Galactose recognition by the carbohydrate-binding module of a bacterial sialidase

Simon L. Newstead,[a]‡
Jacqueline N. Watson,[b]
Andrew J. Bennet[b] and
Garry Taylor[a]*

[a]Centre for Biomolecular Sciences, University of St Andrews, St Andrews, Fife KY16 9ST, Scotland, and [b]Department of Chemistry, Simon Fraser University, 8888 University Drive, Burnaby, British Columbia V5A 1S6, Canada

‡ Present address: Department of Biochemistry, Imperial College London, South Kensington Campus, London SW7 2AZ, England.

Correspondence e-mail: glt2@st-andrews.ac.uk

Glycoside hydrolases often possess carbohydrate-binding modules (CBMs) in addition to their catalytic domains, which help target the enzymes to appropriate substrates and thereby increase their catalytic efficiency. Sialidases hydrolyse the release of sialic acid from a variety of glycoconjugates and play significant roles in the pathogenesis of a number of important diseases. The sialidase from *Micromonospora viridifaciens* has a CBM which recognizes galactose. The CBM is linked to the catalytic domain by an immunoglobulin-like domain, resulting in the galactose binding site sitting above the catalytic site, suggesting an interplay between the two sites. By studying nine crystallographically independent structures of the *M. viridifaciens* sialidase, the relative flexibility of the three domains was analysed. A detailed study is also presented of the recognition of galactose and lactose by the *M. viridifaciens* CBM. The striking structure of this sialidase suggests a role for the CBM in binding to galactose residues unmasked by the adjacent catalytic site.

1. Introduction

Sialidases, or neuraminidases, are a superfamily of enzymes responsible for the release of sialic acid from a variety of glycoconjugates. These enzymes play significant roles in the pathogenesis of a number of important microbial diseases, such as influenza, mumps, cholera and South American trypanosomiasis (Corfield, 1992). Animals also possess siali-dases; the human and mouse genomes reveal four paralogues each, which are involved in the modulation of glycosylation in events as diverse as cell adhesion and apoptosis (Schauer, 2000). Structural studies on members of the sialidase super-family have revealed a conserved catalytic domain consisting of a six-bladed β-propeller fold, which in the case of the non-viral enzymes is commonly associated with one or more additional domains (Taylor, 1996). The *Trypanosoma cruzi trans*-sialidase and *T. rangeli* sialidase both have a lectin-like domain C-terminal to their catalytic domains (Buschiazzo *et al.*, 2000, 2002), although the nature of the carbohydrate recognized by these domains is unknown. Similar lectin-like domains have been observed in the structures of the sialidases from *Vibrio cholerae* (Crennell *et al.*, 1994) and *Macrodella decora* (Lou *et al.*, 1998). In the case of the *V. cholerae* siali-dase, two structurally identical lectin-like domains flank the central catalytic domain, one of which recognizes sialic acid itself, suggesting a role in targeting the enzyme to sialic acid-rich areas on the surface of cells in the small intestine (Moustafa *et al.*, 2004). It has been suggested that the presence of these carbohydrate-binding domains (CBMs) serves to increase the catalytic efficiency of the sialidases, particularly in

the presence of polysaccharide substrates (Thobhani *et al.*, 2002). A wide range of glycoside hydrolyases have additional CBMs, particularly those involved in degradation of insoluble polysaccharides such as cellulose and starch. Structural details are becoming increasingly available for the current 42 families of CBMs (see http://afmb.cnrs-mrs.fr/CAZY/index.html for

details), which show a great diversity of ligand recognition and folds (Boraston *et al.*, 2004).

The soil bacterium *Micromonospora viridifaciens* secretes two sialidases derived from the same gene. If colominic acid (polysialic acid) is used as an inducer, the organism secretes a 41 kDa sialidase consisting of just the catalytic domain. When carbohydrate-rich milk casein is used as an inducer, the organism secretes a full-length 68 kDa protein (Aisaka *et al.*, 1991). The crystal structure of the 41 kDa form revealed the canonical β-propeller fold for the catalytic domain, whilst the structure of the 68 kDa form revealed two additional domains C-terminal to the catalytic domain (Gaskell *et al.*, 1995) (Fig. 1). The first additional domain has an immunoglobulin-like (Ig-like) fold and acts as a linker to a jelly-roll galactose-binding domain that was first observed in the structure of galactose oxidase from *Dactylium dendroides* (Ito *et al.*, 1994). The catalytic domain (residues 47–402) is connected to the Ig-like domain through two adjacent glycines (402 and 403), suggesting relatively free movement between the domains. However, a disulfide bridge connects the two domains close to this glycine hinge (residues 351–405), limiting such movement. The galactose-recognizing CBM sits with its galactose-binding site poised above the active site of the first domain, suggesting some functional synergy between the two sites. A recent study has investigated the binding of carbohydrates to the *M. viridifaciens* CBM, which was subcloned and expressed for use in UV difference spectroscopy and isothermal titration calorimetry (ITC; Boraston *et al.*, 2003). The UV data gave dissociation constants of ~1 mM for both galactose (D-galac-
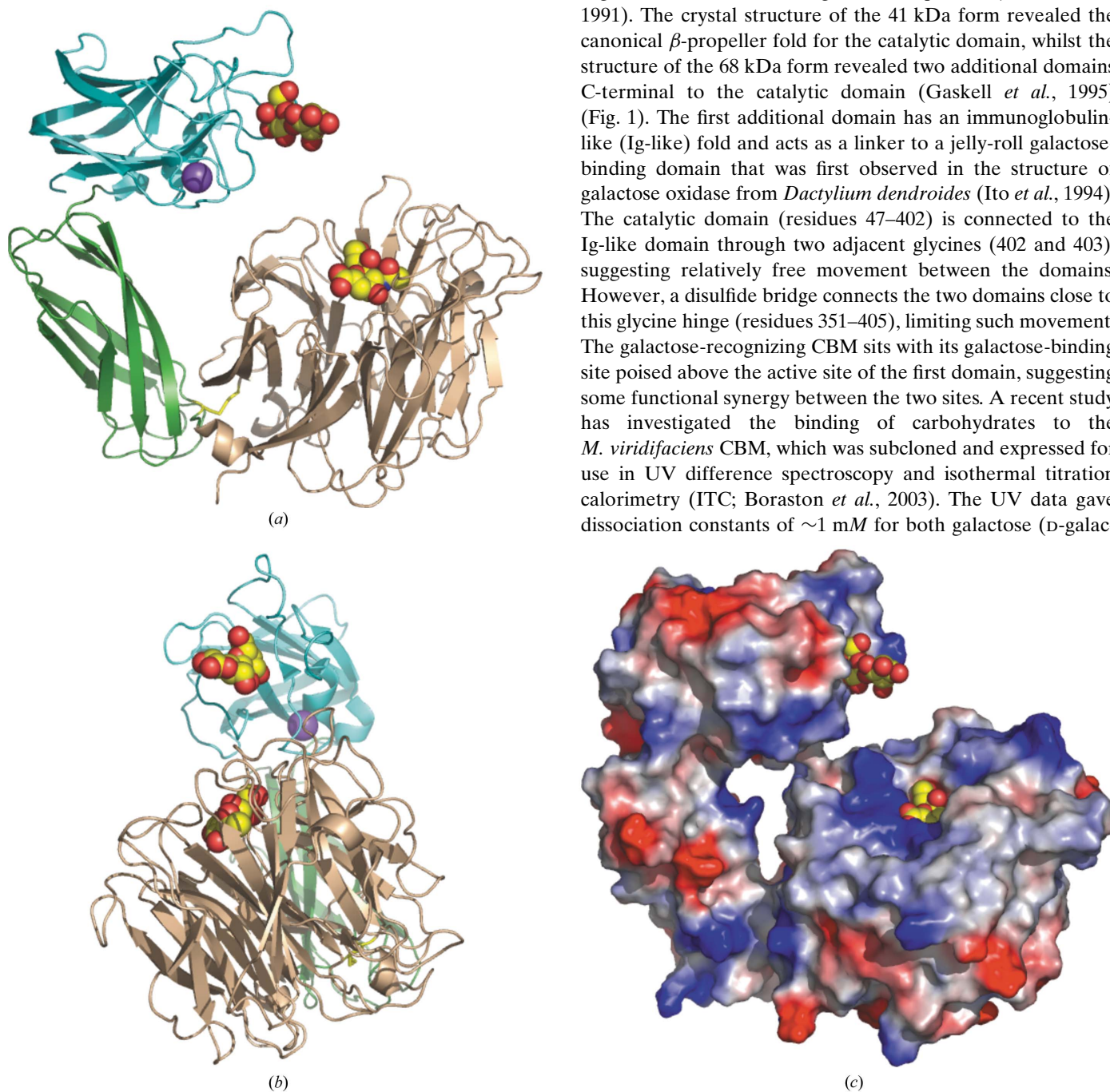

Figure 1
(*a*) and (*b*) Orthogonal views of the *M. viridifaciens* sialidase. The catalytic domain (47–403) is in light brown with Neu5Ac2en bound in the active site, the immunoglobulin-like domain (403–505) is in green and the CBM (505–647) is in cyan with lactose bound. The metal ion is shown as a magenta ball. The disulfide between 351–405 linking the first two domains is also shown. (*c*) Molecular surface coloured according to electrostatic potential; same view as in (*a*).

topyranose) and lactose (4-O-β-D-galactopyranosyl-1-D-glucose), whereas the ITC data gave dissociation constants of 2 and 0.8 mM for galactose and lactose, respectively (Boraston *et al.*, 2003). This study also suggested from UV data that the CBM also recognizes sialic acid. The *M. viridifaciens* galactose-binding CBM has structural similarities to a CBM from *Clostridium stercorarium*, named CsCBM6-3, which recognizes cellulose, xylan and other β-glucans (Boraston *et al.*, 2003), and with the lectin AAA from *Anguilla anguilla*, which recognizes fucose (Bianchet *et al.*, 2002; PDB code 1k12).

The original structural report on the *M. viridifaciens* siali-dase provided limited detail on the galactose interactions with the CBM, owing to incomplete X-ray data collected at room temperature (Gaskell *et al.*, 1995). Here, we report details of complexes with galactose and lactose at 2.1 and 1.7 Å, respectively. These structures reveal the specificity of this CBM for the O4 axial hydroxyl present in the galactose monosaccharide and demonstrate how this domain could bind to any glycoconjugate with an exposed terminal galactose sugar. The nature of the protein–carbohydrate interaction present in this sialidase is remarkably similar to that observed in the AAA lectin from *A. anguilla* and further underlies the 'lectin-like' properties displayed by some members of the CBMs.

In a series of related studies, we have been examining the structures of active-site mutants of *M. viridifaciens* sialidase, which together with the data presented here provide nine crystallographically independent copies of the enzyme. Here, we present a detailed comparison of these structures that shows the extent to which the CBM might move relative to the catalytic domain.

2. Materials and methods

2.1. Crystallization

Full-length 68 kDa *M. viridifaciens* sialidase was expressed and purified as described previously (Watson *et al.*, 2003). The enzyme was concentrated to 10 mg ml^{-1} in 10 mM Tris pH 8.4 and 2 µl of this sample was mixed with an equal volume of mother liquor consisting of 16% PEG 3350, 0.2 M diammonium hydrogen citrate in a sitting-drop vapour-diffusion plate. Large trigonal shaped crystals grew in two to three weeks at 294 K.

2.2. Data collection and structure determination

Crystals were soaked in 10 mM β-D-galactose at room temperature for 1 h before being cryoprotected in mother liquor containing 20% glycerol and flash-frozen in a nitrogen stream at 100 K for data collection. The data-collection statistics are given in Table 1. Diffraction data were collected to 2.0 Å resolution at the ESRF on beamline ID14eh1 fitted with an ADSC Q4R CCD detector system. Data were processed and scaled with *MOSFLM* and *SCALA* from the *CCP*4 (v.5.0.1) suite of programs (Collaborative Computational Project, Number 4, 1994). Several crystal forms of the

Table 1
Data-collection statistics for the *M. viridifaciens* sialidase–galactose complex.

Values in parentheses are for the highest resolution shell.

X-ray source	ID14-2, ESRF
Wavelength (Å)	0.933
Space group	$P3_221$
Unit-cell parameters (Å)	$a = 141.46$, $b = 141.46$, $c = 158.89$
Resolution range (Å)	122–2.0 (2.1–2.0)
Mosaicity (°)	0.5
Unique reflections	121364
Completeness (%)	98.3 (99.6)
Redundancy	5.5 (5.1)
R_{merge}†	0.153 (0.438)
$I/\sigma(I)$	9.9 (2.8)
Refinement statistics	
R_{work}	0.193 (0.246)
R_{free}	0.267 (0.349)
No. of protein atoms	13690
$\langle B_{res}\rangle$‡ for monomers A, B, C (Å^2)	18.1, 17.4, 14.6
No. of galactose atoms	36
$\langle B\rangle$ for monomers A, B, C (Å^2)	60.17, 25.6, 40.61
No. of metal ions	3
$\langle B\rangle$ for monomers A, B, C (Å^2)	20.3, 8.4, 16.9
No. of water atoms	1514
$\langle B\rangle$ (Å^2)	24.95
No. of glycerol atoms	18
$\langle B\rangle$ (Å^2)	29.8
Model quality	
R.m.s.d. bond lengths (Å)	0.016
R.m.s.d. bond angles (°)	1.60

† $R_{merge} = \sum |I(k) - \langle I\rangle|/\sum I(k)$, where $I(k)$ is the value of the kth measurement of the intensity of a reflection, $\langle I\rangle$ is the mean value of the intensity of that reflection and the summation is over all measurements. ‡ B_{res} are the residual B factors of the model.

M. viridifaciens sialidase have been obtained previously: a monoclinic form for a Y370G mutant (Newstead *et al.*, 2005), an orthorhombic form for a D92G mutant (Watson *et al.*, 2004) and a trigonal form for a Y370F mutant (Watson *et al.*, 2005). For the galactose soak, crystals of the active-site mutant (E260A) were grown in conditions similar to the Y370G orthorhombic crystals, yet crystallized in space group $P3_221$, with unit-cell parameters $a = b = 141.46$, $c = 158.89$ Å. A Matthews coefficient (Matthews, 1968) of 2.2 Å^3 Da^{-1} suggested the presence of three molecules in the asymmetric unit with a solvent content of 44%. Molecular replacement using the previously solved monomer of the D92G *M. viridifaciens* sialidase (PDB code 1w8o) was carried out using the program *AMoRe* (Navaza, 1994, 2001) as implemented in the *CCP*4 suite (Collaborative Computational Project, Number 4, 1994). The rotation searches were carried out using data within the resolution range 15–3.0 Å and an integration sphere of 38 Å radius. The translation function readily found three monomers in the asymmetric unit. Following rigid-body fitting of the models in *AMoRe*, the R factor was 47%. Refinement was carried out with *REFMAC* (Murshudov *et al.*, 1997), first carrying out rigid-body refinement of the three domains 47–402, 403–505 and 506–647 within each monomer. TLS and restrained refinement was then carried out, consisting of 15 cycles of TLS refinement with atomic residual isotropic B factors set at 20 Å^2 followed by 20 cycles of coordinate and residual isotropic B-factor refinement. Six TLS groups were defined, representing the three domains of chains A and B

Table 2
Refined TLS parameters for the model of *M. viridifaciens* sialidase.

Orientation and position of the non-intersecting screw axes, together with the rotation and pitch of the motion about these axes.

Group	Screw axis	Direction cosines			Position (Å)			Mean-square $\mathbf{L}$ (deg^2)	Pitch (Å)
1	1	0.757	0.432	0.490	−22.215	58.331	83.146	3.01	3.259
	2	−0.401	0.899	−0.174	−22.011	58.618	83.961	5.56	−2.134
	3	−0.516	−0.065	0.854	−25.834	56.742	81.542	0.88	2.353
2	1	0.810	−0.350	0.471	−2.636	30.461	68.926	6.12	4.768
	2	0.405	0.914	−0.017	−19.207	48.670	77.340	1.07	−10.491
	3	−0.424	0.205	0.882	2.535	49.687	70.929	19.31	−0.931
3	1	−0.580	−0.499	0.643	−15.214	57.479	47.718	3.14	1.304
	2	0.538	0.828	0.157	−13.875	61.615	48.313	6.20	−1.297
	3	0.611	0.255	0.749	−0.151	62.035	40.719	1.05	3.755
4	1	0.870	−0.162	−0.466	57.069	−32.891	38.719	1.16	6.043
	2	−0.016	0.935	−0.354	54.492	−32.318	39.451	2.70	−3.909
	3	0.493	0.315	0.811	55.459	−31.306	40.851	0.68	5.241
5	1	0.843	−0.357	−0.402	39.240	9.999	36.150	8.38	0.541
	2	0.434	0.893	0.116	42.873	6.889	36.563	1.04	−0.255
	3	0.317	−0.272	0.908	27.121	−13.368	35.313	0.74	−5.748
6	1	0.847	−0.229	0.480	27.429	−6.811	38.456	0.78	−10.080
	2	0.076	0.945	0.318	31.173	−8.089	39.404	3.88	2.968
	3	−0.526	−0.232	0.818	22.838	−9.330	44.411	1.42	−2.580

within the asymmetric unit of the crystal. TLS refinement was not performed on chain *C* owing to the appearance of unusually low full *B* values during the analysis of the TLS parameters. The three domains of chains *A* and *B* were chosen based on the information provided by *TLSMD* (Painter & Merritt, 2005) and by assuming that each unique fold could move as a rigid body, with the hinges located between them. The refinement was interspersed with manual inspection and model building in *O* (Jones *et al.*, 1991), with water molecules being located using the automated procedures implemented in *ARP/wARP* (Perrakis *et al.*, 1999). The galactose ligand was manually modelled into the difference electron density visible in the binding site of the C-terminal CBM. There are no residues in disallowed regions of the Ramachandran plot. Molecular images were prepared with *PyMOL* (DeLano, 2002) and *RASTEP* (Merritt & Murphy, 1994).

2.3. Analysis of TLS parameters

The 20 TLS parameters describing the rigid-body motion of each of the six groups were analysed using the program *TLSANL* (Howlin *et al.*, 1993). The TLS parameters were interpreted by representing them as three translations together with three screw displacements along the three mutually perpendicular non-intersecting screw axes that lie parallel to the principal axes of the $\mathbf{L}$ tensor (Schomaker & Trueblood, 1968; Winn *et al.*, 2001). These axes were visualized using the program *PyMOL* and the coordinates output from the program *TLSANL*. In addition to analysing the TLS parameters directly, anisotropic displacement parameters U_{ij} were also calculated for each atom by decomposition of the TLS

tensors and the resulting thermal ellipsoids were visualized using the program *RASTEP* (Merritt & Murphy, 1994).

2.4. Lactose-complex and domain-movement analysis

The complex of the *M. viridifaciens* sialidase with lactose was a serendipitous discovery during an analysis of a D92G active-site mutant of the enzyme (Watson *et al.*, 2004). The enzyme appeared to have picked up lactose during the expression or purification procedures or possibly a longer carbohydrate as no electron density was seen beyond the glucose. Data were collected to 1.7 Å for an orthorhombic crystal form grown under crystallization conditions identical to the galactose complex. Data statistics and structure determination have been described previously (Watson *et al.*, 2004), but no analysis was made of the lactose binding in that study.

The program *ESCET* (Schneider, 2002) was used to make an objective analysis of the conformational variability of the three domains of the *M. viridifaciens* sialidase. Nine structures from five different crystals were used in the analysis, as shown in Table 3.

3. Results and discussion

3.1. Overall structure

Superposition of the bound and unbound CBMs of the *M. viridifaciens* sialidase (residues 502–647) show that the binding site is essentially rigid, with no significant movement occurring upon binding of either galactose or lactose. Least-squares superimposition of the galactose-bound and lactose-bound models with a 1.8 Å resolution structure containing no ligand (PDB code 2ber) gave r.m.s.d. values of 0.32 and 0.29 Å, respectively, for the 146 C$^{\alpha}$ atoms that constitute the CBM.

3.2. Galactose complex

Analysis of the difference electron-density map resulted in the identification of a large positive peak in all three monomers in the asymmetric unit at the site originally identified as binding galactose (Gaskell *et al.*, 1995). The difference density was sufficiently unambiguous to allow the accurate placement of β-D-galactose into the binding sites of the three CBM domains, with the clearest density being in monomer *B*, where upon refinement the average *B* factor for the ligand was 22 Å^2 (Fig. 2*a*). The CBM domain recognizes the galactose ring through only three direct hydrogen bonds to the non-reducing end of the monosaccharide (Fig. 2*c*) and this results in the
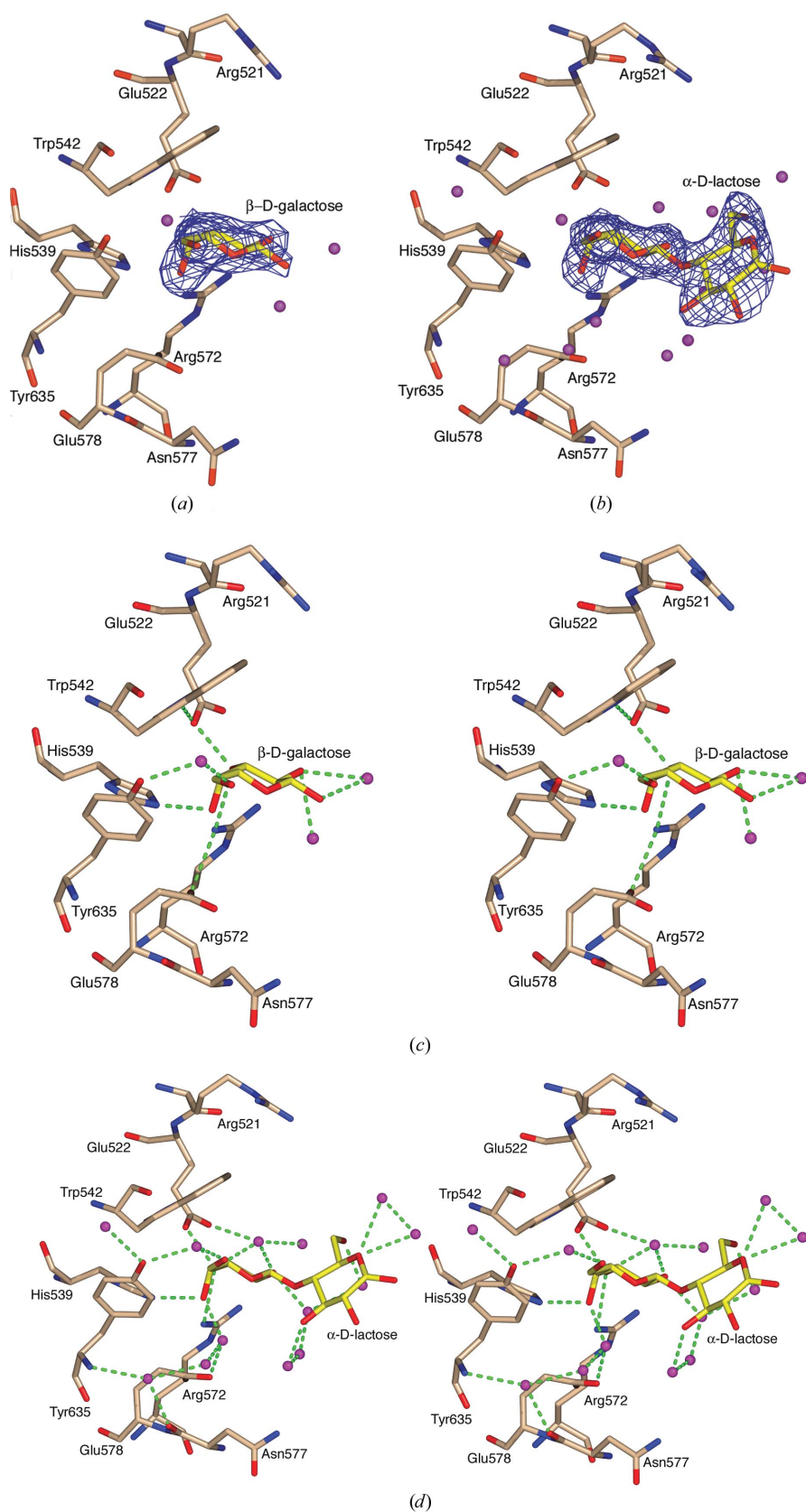
Figure 2
The galactose-binding site of the CBM. (a) and (b) show $2F_{obs} - F_{calc}$ electron-density maps for the galactose and lactose, respectively, contoured at 0.6 e $Å^{-3}$ (approximately 2σ). (c) and (d) show stereoviews of the binding site with galactose and lactose bound, respectively. Hydrogen bonds are drawn in green and waters are shown as magenta spheres.

reducing end of the sugar being exposed to solvent: the 3-OH hydroxyl of galactose hydrogen bonds to Arg572 $N^{\eta 1}$ (2.9 Å) and Glu522 $O^{\varepsilon 2}$ (2.5 Å) and the 4-OH hydroxyl hydrogen bonds to His539 $N^{\varepsilon 2}$ (2.6 Å). There are also four water-mediated hydrogen bonds: the 1-, 2- and 6-hydroxyls all form hydrogen bonds to three water molecules, which in turn hydrogen bond to side chains within the binding pocket.

Another structural feature of this binding site is the stacking interaction between Trp542 and the sugar ring. The C4 atom of the sugar ring is ~3.5 Å from the indole ring. This type of stacking interaction is ubiquitous in CBM protein–carbohydrate interactions (Boraston *et al.*, 2004).

The binding-site topography and ligand conformation strongly suggest that this CBM is specific for galactose. The binding pocket appears to only accommodate the O4 atom in the axial position, with the hydrogen bond to His539. Modelling of the O4 hydroxyl to the equatorial position (glucose) shortens the distance to the atom His539 $N^{\varepsilon 2}$ to ~2.1 Å. The fucose CBM from *A. anguilla* (Bianchet *et al.*, 2002) shares the same topology as the galactose CBM from *M. viridifaciens*; however, the only common interaction is the hydrogen bond between the O4 hydroxyl and the $N^{\varepsilon 2}$ atoms of a histidine. Although galactose and fucose sit in approximately the same position in the CBM-binding site, the fucose is rotated 180° relative to galactose.

3.3. Lactose complex

We previously reported the structure of an active-site point mutant of the *M. viridifaciens* sialidase at 1.7 Å resolution (Watson *et al.*, 2004). Lactose was clearly bound in the galactose CBM, which must have been acquired by the protein during its production, as no exogenous lactose was added (Fig. 2b). The galactose moiety of lactose binds in a manner essentially identical to that of galactose as observed in the galactose complex. The equivalent hydrogen bonds are O3···Arg572 $N^{\eta 1}$ (2.9 Å), O3···Glu522 $O^{\varepsilon 2}$ (2.4 Å) and O4··· His539 $N^{\varepsilon 2}$ (2.7 Å) (Fig. 2d). The remaining hydrogen-bond interactions

are made to a constellation of nine water molecules located within the binding site. Ser575 plays a role in linking this hydrogen-bond network to the glucose unit of the disaccharide.

3.4. Intrinsic flexibility of the *M. viridifaciens* sialidase (displacement of individual domains)

To investigate the spatial relationship between the catalytic and the CBM domains of the *M. viridifaciens* sialidase, a number of different approaches were taken. These included analysing the TLS parameters for chains *A* and *B* in the present model and investigating the different conformations of the protein in nine crystallographically determined models.

Table 2 gives the values associated with the six independent translation and screw motions for each of the six TLS groups refined in this study. These values form an equivalent description of the tensors (Schomaker & Trueblood, 1968; Winn *et al.*, 2001) from the standard **T**, **L** and **S** matrices and

simplify the visualization of the rigid-body motion of the groups. The values indicate that the primary librations modelled by the refinement occur within the immunoglobulin 'arm' domain in both chains *A* and *B*, represented by groups 2 and 5 in Table 2. The non-intersecting screw axes are shown in Fig. 3 for both chains *A* and *B*. The primary librations within the immunoglobulin domain occur along the axis running down the centre of this domain. If we consider that motion about these non-intersecting screw axes is helical, rather than simply rotational in nature, then librations along these axes will result in the displacement of both the catalytic and CBM domains relative to each other. The length of each axis is proportional to the mean-square libration. We believe that the appearance of significant axes in all three TLS groups per chain indicates the likelihood that these domains do indeed behave as rigid bodies *in vivo*. The flexibility of the immunoglobulin domain relative to the other two is also implied by the higher *B* values associated with this group, as shown in Fig. 3.

In order to investigate the dynamics of this model further, nine crystallographically independent models of *M. viridifaciens* sialidase were also analysed using the program *ESCET* (v. 0.7) (Table 3). For each pair of monomers, the *ESCET* program calculates an error-scaled difference distance matrix (Schneider, 2000) that takes into account the different levels of coordinate precision, both for different models and individual atoms, *via* a modified form of Cruickshank's Diffraction Precision Index (DPI; Cruickshank, 1999*a,b*). This modified form of the DPI is used to place the errors in different structures onto a common scale by exploiting the experimentally observed correlation between standard uncertainties (e.s.d.s) and *B* values (Schneider, 2000). Once on a common scale, the significance of the movement can be determined and judgements made about the possible biological implications this may have. Inspection of all pairwise comparisons to establish which conformers were identical according to the criteria used by the *ESCET* program revealed that at the 2σ level (where σ is the uncertainty in the measurement of the difference) there were no consistent sets of identical molecules. This is the default setting used in the *ESCET* program and has been shown to work reasonably well in most cases (Schneider, 2002). The nine models were then subjected to rigid-body analysis as implemented within the *ESCET* program using the genetic algorithm with standard parameters ($n_{hyp} = 20$, $w_p = 20.0$, $\varepsilon_l = 2.0$, $\varepsilon_h = 5.0$, $r_{mut} = 5.0\%$); convergence to homogeneity was reached after 27 genera-
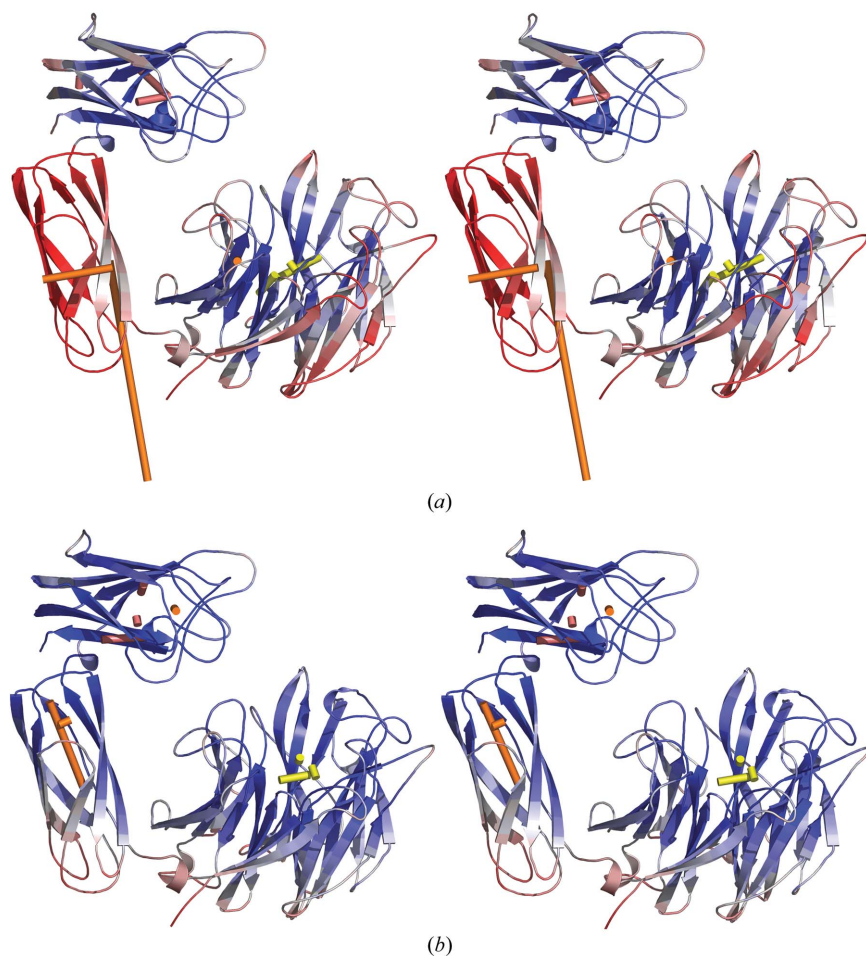

Figure 3
Stereoviews of the TLS-derived non-intersecting screw axes for the *M. viridifaciens* sialidase model superimposed on chains *A* (*a*) and *B* (*b*), which have been coloured according to the full *B* factor output by *TLSANL* from blue (low) to red (high). The three domains of each chain have three axes, which have been coloured accordingly: yellow for the β-barrel domain, orange for the immunoglobulin domain and pink for the CBM domain. The screw axes lie parallel to the libration axes of the TLS group and the length of each axis is proportional to the mean-square libration along the axis.

tions and 158 of 497 C^α atoms were marked as conformationally invariant. Conformationally invariant in this context is defined as the largest subset of atoms for which all interatomic distances are identical, within error, across all models under investigation (Schneider, 2000). The residues selected as rigid

Table 3
Coordinate sets used in the *ESCET* analysis.

Data set	Crystal form	Resolution (Å)	Molecules per AU	PDB code	Reference
Wild type + galactose	Trigonal	2.0	3	2bzd	This study
D92G + lactose	Orthorhombic	1.7	1	1w8o	Watson *et al.* (2004)
D92G + Neu5Ac2en	Orthorhombic	2.1	1	1w8n	Watson *et al.* (2004)
Y370G +β-Neu5Ac	Monoclinic	1.8	1	2ber	Newstead *et al.* (2005)
Y370F + Neu5Ac2en	Trigonal	2.1	3	1wcq	Watson *et al.* (2005)

(A79–A91, A98–A104, A109–A119, A134–A149, A169–A181, A185–A212, A216–A266, A271–A276, A284–A297) are all located within the N-terminal β-propeller domain. For models where TLS refinement was used the full *B* values as output from the program *TLSANL* were used in the *ESCET* analysis.

Superposition of the nine conformers based on the 158 conformationally invariant C^α atoms resulted in mean r.m.s.d. values of between 0.238 and 0.275 Å for the nine structures (Fig. 4*a*). The *ESCET* analysis suggests that the inherent flexibility of the protein originates from the immunoglobulin domain, which acts as a linker allowing the CBM and β-propeller domains to be brought into close proximity to each other and to presumably act in synergy.

Further support for this hypothesis also comes from the analysis of the TLS parameters given in Table 2 and the motion implied by the three non-intersecting screw axes for each of the three domains. We believe the movement of the CBM and β-propeller domains occurs as a result of rotation and translation along the primary screw axes, which are present in the immunoglobulin domains of both chains *A* and *B*. Movement of the immunoglobulin domain along these screw axes would appear to create two 'mechanical hinges' between the three domains of the enzyme (Fig. 4*b*). Intuitively, one would think that the 'hinge' connecting the immunoglobulin domain to the β-propeller would lie orthogonal to the one shown, *i.e.* running from left to right along a line joining the diglycine linker and the disulfide bond. However, it appears that the movement of the immunoglobulin domain is not a straightforward closure motion, but instead appears to resemble that of a screw motion, as implied by the TLS axes and shown graphically in Fig. 3. This causes the immunoglobulin domain to move sideways as well as towards the β-propeller and therefore gives rise to the 'hinge' axis shown. The cause of this sideways motion may be the location of the disulfide bond, which would act to stabilize the immunoglobulin domain in the region close to the diglycine linker. Comparing the nine models of the protein shows that the majority of the
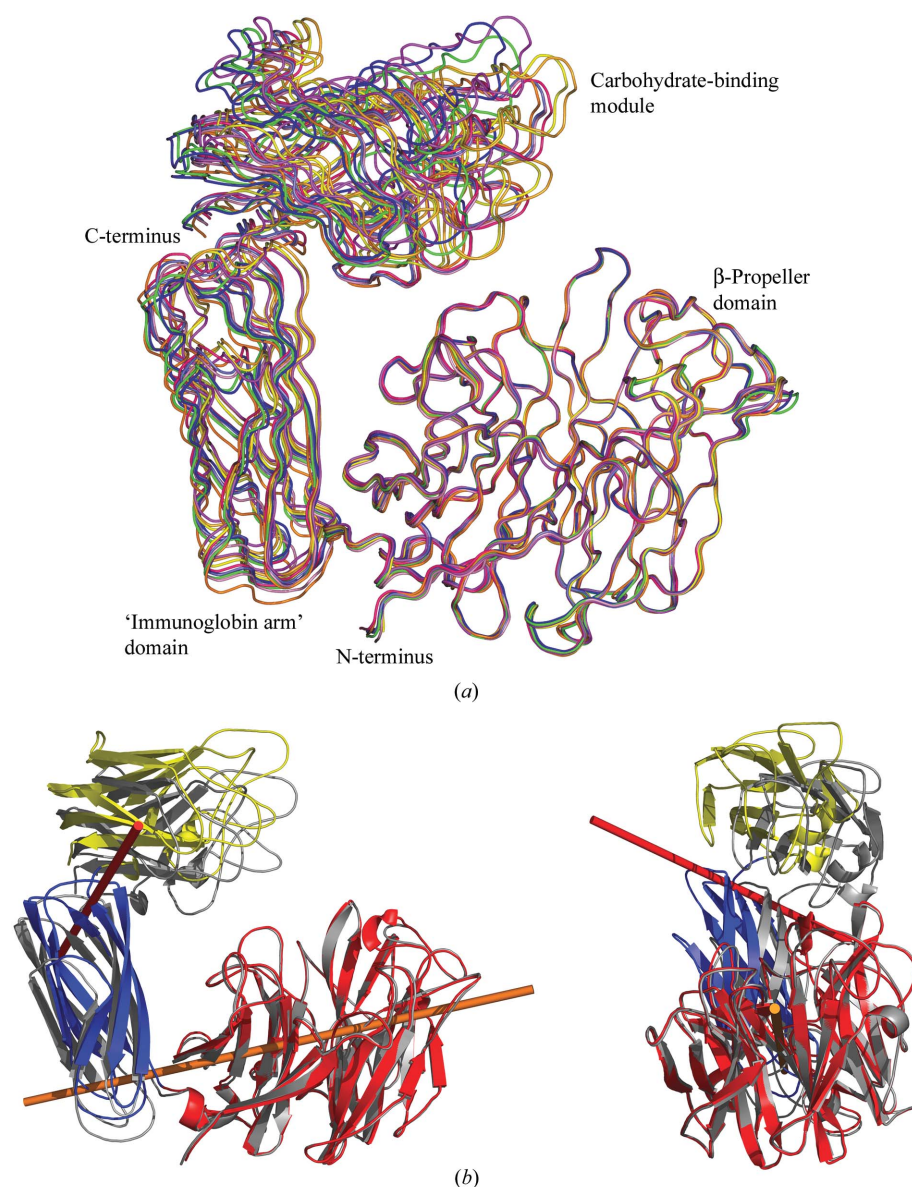

Figure 4
(*a*) Superposition of the C^α traces of the nine models using the 158 C^α atoms identified as being conformationally invariant. (*b*) *DynDom* analysis showing the orientation of the two hinge axes between the domains. The two extreme structures are shown, one coloured grey. The right-hand image represents a 90° rotation of the left-hand image around a vertical axis.

motion occurs in the top half of the immunoglobulin domain.

The identification of these two hinge regions was made using the program *DynDom* (http://www.sys.uea.ac.uk/dyndom/; Berendsen & Hayward, 1998). The two most divergent conformations of the enzyme observed from the crystallographic structures were analysed in this manner. These were represented by the models 1w8n and 1wcq (chain *A*). The program determined that the protein had two hinge axes formed by residues 404–405 and 500–505, which act as two mechanical hinges linking the three domains together.

The results from the comparison show that the C-terminal CBM can rotate at least 12.9° and translate 0.6 Å relative to the immunoglobulin-like domain. The axis along which the CBM moves is illustrated in Fig. 3(*b*) as a red cylinder. The second hinge, connecting the β-propeller to the immuno-globulin arm, causes a smaller rotation of 7.6° and a transla-tion of 0.3 Å relative to the β-propeller domain. This movement results in the CBM domain moving approximately 7.5 Å closer to the active site.

4. Conclusions

The CBM of *M. viridifaciens* sialidase recognizes galactose through direct hydrogen-bond interactions of the O3 hydroxyl with the side chains of Glu522 and Arg572 and of the O4 hydroxyl with the side chain of His539. In addition, the O2 hydroxyl forms a water-mediated hydrogen-bond interaction with the hydroxyl of Tyr635. Further recognition is provided through van der Waals interactions of the sugar ring with Trp542. These interactions make the CBM specific for galac-tose. The complex with lactose shows that there is no direct recognition of the glucose moiety of lactose by the CBM. Galactose and lactose bind the CBM with similar dissociation constants (∼1 m*M*; Boraston *et al.*, 2003) and this is consistent with the structural data presented here. Although the glucose moiety of lactose makes several water-mediated interactions within the CBM and with the galactose, presumably any enthalpic gain made through these hydrogen bonds is balanced by the entropic cost of immobilizing these waters.

Analysis of the nine *M. viridifaciens* sialdiase structures combined with the TLS parameters for the model presented here reveals the flexibility of the CBM relative to the catalytic domain. The consecutive glycines at positions 402 and 403 connecting the catalytic and immunoglobulin-like domains suggest major flexibility at this point; however, a nearby disulfide bond joining cysteines 351 and 405 ensures that the flexibility at this 'hinge' is severely restricted. The analysis shows that within the structures observed to date, the CBM can move by approximately 8 Å.

The results from the TLS parameters show that the primary libration in both chains *A* and *B* is along the non-intersecting screw axis that runs down the centre of the immunoglobulin domains. Movement along these axes is helical, as opposed to rotational, and would act to bring the CBM and immuno-globulin domains closer to the catalytic domain through movement either to the right or left of this libration axis, *i.e.*

not simply a hinge motion closing the domains together along an axis running between the diglycine linker and the disulfide bridge. This would place the effective hinge axis running down the centre of the protein chain, as illustrated in Fig. 4(*b*) and predicted by the *DynDom* program. This predicted movement is also supported by the results from the *ESCET* comparison of the nine crystallographic models. Comparing the two most divergent models shows that the movement is indeed along the primary libration axes calculated from the TLS parameters (Fig. 3). These results support the use of TLS parameters and their associated axes in aiding the understanding of protein dynamics and in particular rigid-body movement within the crystal lattice.

Weak interactions between proteins and carbohydrates are often compensated for by the phenomenon of avidity resulting from multivalent interactions (Liang *et al.*, 1997; Boraston *et al.*, 2004). In these cases, multiple clustered carbohydrate-binding sites interact simultaneously with the carbohydrate ligands, which present multiple recognition elements. This type of interaction results in increased association constants relative to any one of the isolated carbohydrate-binding sites (Kiessling *et al.*, 2000). A recent study investigating the role of multivalency in the mode of action of bacterial sialidases concluded that the extra CBM domains associated with the catalytic β-propeller have dramatic effects on the K_m values for polyvalent substrates compared with their monovalent counterparts (Thobhani *et al.*, 2002). It appears that the extra CBM domains associated with bacterial sialidases have increased the catalytic specificity for polyvalent sialic acid-containing glycoconjugates.

Hydrolysis of sialic acid from cell-surface glycoconjugates will often result in exposure of underlying galactose residues. We can therefore speculate that the presence of the galactose-specific CBM in the *M. viridifaciens* sialidase and its proximity to the catalytic domain confers some functional and catalytic advantage in the environment in which the enzyme needs to operate.

References

Aisaka, K., Igarashi, A. & Uwajima, T. (1991). *Agric. Biol. Chem.* **55**, 997–1004.
Berendsen, H. J. & Hayward, S. (1998). *Proteins*, **30**, 144–154.
Bianchet, M. A., Odom, E. W., Vasta, G. R. & Amzel, L. M. (2002). *Nature Struct. Biol.* **9**, 628–634.
Boraston, A. B., Bolam, D. N., Gilbert, H. J. & Davies, G. J. (2004). *Biochem. J.* **382**, 769–781.
Boraston, A. B., Notenboom, V., Warren, R. A., Kilburn, D. G., Rose, D. R. & Davies, G. (2003). *J. Mol. Biol.* **327**, 659–669.
Buschiazzo, A., Amaya, M. F., Cremona, M. L., Frasch, A. C. & Alzari, P. M. (2002). *Mol. Cell*, **10**, 757–768.
Buschiazzo, A., Tavares, G. A., Campetella, O., Spinelli, S., Cremona, M. L., Paris, G., Amaya, M. F., Frasch, A. C. & Alzari, P. M. (2000). *EMBO J.* **19**, 16–24.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Corfield, A. P. (1992). *Glycobiology*, **2**, 509–521.

Crennell, S. J., Garman, E., Laver, G., Vimr, E. R. & Taylor, G. (1994). *Structure*, **2**, 535–544.

Cruickshank, D. W. (1999a). *Acta Cryst.* D**55**, 583–601.

Cruickshank, D. W. (1999b). *Acta Cryst.* D**55**, 1108.

DeLano, W. L. (2002). *PyMOL*. http://www.pymol.org.

Gaskell, A., Crennell, S. J. & Taylor, G. (1995). *Structure*, **3**, 1197–1205.

Howlin, B., Butler, S. A., Moss, D. S., Harris, G. W. & Driessen, H. P. C. (1993). *J. Appl. Cryst.* **26**, 622–624.

Ito, N., Phillips, S. E., Yadav, K. D. & Knowles, P. F. (1994). *J. Mol. Biol.* **238**, 794–814.

Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. (1991). *Acta Cryst.* A**47**, 110–119.

Kiessling, L. L., Gestwicki, J. E. & Strong, L. E. (2000). *Curr. Opin. Chem. Biol.* **4**, 696–703.

Liang, R., Loebach, J., Horan, N., Ge, M., Thompson, C., Yan, L. & Kahne, D. (1997). *Proc. Natl Acad. Sci. USA*, **94**, 10554–10559.

Lou, Y., Li, S.-C., Chou, M.-Y., Li, Y.-T. & Lou, M. (1998). *Structure*, **6**, 521–530.

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

Merritt, E. A. & Murphy, M. E. P. (1994). *Acta Cryst.* D**50**, 869–873.

Moustafa, I., Connaris, H., Taylor, M., Zaitsev, V., Wilson, J. C., Kiefel, M. J., von Itzstein, M. & Taylor, G. (2004). *J. Biol. Chem.* **279**, 40819–40826.

Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). *Acta Cryst.* D**53**, 240–255.

Navaza, J. (1994). *Acta Cryst.* A**50**, 157–163.

Navaza, J. (2001). *Acta Cryst.* D**57**, 1367–1372.

Newstead, S. L., Watson, J. N., Knoll, T. L., Bennet, A. J. & Taylor, G. (2005). *Biochemistry*, **44**, 9117–9122.

Painter, J. & Merritt, E. A. (2005). *Acta Cryst.* D**61**, 465–471.

Perrakis, A., Morris, R. & Lamzin, V. S. (1999). *Nature Struct. Biol.* **6**, 458–463.

Schauer, R. (2000). *Glycoconj. J.* **17**, 485–499.

Schneider, T. R. (2000). *Acta Cryst.* D**56**, 714–721.

Schneider, T. R. (2002). *Acta Cryst.* D**58**, 195–208.

Schomaker, V. & Trueblood, K. N. (1968). *Acta Cryst.* B**24**, 63–76.

Taylor, G. (1996). *Curr. Opin. Struct. Biol.* **6**, 830–837.

Thobhani, S., Ember, B., Siriwardena, A. & Boons, G. J. (2002). *J. Am. Chem. Soc.* **125**, 7154–7155.

Watson, J. N., Dookhun, V., Borgford, T. J. & Bennet, A. J. (2003). *Biochemistry*, **42**, 12682–12690.

Watson, J. N., Newstead, S. L., Dookhun, V., Taylor, G. & Bennet, A. J. (2004). *FEBS Lett.* **577**, 265–269.

Watson, J. N., Newstead, S. L., Narine, A., Taylor, G. & Bennet, A. J. (2005). In the press.

Winn, M. D., Isupov, M. N. & Murshudov, G. N. (2001). *Acta Cryst.* D**57**, 122–133.