-3FNeuAc and key active site residues. CMP-3FNeuAc is depicted in CPK coloring as in **a** with the fluorine atom in light blue. The Helix F is highlighted as a blue ribbon.

175–187) is ordered only upon the binding of the intact CMP-NeuAc substrate.

From the structures published so far, glycosyltransferases have been categorized into two groups based on overall fold. The so-called GTA (glycosyltransferase A) group is composed of α/β proteins with a single Rossmann domain, a conserved 'DXD' motif and a conical active site cleft formed by two closely associated domains[10–12]. The GTB (glycosyltransferase B) group consists of two Rossmann domains separated by a deep substrate-binding cleft[13–15]. Because CstII is composed of a single Rossmann domain, it belongs to the GTA group. However, marked differences of CstII from other members of the GTA group are observed in terms of the connectivity of the secondary structural elements and the absence of the conserved DXD motif. Structural alignment of CstIIΔ32 against all reported glycosyltransferase structures (both inverting and retaining) was performed with the Secondary Structure Matching server from the European Bioinformatics Institute (http://www.ebi.ac.uk/msd-srv/ssm/ssm data/). The closest match was rabbit muscle glycogenin[16] (PDB entry 1LL2), with a r.m.s. deviation of 4.46 Å for 136 equivalent Cα atoms (primarily the Rossmann domains). A general search of the DALI database (http://www.ebi.ac.uk/dali/index.html) indicates that the closest match in the structural database to CstII is the Rossmann fold–containing mouse thiamin pyrophosphokinase[17] (PDB entry 1IG3) with a weak Z score of 6.4 and a r.m.s. deviation of 3.8 Å for only 127 equivalent Cα atom positions.

CMP binding

CstIIΔ32 was first crystallized in the presence of both the donor sugar, CMP-NeuAc, and the acceptor sugar analog, sialyl lactose. However, in the density map only the CMP was present, suggesting that the sialic acid moiety of the CMP-NeuAc had either been transferred or hydrolyzed by the wild-type enzyme or that the sialic acid moiety was disordered in the crystal packing. Indeed, kinetic analysis clearly shows considerable hydrolysis of CMP-NeuAc in the absence of an acceptor catalyzed by both the truncated and the full-length CstII. CMP binds to a deep cleft in the nucleotide-binding domain at the C-terminal end of the central β-sheet. This location of the active site at the protein-membrane interface presumably facilitates transfer of the sugar onto the terminus of the LOS, which is assembled on the cytoplasmic face of the inner membrane and transported onto the cell surface by means of an ABC transporter. CMP is relatively buried within the active site cleft (~630 Å^2 or 68%) with several favorable interactions formed within

the active site (**Fig. 3a**): the cytidine ring is held in position by aromatic stacking with the conserved Tyr156, and the cytidine carbonyl O2 forms hydrogen bonds with main chain nitrogen atoms of Asp154 and Phe155. N4 of the base donates a hydrogen bond to a main chain carbonyl of Ser161. The ribose ring adopts a C2′ exo conformation with the cytidine base axial to the ring. O2′ interacts with the OG1 of Thr131, whereas O3′ interacts with the main chain amide of Ser132. Both O2′ and O3′ interact with the main chain amide of Gly133. O4′ is within hydrogen-bonding distance of the main chain amide of Asn9. The phosphate is stabilized by hydrogen bonds to the hydroxyl groups of Tyr156 and Tyr162 and to ND2 of Asn31, and by a water-mediated hydrogen bond interaction with ND2 of Asn9. Notably, this mode of CMP binding is reminiscent of the ring stacking and phosphate coordination observed for Tyr75 within the AMP binding site of glycogen phosphorylase[18]. An extended loop of CstIIΔ32 (residues 175–187), which presumably lies adjacent to the active site, is highly disordered in the CstIIΔ32-CMP complex.

CMP-3FNeuAc binding

To overcome the problem of hydrolysis of the donor sugar and to allow stable complex formation with acceptor sugars, the inert donor substrate analog CMP-3FNeuAc (**Figs. 3b** and **4a**) was synthesized and used to form cocrystals with the truncated wild type enzyme. Substitution of the hydrogen at the C3′ position on the sialic acid with an electronegative fluorine atom prevents turnover by inductively destabilizing the oxocarbenium ion-like transition state for the reaction[19]. Moreover, it stabilizes the otherwise labile sugar donor against spontaneous hydrolysis. Kinetic and mass spectroscopic analysis showed that CMP-3FNeuAc is not hydrolyzed by our CstIIΔ32 construct in solution (data not shown). The measured K_i value for the binding of this inert sugar to CstIIΔ32 ($K_i = 657$ μM) is comparable to the K_m value for the natural substrate CMP-NeuAc ($K_m = 460$ μM). The near equivalence of these numbers is not surprising given that the fluorine has replaced a hydrogen and that no close interactions with enzyme are seen at that position (**Fig. 3b**). Although there is very little difference in overall structure between the CMP and CMP-3FNeuAc complexes (r.m.s. deviation ~0.2 Å on 246 Cα atoms), there is a substantial local change involving the ordering of residues 175–187, thereby creating an effective lid that closes over and substantially buries the donor sugar. This lid probably fulfills several functions in catalysis; directly liganding the donor sugar, creating the acceptor sugar binding site and shielding the enzyme active site from bulk
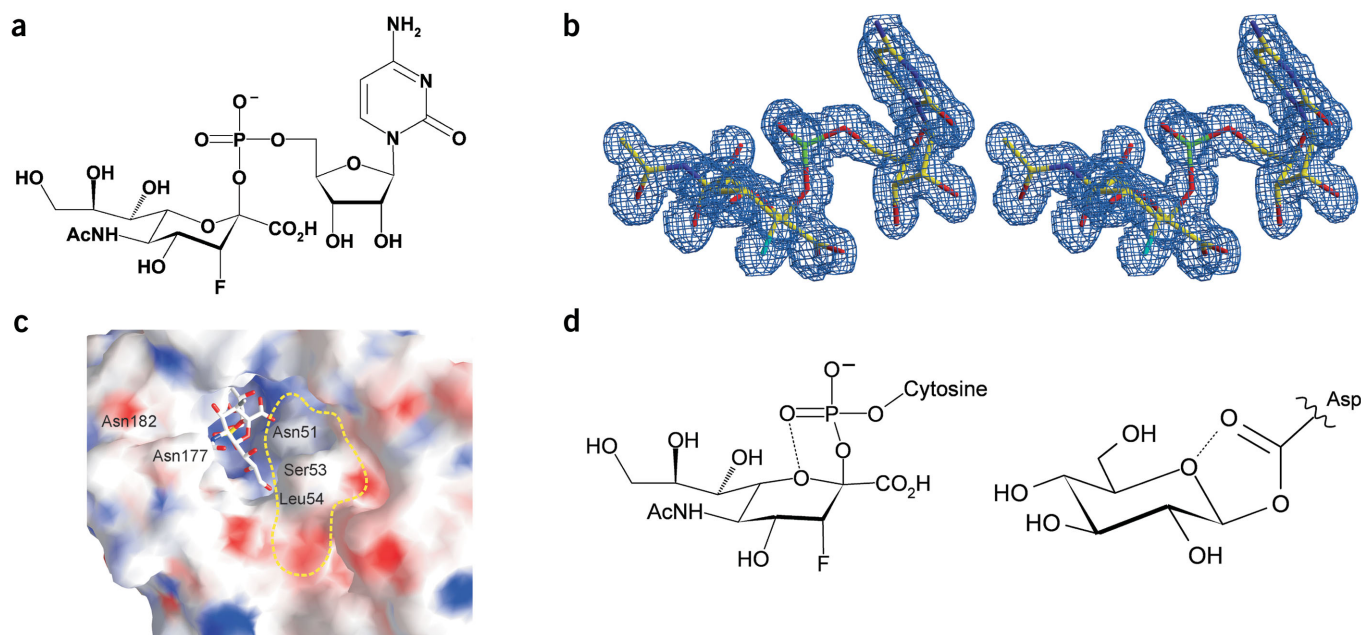
Figure 4 The donor substrate analog CMP-3FNeuAc. (**a**) Chemical drawing of CMP-3FNeuAc. (**b**) Observed electron density of CMP-3FNeuAc in a refined $2F_o - F_c$ map contoured at 1.5σ. (**c**) Molecular surface representation of the active site cleft. The surface is color-coded according to the electrostatic potential, (red, negative potential; blue, positive potential). CMP-3FNeuAc is depicted in stick representation. The yellow dashed-circle indicates the space in the cleft where the acceptor sugars, Gal(β1,3)-GalNAc or NeuAc(α2,3)-Gal(β1,3)-GalNAc, are proposed to bind. Residues in the proximity of the active site that differ between monofunctional (CstII$_{O19}$) and bifunctional (CstII$_{OH4384}$)[6] are labeled. (**d**) Interaction of the ring oxygens of CMP-3FNeuAc with the phosphate oxygen in CstIIΔ32 as observed in this work and of glucose with the carbonyl oxygen of Asp229 in cyclodextrin glucanotransferase[35].

solvent, thereby minimizing undesirable side reactions such as hydrolysis of substrate. Indeed, a similar ordering of the active site has also been observed in other glycosyltransferase systems[11,12].

Well-ordered density defines the CMP-3FNeuAc in the active site of CstIIΔ32 (**Figs. 3b** and **4b**), with the CMP nucleotide moiety binding in a manner very similar to that seen in the CMP complex, despite the presence of the ordered lid domain. The only marked difference is a C3′ endo conformation of the ribose resulting in a more equatorial position of C5′, which consequently positions the remaining chain of phosphate and sialic acid in contact with active site residues.

The sugar ring of the sialic acid, NeuAc, adopts a distorted conformation best described as $^0S_5$ skew boat conformation, with C2″ and the carboxylate group in the same plane as C3″ and the ring oxygen (**Fig. 3b**). Although this conformation of the sugar is of higher energy than the normal chair or boat conformation, it has also been observed in the complex structures of sialic acid with *Trypanosoma rangeli* sialidase[20] (PDB entry 1N1Y) and Influenza B virus neuraminidase[21] (PDB entry 1NSC). The NeuAc carboxylate oxygens in our structure are highly coordinated by neutral amino acid side chains in the enzyme. One oxygen forms hydrogen bonds with the conserved side chains of Asn31, Asn51 and Ser132, and the other oxygen forms hydrogen bonds with the main chain amide of Ser132 as well as with O3′ of the ribose. In addition, the carboxylate resides at the N terminus of the extended α-helix F (initiated by Ser132). This structural feature provides extra electrostatic stabilization of the negatively charged carboxylate through the positive charge arising from the helix dipole. Surprisingly, the observed coordination of the sialic acid carboxylate in our sialyltransferase structure is very different from that observed earlier in sialidase[21–28] and *trans*-sialidase structures[29]. In these enzymes, a highly conserved arginine triad directly coordinates the carboxylate of sialic acid and seems to help

in anchoring the substrate during its distortion along the reaction coordinate.

Other key interactions observed in our CstIIΔ32-CMP-3FNeuAc structure include hydrogen bonds between O4″ and the main chain amide of Tyr185, and between the glycerol moiety of the sugar and the enzyme active site: O7″ with ND2 of Asn51, O8″ and O9″ with each of the side chain atoms of Gln32, and O9″ with NE2 of Gln58. A hydrophobic interaction of Phe178 with the *N*-acetyl group of the sialic acid moiety provides further stabilization, and an ordered water molecule mediates the interaction of the nitrogen of the *N*-acetyl group with OG of Ser183. The introduced substituent F3″ is 2.8 Å away from NE2 of His188. A kinetic analysis of key active site mutants is presented in **Table 1**.

Table 1 Michaelis-Menten parameters for CMP-NeuAc with various CstIIΔ32 mutants

CstIIΔ32	k_{cat} (min^{-1})[a]	K_m (μM)[a]	k_{cat}/K_m (mM^{-1} min^{-1})
WT	43 (± 5)	4.6×10^2 (± 1×10^2)	94
WT hydrolysis[b]	1.7 (± 0.09)	4.7×10^2 (± 0.5×10^2)	3.6
R129A[c]	<0.20 (± 0.01)	2.6×10^2 (± 0.4×10^2)	0.78
H188A[c]	<0.16 (± 0.03)	3.0×10^2 (± 1.1×10^2)	0.53
Y156F[d]	0.58 (± 0.07)	2.2×10^2 (± 0.5×10^2)	2.6
Y162F[d]	0.12 (± 0.01)	4.3×10^2 (± 0.9×10^2)	0.28
Y156F Y162F[e]	<0.040	N.D.	–

[a]Parameters were determined at a constant lactose concentration of 160 mM. Values in parentheses are ± standard error N.D., none detected. [b]For enzyme-catalyzed hydrolysis, kinetic parameters were determined in the absence of acceptor. [c]No significant activity above that of enzyme-catalyzed hydrolysis (measured in the absence of acceptor) was observed; thus kinetic parameters reflect the hydrolytic process, as confirmed by product analysis using TLC and mass spectrometry (data not shown). [d]Because no hydrolytic activity could be detected, the rate represents transferase activity. [e]No activity above that of the considerable rate of spontaneous background hydrolysis was detected up to an enzyme concentration of 500 μg ml^{-1}.

Table 2 Michaelis-Menten parameters for acceptors with wild type CstIIΔ32

Acceptor	k_{cat} (min^{-1})	K_m (mM)[a]	k_{cat} / K_m (mM^{-1} min^{-1})
Lactose	38 (±1)	35 (±3)	1.1
3′-Sialyl lactose	55 (±1)	3.5 (±0.3)	16

[a]Kinetic parameters were determined at a constant CMP-NeuAc concentration of 1 mM. Values in parentheses are ±s.e.

Several of the interactions of the sugar moiety with the protein involve residues of the flexible lid domain that is disordered in the CstIIΔ32 complex with CMP alone. These interactions include electrostatic as well as hydrophobic contacts (**Fig. 3b**), explaining why this loop is disordered in the absence of the sugar moiety. Ordering of the lid domain probably creates the acceptor sugar binding site, an obvious cleft adjacent to the NeuAc in our structure (**Fig. 4c**). Notably, the I53S mutation that promotes stabilization of our CstII construct lies on the periphery of this cleft, where it may stabilize the enzyme through interactions with the acceptor sugar.

A sialyl-binding motif L has been proposed on the basis of sequence analysis of mammalian sialyltransferases and mutagenesis studies in rat liver α-2,6-sialyltransferase[30]. The proposed motif contains eight residues (in rat, Cys181, Val184, Leu190, Gly196, Arg207, Asp219, Val220 and Gly221). However, our structure shows that the majority of these positions would not be predicted to line the sialic acid binding site (Val4, Ala7, Leu13, Ser19, Arg29, Tyr40, Leu41, and Gly42 in CstII). Only Ala7 (Val184 in rat) and Arg29 (Arg207 in rat) are in relative proximity to the active site, but still too far to bind substrate directly (7 and 9 Å from the donor sugar, respectively). Analysis of our structure indicates that at least some of the conserved residues in the sialyl motif L may may have an indirect structural role in substrate binding, potentially aiding in correct positioning or stability of adjacent but less conserved amino acids in the active site cleft. For example, residues 185–187, which follow the conserved Val184 in the rat enzyme, correspond to Gly8, Asn9 and Gly10 in our CstIIΔ32 structure, all of which line the bottom of the sialic acid cleft. Similarly, Phe208–Asn209, which follow the invariant Arg207 in the rat enzyme, would correspond to Cys30 and Asn31 in CstII, both of which line the side of the CMP binding pocket.

Unlike other glycosyltransferases with a GTA fold, no bound metal was observed in the active site of CstIIΔ32. Furthermore, CstII lacks the DXD sequence motif found in a wide range of glycosyltransferases. This motif is known to coordinate the divalent cation involved in binding of the nucleotide sugar through interaction with the diphosphate moiety[31,32]. The metal is generally considered to act as a Lewis acid catalyst in the reaction mechanism by stabilization of the leaving nucleoside diphosphate. Mutagenesis studies of the conserved Asp residues in the DXD motif of glycosyltransferases from various species have shown that the removal of the Asp residues, and hence the metal ions, completely eliminates the transferase activity[31–35]. Although kinetic studies with CstIIΔ32 indicate that both magnesium and manganese enhance the activity slightly(~50% increase for both metal ions), they are not essential for catalysis. Such differences are not surprising given that the donor substrate in this case is a nucleoside monophosphate sugar (CMP-NeuAc), so that no diphosphate moiety is present for metal ion binding. CstII presumably uses other means to activate or stabilize the leaving nucleotide (see later).

DISCUSSION

All structural and kinetic evidence for inverting glycosyltransferases thus far supports a direct displacement mechanism through an oxocarbenium ion-like transition state with general base assistance. Thus CstII presumably catalyzes two such direct displacements. In the first reaction, the 3-hydroxyl of the galactose acts as the nucleophile, directly attacking the anomeric center (C2) of CMP-NeuAc to form NeuAc(α2,3)-Gal(β1,3)-GalNAc with the release of CMP. In the second reaction, the 8-hydroxyl from the sialyl moiety of this newly formed sugar now acts as the nucleophile, attacking the anomeric center of a second CMP-NeuAc bound in the active site (**Fig. 1** and **Table 2**). Key questions about this mechanism, then, are the identities of the general base catalyst that deprotonates the attacking nucleophile and of any acid catalyst that assists CMP departure, as well as how the anticipated oxocarbenium ion-like transition state is stabilized.

The structure of CstIIΔ32 in complex with CMP-3FNeuAc provides some answers. The distorted skew boat conformation of the sialyl moiety places the leaving-group phosphate in a pseudo axial position suitable for departure upon attack of the nucleophilic hydroxyl. Such an axial disposition for the leaving group is a universally observed feature for both transferases and glycosidases at each step in catalysis. Distortion of the sialyl moiety to a skew boat places O6, C2, C1, C3 and C4 in a planar arrangement as required for oxocarbenium ion formation. The leaving phosphate is twisted above the ring plane such that one of the nonbridging oxygens—the pro-*R*—interacts with the ring oxygen 3.0 Å (reminiscent of the interaction of the carbonyl oxygen of the nucleophile Asp229 with the ring oxygen in the glycosyl-enzyme intermediate formed on cyclodextrin glucanotransferase[35]; **Fig. 4d**). In both cases, the breakage of the glycosidic bond could be facilitated by the negative-charge buildup on the leaving-group oxygen, which serves to stabilize the developing oxonium ion character at O6. Departure of the phosphate could also be facilitated by the hydrogen-bonding interactions of Tyr156 and Tyr162 with the nonbridging pro-*S* oxygen. Discrete acid catalysis is not required for transferases, because the second ionization of the leaving-group monophosphate has a pK_a of around neutrality, in contrast to the alcohol leaving group of p$K_a \approx$16 released by glycosidases. Indeed, the situation is again more similar to the second step of a retaining glycosidase, wherein the leaving group is an enzymatic carboxylate. In those cases, tyrosine residues also often act in stabilizing the departing leaving group[36]. Mutational analysis shows that k_{cat} values for the Y156F and Y162F mutants of CstII are 75-fold and 360-fold lower than those for the wild type enzyme, and their simultaneous mutation results in the complete loss of any activity above that of spontaneous hydrolysis (**Table 1**).

The identity of the general base catalyst is unclear. In glycosidases, including sialidases, this role is played by the carboxylate group of a Glu or Asp residue. However, there are no suitably disposed acidic groups in the active site of CstIIΔ32, the closest being Glu57, which is a full 14 Å away. The closest side chains to the anomeric carbon are Asn31 (3.9 Å), Asn51 (3.9 Å), Ser132 (4.6 Å) and His188 (4.8 Å). Of these, the only feasible candidate is His188, which lies appropriately on the α face of the sugar and adjacent to an open cleft in the enzyme, which would be the most obvious site to bind the acceptor sugar (**Fig. 3b**). It is thus well-positioned for the imidazole side chain to act as the general base in catalysis, deprotonating the incoming hydroxyl group of the acceptor galactose during nucleophilic attack at the anomeric carbon of the donor sugar. Despite the lack of overall structural similarity and sequence identity, a comparison of the active site of CstIIΔ32 to that of the inverting human β1,3-glucuronyltransferase (GlcAT-I; PDB entry 1FGG) based on the position of the anomeric centers of the donor sugars shows that His188 is in a position equivalent to Glu281, the proposed catalytic base of GlcAT-I (ref. 37), ~5.6 Å away from the anomeric center. Our kinetic analysis shows the optimal activity of CstII occurs at approximately pH 8. At this pH, the

Table 3 Data collection and refinement statistics

| | Derivative (Se-Met) | | | Complex |
	Peak	Inflection	Remote	CMP-3FNeuAc
Data collection				
Space group	$P4$			$P2_1$
Beamline	X25			X8-C
Resolution (Å)	30–2.1	30–2.3	30–2.3	30–1.8
Wavelength (Å)	0.97912	0.97961	0.97900	1.00000
Total reflections	363,065	98,127	99,539	429,958
Unique reflections	62,730	41,979	42,343	99,748
Redundancy	5.8	2.3	2.4	4.3
Completeness (%)[a]	93.7 (68.8)	88.9 (55.3)	89.7 (58.7)	99.6 (98.5)
$<I / \sigma (I)>$[a]	23.9 (10.6)	26.6 (11.2)	28.3 (11.5)	23.6 (2.0)
R_{sym}[a,b]	5.7 (9.1)	3.1 (6.3)	3.1 (6.4)	5.4 (52.6)
Refinement statistics				
Resolution (Å)	30–2.1			30–1.8
Number of atoms				
Protein	3,948			8,164
Substrate	41			167
Water	89			334
R_{cryst} (%)[c]	21.9			21.9
R_{free} (%)[c]	25.6			25.2
R.m.s. deviations				
Bonds (Å)	0.0062			0.0067
Angles (°)	1.17			1.39
Average B-factor (Å²)				
Protein	39.4			34.7
Substrate	44.2			43.3
Water	35.8			33.0

[a]Highest-resolution shell (2.18–2.10 Å for $P4$ and 1.86–1.80 Å for $P2_1$) statistics are given in parentheses. [b]$R_{sym} = \Sigma(I_{hkl}) - <I> / \Sigma(I_{hkl})$, where I_{hkl} is the integrated intensity of a given reflection. [c]$R_{cryst} = (\Sigma|F_o - F_c|) / (\Sigma F_o)$, where F_o and F_c are observed and calculated structure factors. 5% of total reflections were excluded from the refinement to calculate R_{free}.

imidazole of His188 would be expected to be deprotonated as required for a general base catalyst. Upon proton transfer, the positive charge that develops on the imidazole side chain could be stabilized through electrostatic interactions with the negatively charged phosphate and carboxylate groups of the donor sugar substrate (~5 Å away). Mutation of His188 to alanine results in a loss of all detectable transferase activity, indicating its key catalytic role in the transferase mechanism of CstII (**Table 1**).

Our structure shows a conserved arginine residue (Arg129) that lies ~5.5 Å away from His188 and thus could act to provide an electrostatic shield favoring the deprotonated form of the putative general base His188. At physiological pH, the imidazole side chain may not be completely deprotonated without such assistance depending on its intrinsic pK_a. Arg129 is also suitably positioned within the active site to potentially help in orienting the incoming acceptor sugar. Accordingly, our kinetic analysis shows that mutation of Arg129 to alanine decreases the k_{cat} value by 210-fold and, similarly to His188, that the residual activity results from enzyme-catalyzed hydrolysis (**Table 1**). Notably, rate constants for hydrolysis by both H188A and R129A variants are ten-fold lower than those for the wild type enzyme, implying that the normal catalytic machinery is being recruited for hydrolysis (**Table 1**).

Another possible candidate for the general base catalyst could be the carboxylate moiety of the substrate itself. In support of this contention might be the complete absence of any basic amino acid side chains

from the immediate vicinity of the carboxylate, in contrast to the three arginines seen in sialidases. However, the in-plane conformation of the carboxylate seen here is not consistent, because a role as general base catalyst would require that the carboxylate be bound perpendicularly to the ring plane. Although it is still possible that the conformation would change upon formation of a ternary complex with the acceptor sugar, it is notable that this near-planar conformation of the carboxylate is what has been predicted by *ab initio* calculations and supported by kinetic isotope effect (KIE) experiments to exist during the transition state for the formation of a sialyloxocarbenium ion during the spontaneous hydrolysis of CMP-NeuAc[38]. Additionally, more recent KIE experiments have cast serious doubt on the ability of the carboxylate of CMP-NeuAc to function as an intramolecular base catalyst during spontaneous hydrolysis[39].

This structure of a sialyltransferase in complex with an inert substrate analog provides our first insights into the architecture and mechanism of this important class of enzyme and will be very useful for both understanding the basis of the regiospecificity and acceptor specificity in the various types of sialyltransferases and as a foundation for the design of novel sialyltransferase inhibitors.

METHODS

Cloning, overexpression and purification. Cloning of CstII from *C. jejuni* strain OH4384 was done as described[6]. Primers with sequences 5′-GCAT-TACGCATATGAAGAAAGTTATTATTGC-3′ and 5′-GCATTACGTCGACT-TAATTAATATTTTTTG-3′ were used to remove the C-terminal 32 residues. The PCR product was inserted into pET-28a and pET-41a (Novagen) vectors, transformed into *Escherichia coli* strain BL21(DE3) and grown at 37 °C. Overexpression of the protein was induced with 0.5 mM IPTG when the OD$_{600}$ was ~0.60, and the cells were harvested after overnight incubation at 20 °C. His-tagged CstIIΔ32 was purified with a Ni^{2+}-chelating Sepharose column (Pharmacia). The tag was cleaved at 4 °C overnight with 1:1,000 thrombin (Sigma), leaving three extra residues at the N terminus (Gly-Ser-His). Cleaved product was purified by Mono-Q anion-exchanger (Pharmacia) using a linear gradient of NaCl in 20 mM Tris-HCl buffer, pH 8.3. A tagless CstIIΔ32 construct was expressed in the same way as before but was selectively precipitated with 1.0 M ammonium sulfate to aid in purification. The pellet was resuspended in and dialyzed overnight against 20 mM Tris-HCl, pH 8.3. The dialysate was then applied to a Q-Sepharose column, followed by Mono-Q and Superdex200 columns (Pharmacia). Purified proteins were concentrated to ~12 mg ml⁻¹ for crystallization trials. Selenomethionyl CstIIΔ32 was prepared using described protocols[40].

Crystallization and data collection. Protein mixtures containing 10 mg ml⁻¹ protein, 10 mM MgCl$_2$ and different substrates at 10 mM (for donor, CMP-NeuAc or CMP-3FNeuAc; for acceptor, lactose, 3′-deoxylactose or sialyllactose) were used for hanging-drop vapor diffusion crystallization trials. Tetragonal crystals of His-tagged protein crystallized in 100 mM HEPES, pH 7.5, 10% (w/v) PEG 6000 and 5% (v/v) 2-methyl-2,4-pentanediol (MPD) and monoclinic crystals of nontagged protein crystallized in 100 mM bicine, pH 9.0, 10% (w/v) PEG MME 2000 and 100 mM NaCl. Data were collected at 100 K with 15% (v/v) MPD plus mother liquor as cryoprotectant. The CMP complex crystal ($P4$) has unit cell dimensions $a = b = 115.43$ Å, $c = 41.06$ Å. A multi-wavelength data set was collected at the National Synchrotron Light Source (NSLS), beamline X25, using an ADSC Quantum Q315 CCD detector. CMP-3FNeuAc was synthesized as described[41]. For the CMP-NeuAc complexes ($P2_1$), data sets were collected at NSLS beamline X8-C using an ADSC Quantum Q4R CCD detector. The $P2_1$ crystal has unit cell dimensions $a = 83.66$ Å, $b = 66.03$ Å, $c = 99.17$ Å, $\beta = 94.42°$. Data were processed using DENZO and SCALEPACK[42]. Statistics for data collection and processing are summarized in **Table 3**. We note that cocrystallization and soaking attempts with various acceptor substrate analogs (namely lactose, sialyl lactose and 3′-deoxy-lactose; **Table 2**) have thus far been unsuccessful, although the non-σA-weighted map reveals density in the acceptor cleft with very low occupancy. Analysis of crystal packing suggests that our inability to form a ternary complex of the enzyme may result from a

glutamic acid side chain that protrudes near the acceptor sugar binding site from a neighboring molecule in the crystal.

Structure determination and refinement. The CstIIΔ32 structure was solved by MAD phasing from tetragonal crystals grown with protein containing incorporated SeMet. Se atom positions were determined using CNS 1.1 (ref. 43), initial phases improved by RESOLVE[44] and electron density maps improved with DM[45]. The initial model was built using XTALVIEW[46], and the final model was obtained after iterations of refinement with CNS[43]. The sequence differs from the published sequence at position 53 (I53S). In the His-tagged construct of CstIIΔ32, a second random mutation occurred at position 222 (E222G). This latter mutant did not occur in the non-His-tagged construct, and overlap of the two structures showed that this surface-localized amino acid (on a solvent-exposed disordered loop) had little effect on the structure. Sequence differences were confirmed with DNA sequencing. The refined CstIIΔ32 structure from the P4 crystal (solvent and CMP removed) was used as the starting model for the CMP-NeuAc complex in the P2$_1$ crystal form with AMoRe[47] and refined with CNS 1.1 (ref. 43). The dictionaries for substrates were generated by XPLO2D (ref. 48). Quality of the models was analyzed with PROCHECK[49] (88% in the most favorable region of the Ramachandran plot). Residues 31 and 158, well-ordered residues in the active site, adopt disallowed main chain parameters. **Figures 2** and **3** were prepared with MOLSCRIPT[50], **Figure 4b** with XtalView[46] and **Figure 4c** with GRASP[51]. Figures were rendered with Raster3D (ref. 52).

Site-directed mutagenesis by PCR. CstIIΔ32 mutants were generated using two-stage PCR mutagenesis (primers containing *Nde*I and *Sal*I restriction sites were used to subclone the final PCR product into pET-28a) and using QuikChange site-directed mutagenesis (coding and antisense primers containing a single mutagenic site were used for PCR amplification).

Kinetic assays. Careful attention was required to ensure that rates being determined reflect the transferase activity of CstII. Background hydrolysis of CMP-NeuAc in the absence of enzyme was subtracted from all enzyme-catalyzed activities to arrive at true enzyme-catalyzed turnover rates. CstIIΔ32 was also found to catalyze the hydrolysis of CMP-NeuAc at an appreciable rate, as determined by incubating CstIIΔ32 with CMP-NeuAc in the absence of acceptor (**Table 1**). To certify that the observed hydrolytic activity was not due to a transfer of sialic acid from CMP-NeuAc to an additional molecule of donor to form CMP-NeuAc-NeuAc (a possible result, considering the known bifunctional acceptor specificity), mass spectrometric product analysis was conducted, confirming that no such transfer product formed but rather sialic acid and CMP (data not shown). The quantification of each of these three possible transformations of CMP-NeuAc proved particularly important during the analysis of several CstIIΔ32 mutants. Studies were carried out at 37 °C in HEPES, pH 7.5, containing 0.1% (w/v) BSA and 50 mM KCl. Transferase activity was monitored using a continuous coupled assay analogous to that described[53]. Release of CMP was coupled to the oxidation of NADH (λ = 340 nm, ε = 6.22 mM^{-1} cm^{-1}) using the enzymes nucleoside monophosphate kinase, pyruvate kinase and lactate dehydrogenase. Measures were undertaken to ensure the consumption of contaminating nucleotide phosphates and that a stable rate of the considerable spontaneous hydrolysis of CMP-NeuAc was established before CstIIΔ32 was added to initiate assays. Absorbance measurements were obtained using a Cary 300 UV-vis spectrophotometer equipped with a circulating water bath. Grafit 4.0 (Erithacus Software) was used to calculate kinetic parameters by direct fit of initial rates to the respective equations.

Static light scattering. Static light scattering experiments were done at 25 °C on a Shodex Protein KW-803 column using 50 mM HEPES, pH 7.5, 100 mM NaCl. Refractive index and Mini-dawn light scattering detectors (Wyatt Technology) were calibrated using BSA (Sigma).

Coordinates. Coordinates have been deposited in the Protein Data Bank (accession codes 1RO7 and 1RO8).

1. Vyas, A.A. & Schnaar, R.L. Brain gangliosides: functional ligands for myelin stability and the control of nerve regeneration. *Biochimie* **83**, 677–682 (2001).
2. Lloyd, K.O. & Furukawa, K. Biosynthesis and functions of gangliosides: recent advances. *Glycoconj. J.* **15**, 627–636 (1998).
3. Guerry, P. *et al.* Phase variation of *Campylobacter jejuni* 81–176 lipooligosaccharide affects ganglioside mimicry and invasiveness in vitro. *Infect. Immun.* **70**, 787–793 (2002).
4. Penner, J.L. & Aspinall, G.O. Diversity of lipopolysaccharide structures in *Campylobacter jejuni*. *J. Infect. Dis.* **176** Suppl 2, S135–S138 (1997).
5. Endtz, H.P. *et al.* Molecular characterization of *Campylobacter jejuni* from patients with Guillain-Barre and Miller Fisher syndromes. *J. Clin. Microbiol.* **38**, 2297–2301 (2000).
6. Gilbert, M. *et al.* Biosynthesis of ganglioside mimics in *Campylobacter jejuni* OH4384. Identification of the glycosyltransferase genes, enzymatic synthesis of model compounds, and characterization of nanomole amounts by 600-MHz ^{1}H and ^{13}C NMR analysis. *J. Biol. Chem.* **275**, 3896–3906 (2000).
7. Gilbert, M. *et al.* The genetic bases for the variation in the lipo-oligosaccharide of the mucosal pathogen, *Campylobacter jejuni*. Biosynthesis of sialylated ganglioside mimics in the core oligosaccharide. *J. Biol. Chem.* **277**, 327–337 (2002).
8. Coutinho, P.M., Deleury, E., Davies, G.J. & Henrissat, B. An evolving hierarchical family classification for glycosyltransferases. *J. Mol. Biol.* **328**, 307–317 (2003).
9. Rossmann, M.G. & Argos, P. The taxonomy of binding sites in proteins. *Mol. Cell. Biochem.* **21**, 161–182 (1978).
10. Charnock, S.J. & Davies, G.J. Structure of the nucleotide-diphospho-sugar transferase, SpsA from *Bacillus subtilis*, in native and nucleotide-complexed forms. *Biochemistry* **38**, 6380–6385 (1999).
11. Persson, K. *et al.* Crystal structure of the retaining galactosyltransferase LgtC from *Neisseria meningitidis* in complex with donor and acceptor sugar analogs. *Nat. Struct. Biol.* **8**, 166–175 (2001).
12. Unligil, U.M. *et al.* X-ray crystal structure of rabbit *N*-acetylglucosaminyltransferase I: catalytic mechanism and a new protein superfamily. *EMBO J.* **19**, 5269–5280 (2000).
13. Vrielink, A., Ruger, W., Driessen, H.P. & Freemont, P.S. Crystal structure of the DNA modifying enzyme β-glucosyltransferase in the presence and absence of the substrate uridine diphosphoglucose. *EMBO J.* **13**, 3413–3422 (1994).
14. Ha, S., Walker, D., Shi, Y. & Walker, S. The 1.9 Å crystal structure of *Escherichia coli* MurG, a membrane-associated glycosyltransferase involved in peptidoglycan biosynthesis. *Protein Sci.* **9**, 1045–1052 (2000).
15. Gibson, R.P., Turkenburg, J.P., Charnock, S.J., Lloyd, R. & Davies, G.J. Insights into trehalose synthesis provided by the structure of the retaining glucosyltransferase OtsA. *Chem. Biol.* **9**, 1337–1346 (2002).
16. Gibbons, B.J., Roach, P.J. & Hurley, T.D. Crystal structure of the autocatalytic initiator of glycogen biosynthesis, glycogenin. *J. Mol. Biol.* **319**, 463–477 (2002).
17. Timm, D.E., Liu, J., Baker, L.J. & Harris, R.A. Crystal structure of thiamin pyrophosphokinase. *J. Mol. Biol.* **310**, 195–204 (2001).
18. Stura, E.A. *et al.* Comparison of AMP and NADH binding to glycogen phosphorylase b. *J. Mol. Biol.* **170**, 529–565 (1983).
19. Burkart, M.D. *et al.* Chemo-enzymatic synthesis of fluorinated sugar nucleotide: useful mechanistic probes for glycosyltransferases. *Bioorg. Med. Chem.* **8**, 1937–1946 (2000).
20. Buschiazzo, A. *et al.* Structural basis of sialyltransferase activity in trypanosomal sialidases. *EMBO J.* **19**, 16–24 (2000).
21. Burmeister, W.P., Ruigrok, R.W. & Cusack, S. The 2.2 Å resolution crystal structure of influenza B neuraminidase and its complex with sialic acid. *EMBO J.* **11**, 49–56 (1992).
22. Crennell, S., Garman, E., Laver, G., Vimr, E. & Taylor, G. Crystal structure of *Vibrio cholerae* neuraminidase reveals dual lectin-like domains in addition to the catalytic domain. *Structure* **2**, 535–544 (1994).
23. Crennell, S., Takimoto, T., Portner, A. & Taylor, G. Crystal structure of the multifunctional paramyxovirus hemagglutinin-neuraminidase. *Nat. Struct. Biol.* **7**, 1068–1074 (2000).
24. Crennell, S.J., Garman, E.F., Laver, W.G., Vimr, E.R. & Taylor, G.L. Crystal structure of a bacterial sialidase (from *Salmonella typhimurium* LT2) shows the same fold as an influenza virus neuraminidase. *Proc. Natl. Acad. Sci. USA* **90**, 9852–9856 (1993).
25. Crennell, S.J. *et al.* The structures of *Salmonella typhimurium* LT2 neuraminidase and its complexes with three inhibitors at high resolution. *J. Mol. Biol.* **259**, 264–280 (1996).
26. Gaskell, A., Crennell, S. & Taylor, G. The three domains of a bacterial sialidase: A β-propeller, an immunoglobulin module and a galactose-binding jelly-roll. *Structure* **3**, 1197–1205 (1995).

27. Luo, Y., Li, S.C., Chou, M.Y., Li, Y.T. & Luo, M. The crystal structure of an intramolecular trans-sialidase with a NeuAc[α]2→3Gal specificity. *Structure* **6**, 521–530 (1998).
28. Luo, Y., Li, S.C., Li, Y.T. & Luo, M. The 1.8 Å structures of leech intramolecular trans-sialidase complexes: evidence of its enzymatic mechanism. *J. Mol. Biol.* **285**, 323–332 (1999).
29. Amaya, M.F., Buschiazzo, A., Nguyen, T. & Alzari, P.M. The high resolution structures of free and inhibitor-bound *Trypanosoma rangeli* sialidase and its comparison with *T. cruzi* trans-sialidase. *J. Mol. Biol.* **325**, 773–784 (2003).
30. Datta, A.K. & Paulson, J.C. The sialyltransferase "sialylmotif" participates in binding the donor substrate CMP-NeuAc. *J. Biol. Chem.* **270**, 1497–1500 (1995).
31. Wiggins, C.A.R. & Munro, S. Activity of the yeast MNN1 α-1,3-mannosyltransferase requires a motif conserved in many other families of glycosyltransferases. *Proc. Natl. Acad. Sci. USA* **95**, 7945–7950 (1998).
32. Hagen, F.K., Hazes, B., Raffo, R., deSa, D. & Tabak, L.A. Structure-function analysis of the UDP-*N*-acetyl-D-galactosamine:polypeptide *N*-acetylgalactosaminyltransferase. Essential residues lie in a predicted active site cleft resembling a lactose repressor fold. *J. Biol. Chem.* **274**, 6797–6803 (1999).
33. Busch, C. *et al.* A common motif of eukaryotic glycosyltransferases is essential for the enzyme activity of large clostridial cytotoxins. *J. Biol. Chem.* **273**, 19566–19572 (1998).
34. Shibayama, K., Ohsuka, S., Tanaka, T., Arakawa, Y. & Ohta, M. Conserved structural regions involved in the catalytic mechanism of *Escherichia coli* K-12 WaaO (RfaI). *J. Bacteriol.* **180**, 5313–5318 (1998).
35. Uitdehaag, J.C.M. *et al.* X-ray structures along the reaction pathway of cyclodextrin glycosyltransferase elucidate catalysis in the α-amylase family. *Nat. Struct. Biol.* **6**, 432–436 (1999).
36. Gebler, J.C. *et al.* Substrate-induced inactivation of a crippled β-glucosidase mutant: identification of the labeled amino acid and mutagenic analysis of its role. *Biochemistry* **34**, 14547–14553 (1995).
37. Pedersen, L.C. *et al.* Heparan/chondroitin sulfate biosynthesis. Structure and mechanism of human glucuronyltransferase I. *J. Biol. Chem.* **275**, 34580–34585 (2000).
38. Horenstein, B.A. Quantum mechanical analysis of an α-carboxylate-substituted oxocarbenium ion. Isotope effects for formation of the sialyl cation and the origin of an unusually large ^{14}C secondary isotope effect. *J. Am. Chem. Soc.* **119**, 1101–1107 (1997).
39. Horenstein, B.A. & Bruner, M. The *N*-acetyl neuraminyl oxocarbenium ion is an intermediate in the presence of anionic nucleophiles. *J. Am. Chem. Soc.* **120**, 1357–1362 (1998).
40. Doublie, S. Preparation of selenomethionyl proteins for phase determination. *Methods Enzymol.* **276**, 523–530 (1997).
41. Watts, A.G. & Withers, S.G. The synthesis of some mechanistic probes for sialic acid processing enzymes and the labeling of a sialidase from *Trypanosoma rangeli*. *Can. J. Chem.* (in the press).
42. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
43. Brunger, A.T. *et al.* Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr. A* **54**, 905–921 (1998).
44. Terwilliger, T.C. Maximun likelihood density modification. *Acta Crystallogr. D.* **56**, 965–972 (2000).
45. Cowtan, K. DM: An automated procedure for phase improvement by density modification. *Joint CCP4 and ESF-EACBM Newslett. Protein Crystallogr.* **31**, 34–38 (1994).
46. McRee, D.E. XtalView/Xfit—a versatile program for manipulating atomic coordinates and electron density. *J. Struct. Biol.* **125**, 156–165 (1999).
47. Navaza, J. AMoRe: an automated package for molecular replacement. *Acta Crystallogr A* **50**, 157–163 (1994).
48. Kleywegt, G.J. & Jones, T.A. Model building and refinement practice. *Methods Enzymol.* **277**, 208–230 (1997).
49. Laskowski, R.A., MacArthur, M.W., Moss, D.S. & Thornton, J.M. PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **26**, 283–291 (1993).
50. Kraulis, P. MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.* **24**, 946–950 (1991).
51. Nicholls, A., Sharp, K.A. & Honig, B. Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins* **11**, 281–296 (1991).
52. Merritt, E.A. & Bacon, D.J. Raster3D: Photorealistic molecular graphics. *Methods Enzymol.* **277**, 505–524 (1997).
53. Gosselin, S., Alhussaini, M., Streiff, M.B., Takabayashi, K. & Palcic, M.M. A continuous spectrophotometric assay for glycosyltransferases. *Anal. Biochem.* **220**, 92–97 (1994).