Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Sugar-complex structures of the C-half domain of the galactose-binding lectin EW29 from the earthworm *Lumbricus terrestris*

Ryuichiro Suzuki,[a,b] Atsushi Kuno,[b,c] Tsunemi Hasegawa,[b] Jun Hirabayashi,[c] Ken-ichi Kasai,[d] Mitsuru Momma[a] and Zui Fujimoto[a]*

[a]Protein Research Unit, National Institute of Agrobiological Sciences, Tsukuba 305-8602, Japan, [b]Department of Material and Biological Chemistry, Faculty of Science, Yamagata University, Yamagata 990-8560, Japan, [c]Research Center for Medical Glycoscience, National Institute of Advanced Industrial Science and Technology, Tsukuba 305-8568, Japan, and [d]Department of Biological Chemistry, Faculty of Pharmaceutical Sciences, Teikyo University, Sagamihara 199-0195, Japan

Correspondence e-mail: zui@affrc.go.jp

Received 12 September 2008
Accepted 12 November 2008

PDB References: rC-half, complex with Lac, 2zqn, r2zqnsf; complex with GalNAc, 2zqo, r2zqosf.

R-type lectins are one of the most prominent types of lectin; they exist ubiquitously in nature and mainly bind to the galactose unit of sugar chains. The galactose-binding lectin EW29 from the earthworm *Lumbricus terrestris* belongs to the R-type lectin family as represented by the plant lectin ricin. It shows haemagglutination activity and is composed of a single peptide chain that includes two homologous domains: N-terminal and C-terminal domains. A truncated mutant of EW29 comprising the C-terminal domain (rC-half) has haemagglutination activity by itself. In order to clarify how rC-half recognizes ligands and shows haemagglutination activity, X-ray crystal structures of rC-half in complex with D-lactose and *N*-acetyl-D-galactosamine have been determined. The structure of rC-half is similar to that of the ricin B chain and consists of a β-trefoil fold; the fold is further divided into three similar subdomains referred to as subdomains α, β and γ, which are gathered around the pseudo-threefold axis. The structures of sugar complexes demonstrated that subdomains α and γ of rC-half bind terminal galactosyl and *N*-acetylgalactosaminyl glycans. The sugar-binding properties are common to both ligands in both subdomains and are quite similar to those of ricin B chain–lactose complexes. These results indicate that the C-terminal domain of EW29 uses these two galactose-binding sites for its function as a single-domain-type haemagglutinin.

1. Introduction

Carbohydrate-binding proteins, defined as lectins, with haemagglutination activity exist ubiquitously in nature. Ricin, one of the best known examples of a lectin, is a cytotoxic heterodimeric protein that is found in the seeds of the castor bean plant *Ricinus communis* and belongs to the type II ribosome-inactivating protein family (Hartley & Lord, 2004). It is formed by two polypeptide chains, the A chain and B chain, which are linked by a single disulfide bond. Ricin toxin A chain is a highly specific *N*-glycosidase that catalyzes the hydrolysis of an adenosine residue of 28S ribosomal RNA (Endo & Tsurugi, 1987). Ricin toxin B chain (RTB) is known to be a lectin with hemagglutination activity that binds to sugar chains on the cell surface that contain galactose or *N*-acetylgalactosamine residues (Olsnes & Pihl, 1973). The crystal structure of ricin revealed that RTB had two distinct globular domains (RTB1 and RTB2) with identical folding topologies which were arranged in tandem (Montfort *et al.*, 1987). Each domain can be further divided into three subdomains (α, β and γ) with a fold consisting of two two-stranded hairpins and involving a representative Q*X*W motif.

The three subdomains assemble around a pseudo-threefold axis in a similar manner to the β-trefoil fold (Rutenber & Robertus, 1991; Murzin *et al.*, 1992). The crystal structure of ricin demonstrated that two of the subdomains (1α and 2γ) retain a major sugar-binding site for galactose (Rutenber & Robertus, 1991); subdomain 1β may also have minor sugar-binding activity (Frankel *et al.*, 1996). The multiple sugar-binding sites enable RTB to bind two sugar chains from different cells, resulting in haemagglutination activity.

Lectins with a $(QXW)_3$ motif have been defined as R-type lectins, as described previously (Rutenber & Robertus, 1991; Villafranca & Robertus, 1981; Rutenber *et al.*, 1987; Hazes, 1996), and are found in animals, plants and bacteria. Based on function, the R-type lectin family has been further classified into a lectin group and an enzyme group (Hirabayashi *et al.*, 1998). The lectin group consists of ricin-type lectins, in which two β-trefoil domains are arranged in tandem. To date, several crystal structures of the lectin group have been determined, including those of abrin (Tahirov *et al.*, 1995), ebulin (Pascal *et al.*, 2001), HA1 (Inoue *et al.*, 2003), HA33 (Arndt *et al.*, 2005), CDT (Nešić *et al.*, 2004) and CEL-III (Uchida *et al.*, 2004; Hatakeyama *et al.*, 2007). The main ligand of the lectin group is galactose (Gal) and/or *N*-acetylgalactosamine (GalNAc) (the R-type lectins were formerly known as Gal/GalNAc-binding lectins), but some proteins have different binding preferences; for example, *Sambucus sieboldina* agglutinin (SSA) binds to sialic acid (Neu5Ac; Kaku *et al.*, 1996, 2007). In contrast, the enzyme-group proteins usually consist of a single chain with a multiple-domain architecture containing a catalytic domain and a single R-type lectin fold domain. In the enzyme group, the role of the R-type lectin domain has been revealed for some proteins. The R-type lectin domains in *Streptomyces olivaceoviridis* E-86 xylanase *So*CBM13 (Fujimoto *et al.*, 2000, 2002) and *S. lividans* xylanase *Sl*CBM13 (Notenboom *et al.*, 2002) play a role in substrate binding, showing specificity for xylooligosaccharides. The polypeptide α-*N*-acetylgalactosaminyltransferase (ppGalNAc-T), an important glycosyl transferase that initiates the biosynthesis of O-linked glycan, also has a R-type lectin domain (Fritz *et al.*, 2004; Kubota *et al.*, 2006). Since these domains act as a substrate-binding domain in enzymes, the R-type lectin family constitutes a 13th member of the carbohydrate-binding module (CBM) family as defined in the Carbohydrate-Active Enzyme Database (http://www.cazy.org/).

EW29 is a galactose-binding lectin from the earthworm *Lumbricus terrestris* (Hirabayashi *et al.*, 1998). EW29 is composed of two homologous domains (14.5 kDa each, 27% identity), an N-terminal domain (N-half) and C-terminal domain (C-half), that are arranged into a tandem-repeat structure like RTB; it belongs to the R-type lectin family based on amino-acid sequence. It has significant haemagglutination activity, which is inhibited by a wide range of galactose-containing saccharides. When each of the domains was separately expressed in *Escherichia coli*, the C-terminal domain (rC-half) was found to bind to an asialofetuin–agarose column as strongly as the whole protein, whereas the N-terminal domain did not bind. Further biochemical experiments showed that rC-half shows haemagglutination activity by itself, indicating that rC-half functions as a single-domain-type haemagglutinin with at least two sugar-binding sites.

Unfortunately, the role of EW29 is unclear. Our research has been focused on elucidation of the biological and biochemical functions of EW29. In order to obtain the structural basis of lectin–sugar interactions in EW29, determination of the crystal structure of rC-half in complex with its typical ligands was required. The three-dimensional structure of rC-half might provide insight into the structure–function relationship and the substrate-recognition mechanism of the R-type lectin family. Therefore, X-ray crystallographic and NMR studies of rC-half have been conducted (Suzuki *et al.*, 2004; Hemmi *et al.*, 2004). Recently, we created a mutant of rC-half that showed novel sugar-binding specificity for Neu5Ac by means of reinforced ribosome display (Yabe *et al.*, 2007). The altered sugar-binding specificity was discussed with regard to the structural information presented. Here, we describe the crystal structures of rC-half of EW29 in complex with its main ligands, *i.e.* lactose (Lac) or GalNAc, and reveal the sugar-binding characteristics of two sugar-binding sites in subdomains α and γ.

2. Materials and methods

2.1. Protein preparation and crystallization

The overexpression vector for recombinant C-half (rC-half) has been constructed previously (Hirabayashi *et al.*, 1998). The plasmid was transformed into *E. coli* BL21 (DE3) strain. The cell lysate of the transformant was subjected to lactose-Sepharose affinity chromatography (Ito *et al.*, 2004). The purified protein solution was desalted using a PD-10 desalting column (GE Healthcare, Buckinghamshire, England) and then concentrated with Centriprep (Millipore Corporation, Massachusetts, USA) to a protein concentration of approximately 20 mg ml^{-1}. Crystallization of the rC-half–Lac complex was performed using the hanging-drop vapour-diffusion method according to previously reported procedures (Suzuki *et al.*, 2004). A droplet composed of 5 µl protein solution containing 15 mg ml^{-1} lactose and 5 µl reservoir solution (0.2 *M* sodium chloride, 1.6 *M* dipotassium hydrogen phosphate, 0.4 *M* sodium dihydrogen phosphate, 0.1 *M* imidazole buffer pH 8.0) was mixed and equilibrated against 1 ml reservoir solution at 293 K. Rod-shaped crystals of rC-half–Lac were obtained within one week.

For the rC-half–GalNAc complex, crystallization trials were performed by the sitting-drop vapour-diffusion method using the same precipitant condition as used for the rC-half–Lac complex; tiny needle-shaped crystals appeared within a few days. Further additive screening was performed and the addition of cadmium ions improved the size and quality of the crystals. Finally, 5 µl protein solution containing 15 mg ml^{-1} GalNAc, 5 µl reservoir solution (0.2 *M* sodium chloride, 1.6 *M* dipotassium hydrogen phosphate, 0.4 *M* sodium dihydrogen phosphate, 0.1 *M* imidazole buffer pH 8.0) and 0.1 µl 0.1 m*M* cadmium chloride solution were mixed and equilibrated

Table 1
Summary of data-collection and refinement statistics.

	rC-half–Lac	rC-half–GalNAc
Data collection		
Space group	$P4_32_12$	$C2$
Unit-cell parameters		
a (Å)	61.2	123.2
b (Å)	61.2	34.8
c (Å)	175.6	61.6
β (°)		104.2
Resolution (Å)	50–1.90 (1.97–1.90)	50–1.80 (1.86–1.80)
Unique reflections	27297 (2660)	23141 (2213)
R_{merge} (%)	6.5 (29.4)	9.8 (30.8)
Completeness (%)	99.9 (100)	94.0 (97.2)
Average $I/\sigma(I)$	22.2 (3.3)	36.1 (8.6)
Average redundancy	7.0 (7.2)	7.2 (7.0)
Structure refinement		
Resolution (Å)	42.29–1.90	48.62–1.80
R factor (%)	17.9	19.5
R_{free}† (%)	22.5	26.0
R.m.s.d. from ideal		
Bond lengths (Å)	0.015	0.014
Bond angles (°)	1.403	1.423
Average B factors (Å^2)		
Protein (chain A/B)	20.3/20.3	13.4/13.9
Water molecules	33.0	23.9
Lac or GalNAc bound to chain A/B	31.7/24.2	12.9/13.5
Ramachandran plot (%)		
Favoured regions (chain A/B)	90.2/90.2	87.5/87.5
Allowed regions (chain A/B)	9.8/9.8	12.5/12.5
Disallowed regions (chain A/B)	0.0/0.0	0.0/0.0

† R_{free} factors were calculated using 5% of the unique reflections.

against 1 ml reservoir solution at 293 K. Rod-shaped crystals of rC-half–GalNAc were obtained within one week.

2.2. Data collection

Diffraction experiments using the crystals of the rC-half–Lac complex were conducted on beamline NW12, Photon Factory, Advanced Ring (PF-AR), Tsukuba, Japan (λ = 0.978 Å). The crystals were mounted in nylon loops (Hampton Research, California, USA) after soaking in cryoprotectant solution (20% or 3% glycerol in the precipitant solution) and then flash-cooled in a nitrogen stream at 95 K. Diffraction data were collected using a Quantum 210 CCD X-ray detector (Area Detector Systems Corporation, California, USA) in 1° oscillation steps over a range of 180°. The data set was processed and scaled using the programs *DENZO* and *SCALEPACK* from the *HKL*-2000 package (Otwinowski & Minor, 1997). The crystals belonged to the tetragonal space group $P4_32_12$, with unit-cell parameters $a = b = 61.2$, $c = 175.6$ Å, and diffracted to beyond 1.9 Å resolution. The data-collection statistics are shown in Table 1. Matthews coefficient calculations suggested that the V_M value of the crystals was 2.9 Å^3 Da^{-1}, assuming the presence of two rC-half molecules in the asymmetric unit (Matthews, 1968), with a solvent content of 58%.

Diffraction data from crystals of the complex with GalNAc were obtained on beamline BL-6A, Photon Factory, Tsukuba, Japan (λ = 1.0 Å); the crystals diffracted to beyond 1.8 Å resolution. Crystals were mounted in a quartz glass capillary of 0.3 mm diameter and were then flash-cooled in a nitrogen stream at 95 K. This capillary cryofreezing method was adopted in order to improve the mosaicity and the resolution of the data. Diffraction data were collected using a Quantum R4 CCD X-ray detector (Area Detector Systems Corporation, California, USA) in 1° oscillation steps over a range of 360°. The data set was processed to 1.8 Å resolution using the program *DENZO* and scaled using the program *SCALEPACK* (Otwinowski & Minor, 1997). Crystals of the rC-half–GalNAc complex grew under similar conditions to the crystals of the rC-half–Lac complex, but belonged to the monoclinic space group $C2$, with unit-cell parameters a = 123.2, b = 34.8, c = 61.6 Å, β = 104.2°. The V_M value for the GalNAc-complex crystal was calculated to be 2.2 Å^3 Da^{-1}, assuming the presence of two molecules in the asymmetric unit (the solvent content was approximately 44%; Matthews, 1968).

2.3. Model building and structure refinement

Structural analysis of each crystal was performed using the molecular-replacement method (Rossmann, 1990). For the rC-half–Lac complex data, structural models of two rC-half molecules A and B (residues 130–260) were initially produced using the xylan-binding domain (*So*CBM13) of *So*Xyn10A (PDB code 1xyf) as a search model with the program *MOLREP* (Vagin & Teplyakov, 1997) from the *CCP*4 package (Collaborative Computational Project, Number 4, 1994). After one trial of unrestrained refinement using the program *REFMAC*5 (Murshudov *et al.*, 1997, 1999), further model building was conducted using the program *ARP*/*wARP* (Perrakis *et al.*, 1999) starting from the phases output by *REFMAC*5; the R factor improved to 27.0% after 100 cycles of calculation. Several cycles of manual model rebuilding, introduction of water molecules with the programs *XtalView* (McRee, 1999) and *QUANTA*2000 (Accelerys, San Diego, California, USA) and refinement using the programs *CNS* and *REFMAC*5 gradually improved the structure model (Murshudov *et al.*, 1997, 1999; Brünger *et al.*, 1998). Clear densities corresponding to galactose molecules were found in the $F_{obs} - F_{calc}$ maps around the two sugar-binding sites of rC-half molecules. Two lactose molecules, two imidazole molecules and two phosphate ions were added to the model. The NCS constraint was not used in the refinement cycles. 5% of the observed reflections were set aside throughout refinement for cross-validation (Brünger, 1992). The crystallographic R and R_{free} factors of the rC-half–Lac complex improved to 17.9% and 22.5%, respectively.

Model building of the rC-half–GalNAc complex (residues 131–260) was conducted by the molecular-replacement method using the rC-half–Lac complex as a search model with the program *MOLREP* (Vagin & Teplyakov, 1997). Structure refinement was conducted using the program *REFMAC*5 (Murshudov *et al.*, 1997, 1999). The resultant $F_{obs} - F_{calc}$ or $2F_{obs} - F_{calc}$ maps yielded electron density corresponding to the bound GalNAc molecules. Three phosphate ions, one imidazole molecule and one Cd^{2+} ion were also added to the

model. The crystallographic R and R_{free} factors improved to 19.5% and 26.0%, respectively.

Unit-cell parameters and refinement statistics are shown in Table 1. The stereochemistry of all models was analyzed with the programs *PROCHECK* (Laskowski *et al.*, 1993) and *WHATCHECK*. The figures were drawn using the programs *MOLSCRIPT* and *RASTER3D* (Kraulis, 1991; Merritt & Bacon, 1997).

2.4. Preparation of ligand-docking models

Ligand models of the type I and type II lactosamines were prepared using the program *SWEET*2 (Bohne *et al.*, 1999). The generated models were manually superposed onto the position of the galactose moiety of the rC-half–Lac complex structure using the program *XtalView* (McRee, 1999). GalNAc-docked models of RTB, abrin, ebulin and *So*CBM13 were constructed similarly. The ligand-docked models were energy-minimized with the program *CNS* (Brünger *et al.*, 1998).

3. Results and discussion

3.1. Overall structure of rC-half

The rC-half–Lac and rC-half–GalNAc complex structures (Fig. 1) were determined at 1.9 and 1.8 Å resolution, respectively, using the molecular-replacement method. Crystal parameters and structural refinement statistics are shown in Table 1. Each crystal contained two molecules, *A* and *B*, in the asymmetric unit; the root-mean-square (r.m.s.) differences between them were 0.21 and 0.23 Å for C^{α}-atom pairs of the rC-half–Lac and rC-half–GalNAc complexes, respectively. In both complexes, electron density was observed throughout the polypeptide chains. The Ramachandran plot showed that all amino-acid residues were in the most favoured or additionally allowed regions. There were only slight differences between the assigned structures of molecules *A* and *B* in both complexes and subsequent descriptions will primarily be focused on molecule *A*.

The amino-acid residue numbering of rC-half (residues 130–260) corresponds to that of the full-length protein (Hirabayashi *et al.*, 1998). The models of rC-half are made up of a β-trefoil fold that displays characteristic pseudo-threefold symmetry (Murzin *et al.*, 1992). Three repeated peptides form a similar β-hairpin subdomain structure. Structure-based sequence alignment of rC-half with the β-trefoil domains of RTB and *So*CBM13 shows that the key $(QXW)_3$ motifs are well conserved among the subdomains (Fig. 2). 24 and 23 amino-acid residues are conserved in RTB1 and RTB2, respectively, and most of them occupy a similar spatial position in all R-type lectins, serving as conformational determinants that specify the fold. The overall structures of these domains are quite similar to each other and the r.m.s. deviations between the rC-half–Lac complex (molecule *A*) and RTB1, RTB2 and *So*CBM13 are 1.5, 1.7 and 1.3 Å for 117, 112 and 116 C^{α} atoms, respectively, although their primary sequence homologies are below 20%. An obvious difference between the structures of rC-half and the other R-

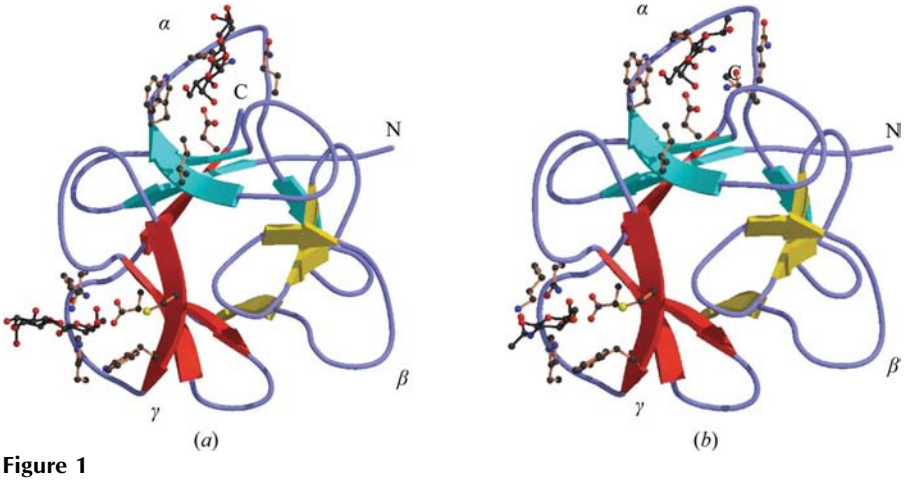

Figure 1
Overall structures of rC-half in complex with (*a*) lactose and (*b*) GalNAc. rC-half is composed of three repeated peptide segments referred to as subdomains α, β and γ, the β-strands of which are coloured blue, yellow and red, respectively. Bound ligands and the coordinating residues are shown as ball-and-stick drawings.

```
                                  *          * *          *QxW
N-half α      1   MA-GRPFLLVSKHSGHALSDPGSHGAPVIVNTRNGNDPRQLWYEDAST  47
N-half β     48   GTIRNQANQLALHV---GGDN----ILYLTPAEH-GQDAQQWK-VG-K  86
N-half γ     87   DVVHHRNDKEKVFDLCRGSKDVGAQVCAWKYH---GGTNQQWEAVHLK 132

rC-half α   134   FYIKSE-LNGKVLDIEGQNPAPGSKIITWDQKKGPTAVNQLWYTDQQ- 179
rC-half β   180   GVIRSK-LNDFAIDA---SHE---QIETQPFD--PNNPKRAWIVS-G- 216
rC-half γ   217   NTIAQLSDRDIVLDIIKSDKEAGAHICAWKQH---GGPNQKFIIESE- 260

Ricin B1α    11   VRIVG--RNGLCVDVRDGRFHNGNAIQLWPCKS-NTDANQLWTLKRD-  54
Ricin B1β    55   NTIRS--NG-KCLTTY--GYSPGVYVMIYDCNT-AATDATRWQIWDN-  95
Ricin B1γ    96   GTIINP-RSSLVLAAT--SGNSGTTLTVQTNI---YAVSQGWLPTNN- 136
Ricin B2α   142   TTIVG--LYGLCLQAN--S----GQVWIEDCSS--EKAEQQWALYAD- 178
Ricin B2β   179   GSIRPQQNRDNCLTS-DSN-IRETVVKILSCGPA--SSGQRWMFKND- 221
Ricin B2γ   222   GTILNL-YSGLVLDVRASDPSLK-QIILYPLH---GDPNQIWLPLF   262

SoCBM13 α   313   GQIKGV-GSGRCLDVPNASTTDGTQVQLYDCH---SATNQQWTYTDA- 355
SoCBM13 β   356   GELRV--YGDKCLDAA--GTGNGTKVQIYSCW---GGDNQKWRLNSD- 395
SoCBM13 γ   396   GSIVGV-QSGLCLDAVGGGTANGTLIQLYSCS---NGSNQRWTRT    436
```

Figure 2
Amino-acid sequence of rC-half and N-half, topologically aligned with RTB and *So*CBM13 (Rutenber & Robertus, 1991; Fujimoto *et al.*, 2000). Residues that belong to 3_{10}-helices and β-strands are highlighted in yellow and green, respectively. Key residues involved in ligand binding of rC-half and corresponding residues are marked with asterisks and coloured red.

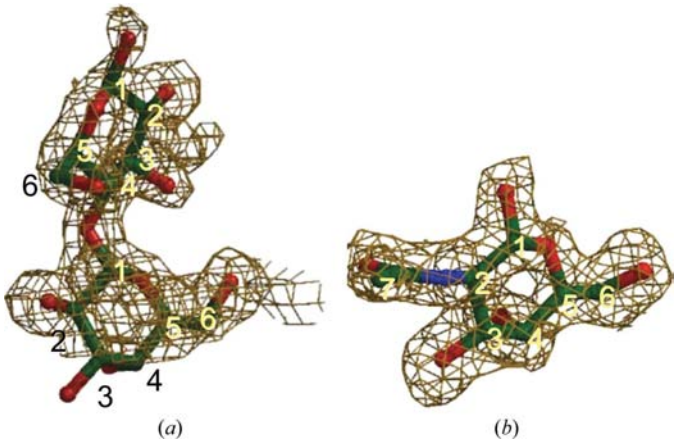

Figure 3
$F_{obs} - F_{calc}$ OMIT electron-density maps of (*a*) lactose and (*b*) GalNAc bound to subdomain α. The C atoms are numbered.

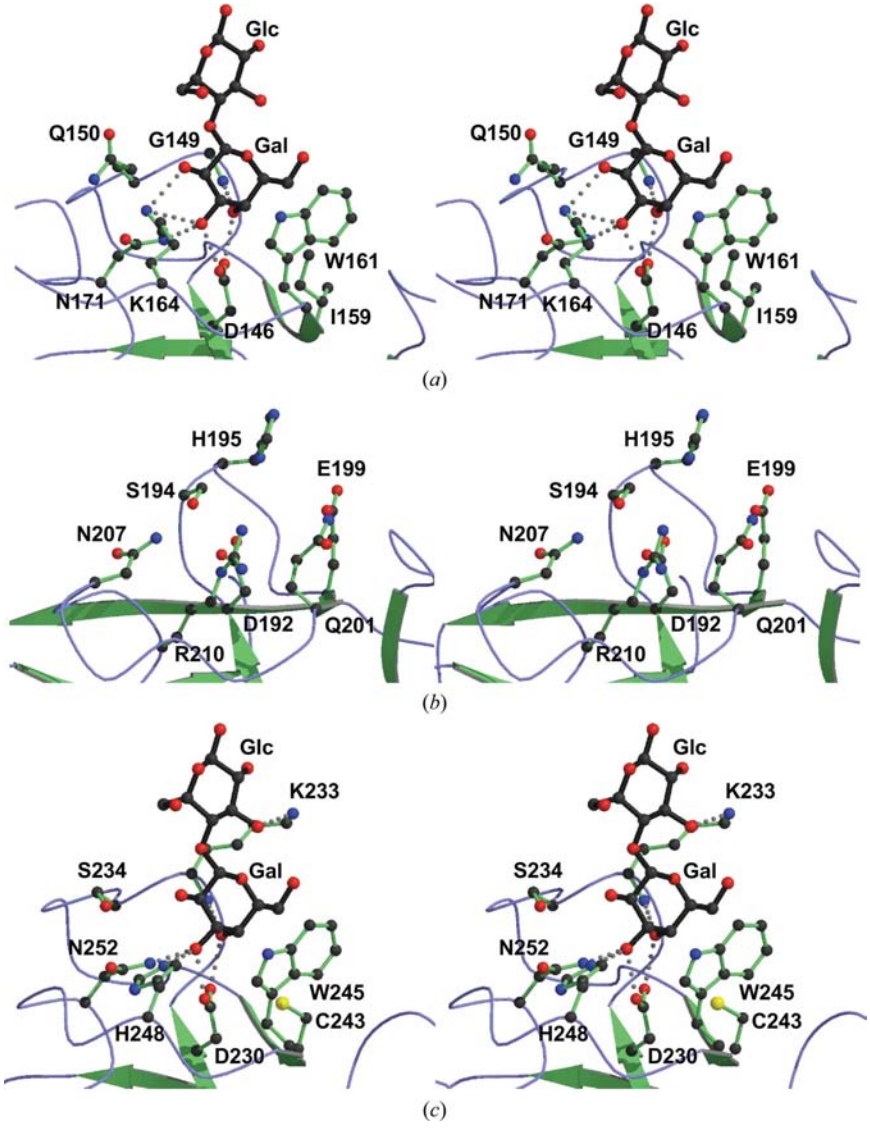

Figure 4
Stereoview of the sugar-binding sites of rC-half complexed with lactose (black) in (*a*) subdomain α, (*b*) subdomain β and (*c*) subdomain γ. Labelled residues participate in lactose binding. Hydrogen bonds are indicated by broken lines.

type lectins is that rC-half contains no disulfide bridges. This indicates that the cylindrical hydrophobic core located at the centre of the β-trefoil fold of C-half effectively acts as a stabilizer of the peptide fold.

The cocrystallized sugars lactose and GalNAc were found in subdomains α and γ of rC-half in both complex structures. $F_{obs} - F_{calc}$ OMIT electron-density maps for lactose and GalNAc molecules bound in subdomain α are shown in Fig. 3. The sugar-binding sites are located on the surfaces of these subdomains, facing in different directions.

3.2. Sugar-binding structure in the rC-half–Lac complex

The crystal structure of the rC-half–Lac complex is shown in Fig. 1(*a*). Cocrystallized lactose molecules were found in the sugar-binding sites of subdomains α and γ. These sites corresponded to those in RTB subdomains 1α and 2γ (Rutenber & Robertus, 1991) or to the xylan-binding sites of *So*CBM13 subdomains α, β and γ (Fujimoto *et al.*, 2002). Subdomain α binds the non-reducing-end galactose moiety of lactose through several polar and nonpolar residues (Fig. 4*a*). The Asp146 $O^{\delta2}$, Asn171 $N^{\delta2}$ and Lys164 N^{ζ} atoms form hydrogen bonds to the O3 atom of the galactose unit with distances of 2.5, 2.9 and 3.0 Å, respectively. The Asp146 $O^{\delta1}$ atom and main-chain N atom of Gly149 form hydrogen bonds to the O4 atom and the Lys164 N^{ζ} atom also interacts with the O2 atom of the galactose moiety (distances of 2.5, 2.8 and 2.8 Å, respectively). In addition to the polar contacts, van der Waals interactions also occur between the sugar and protein molecules. An important interaction is made by Trp161. The indole plane of the tryptophan residue makes a partial stacking interaction with the C3, C4, C5, C6, O3 and O6 atoms of the galactose moiety, which are almost in a single plane. The Gly149 C^{α} atom and the side chain of Gln150 also seem to make van der Waals interactions with the galactose moiety. There are no direct interactions between the glucose moiety of lactose and the amino-acid residues of subdomain α, although the O1 and O2 atoms make hydrogen bonds to adjacent protein molecules.

Subdomain γ binds the lactose in a similar manner as subdomain α (Fig. 4*c*). The O3 atom of the galactose moiety is hydrogen bonded to the Asp230 $O^{\delta2}$, Asn252 $N^{\delta2}$ and His248 $N^{\varepsilon2}$ atoms of rC-half, with distances of 2.5, 3.0 and 2.9 Å, respectively. The O4 atom forms hydrogen bonds to the Asp230 $O^{\delta1}$ atom and the main-chain N atom of Lys233, with distances of 2.6 and

3.0 Å, respectively. The aromatic ring of Trp245 forms a stacking interaction with the C3, C4, C5 and C6 atoms of the galactose unit. In subdomain γ, Lys164 in subdomain α is replaced by His248 and the presence of one hydrogen bond to the O2 atom is deduced. However, the relative position of the galactose moiety in the binding site is conserved between the two subdomains. Another difference is that Gly149 in subdomain α is replaced by Lys233. The hydrogen bond linking the main-chain N atom to the O4 atom of the galactose moiety is conserved, but intriguingly the side chain of Lys233 makes additional contacts with the glucose moiety of the bound lactose. They are hydrogen bonds between the N^{ζ} atom and the O2 and O3 atoms of the glucose unit (distances of 3.1 Å in both cases) and van der Waals interactions between the methylene groups and the O3 and C3 atoms. Despite these contacts, the average *B* factor of the glucose unit is higher than that of the galactose unit for all bound lactose molecules, indicating that the glucose moiety is slightly disordered. Although the electron-density map for the glucose unit of the lactose molecule in subdomain γ of molecule *A* was poorly defined, the sugar model could be constructed into the low-level electron density.

Bound sugar was not observed in subdomain β. Compared with the other subdomains, the aspartic acid Asp192 is conserved but tryptophan is absent in subdomain β (Fig. 2). Furthermore, the loop after Asp192 is shorter than in the other subdomains and the loop structure is different (Fig. 4*b*). In subdomains α and γ the tight turn involving the short 3_{10}-helix occurs after the conserved aspartic acid, providing one hydrogen-bond donor and the van der Waals wall of the binding pocket, while subdomain β does not have such a structure. Consequently, the sugar-binding pocket does not seem to exist in subdomain β, which lacks the ability to bind sugars.

Fig. 5 shows the superimposed lactose-binding structures of rC-half subdomain γ and RTB subdomain 2γ. This figure shows that rC-half and RTB have a similar binding manner for lactose. The positions of the conserved residues of rC-half, Asp230 and Asn252, the hydrogen-bond donor His248 and the stacking-platform provider Trp245 are conserved by the corresponding residues of RTB. The position of the galactose is strictly conserved in the binding pocket as well as the hydrogen-bond network around it. Fig. 2 indicates that the subdomains of RTB that can accommodate galactose also maintain the 3_{10}-helix turn structures after the conserved aspartic acid. These structural features are not only in present in RTB; they are mostly conserved in all sugar-binding sites of R-type lectins that bind to galactose or galacto-oligosaccharides and are considered to be indispensable for binding the galactose molecule in R-type lectins. In contrast, the position of the glucose moiety of the bound lactose varies slightly between the proteins. In rC-half subdomain α there is no contact between the glucose unit and the binding site, while there are two hydrogen bonds to the glucose unit that originate from the terminal N atom of Lys233 in subdomain γ. In RTB subdomains 1α and 2γ the glucose units are recognized by arginines. Thus, glucose is only recognized by long flexible side chains in these structures. The average thermal factors for the galactosyl pyranose ring of the bound lactose are 18.5, 32.4, 19.4 and 18.6 $Å^2$ in subdomains α and γ of molecule *A* and subdomains α and γ of molecule *B*, respectively, while those for the glucosyl pyranose ring are 27.3, 47.7, 29.8 and 28.1 $Å^2$, respectively, indicating that the glucose unit is not as tightly bound as the galactose unit in each subdomain. These facts demonstrate that the galactose-binding sites of R-type lectins bind sugars by recognizing the single galactose unit of the galactosyl oligosaccharides.

The N-half did not to bind to asialofetuin–agarose, but recent biochemical analysis has indicated that the N-half itself also shows binding ability towards galacto-oligosaccharides (Hirabayashi *et al.*, unpublished data). Amino-acid sequence alignment with the C-half shows that subdomain γ of the N-half is likely to contain the structures that are indispensable for galactose binding as described above (Fig. 2). The N-half subdomain γ has a conserved aspartic acid, aspargine and tryptophan and also the 3_{10}-helix structure region after the conserved aspartic acid; therefore, the N-half seems to recognize galactosyl oligosaccharides through subdomain γ and may be able to participate in the haemagglutination activity of EW29.

3.3. Sugar-binding structure in the rC-half–GalNAc complex

The crystal structure of the rC-half–GalNAc complex is shown in Fig. 1(*b*). Bound GalNAc molecules were observed in subdomains α and γ of the rC-half–GalNAc complex (Fig. 6). Both subdomains showed a similar sugar binding to that observed in the rC-half–Lac complex. The position of the bound GalNAc can mostly be superimposed on the galactose moiety of the bound lactose. In subdomain α, differences between the rC-half–Lac and rC-half–GalNAc complexes were found in the hydrogen-bond interactions from residues Gln150 and Lys164. The hydrogen bond between the O2 atom of the galactose moiety of the bound lactose and the Lys164 N^{ζ} atom was lost because of structural differences between the two sugars: the 2-N-acetyl group and the 2-hydroxyl group. A new long-distance hydrogen-bond interaction was found between the O7 atom of GalNAc and the Gln150 $O^{\varepsilon 1}$ atom, which did not form a hydrogen bond to the bound lactose in the rC-half–Lac complex, with a distance of 3.2 Å. In subdomain γ, rC-half binds to GalNAc in the same manner as the galactose moiety of the lactose molecule, with no hydrogen-bond interactions with the C2-substituent. These sugar-bound structures indicate that the binding ability of rC-half for GalNAc is expected to be almost the same as that for galactose; this is consistent with the biochemical data, in which no remarkable difference was detected in the binding ability of rC-half for galactosyl and *N*-acetylgalactosaminyl oligosaccharides (Hirabayashi *et al.*, 1998).

3.4. Crystal packing *via* bound sugars

Since the crystals of the two sugar complexes were obtained using the cocrystallization method, the sugar-binding sites revealed by this study could be the most plausible ligand-

binding sites. Although similar crystallization conditions were adopted to produce the crystals of the two complexes, the resulting crystal packing was different. This may be because the glucose moieties of the lactose molecule are involved in crystal packing in the rC-half–Lac complex. For the bound lactose in subdomain α of molecule *A*, the O1 and the O2 atoms make direct hydrogen bonds to the side-chain carboxylate of Glu237 of the adjacent molecule and the O1 atom also makes a further hydrogen bond to the main-chain N atom of Lys130 of the other molecule. The O1 atom of the lactose in subdomain α of molecule *B* also makes a hydrogen bond to the main-chain N atom of the adjacent Lys130. The glucose moiety of the bound lactose in subdomain γ of molecule *B* makes a stacking interaction with the imidazole ring of His248 of a symmetry-related protein molecule. The lactose molecule in subdomain γ of molecule *A* does not make direct interactions with other protein molecules, although several water-mediated interactions can be observed. This might be one of the reasons why the thermal factor for this lactose molecule is higher than those of the other identified lactose molecules. Conversely, the bound GalNAc molecules in the rC-half–GalNAc complex make no direct contacts with adjacent molecules.

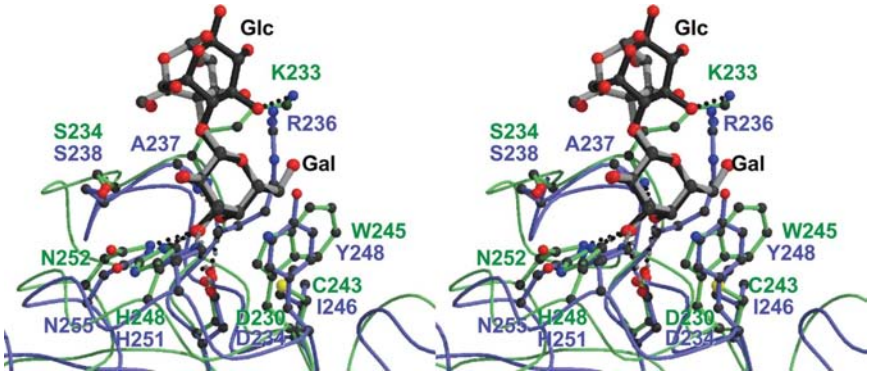

Figure 5
Stereoview of the superposition of subdomain γ of rC-half (green) on subdomain γ of RTB (RTB2γ; ultramarine).

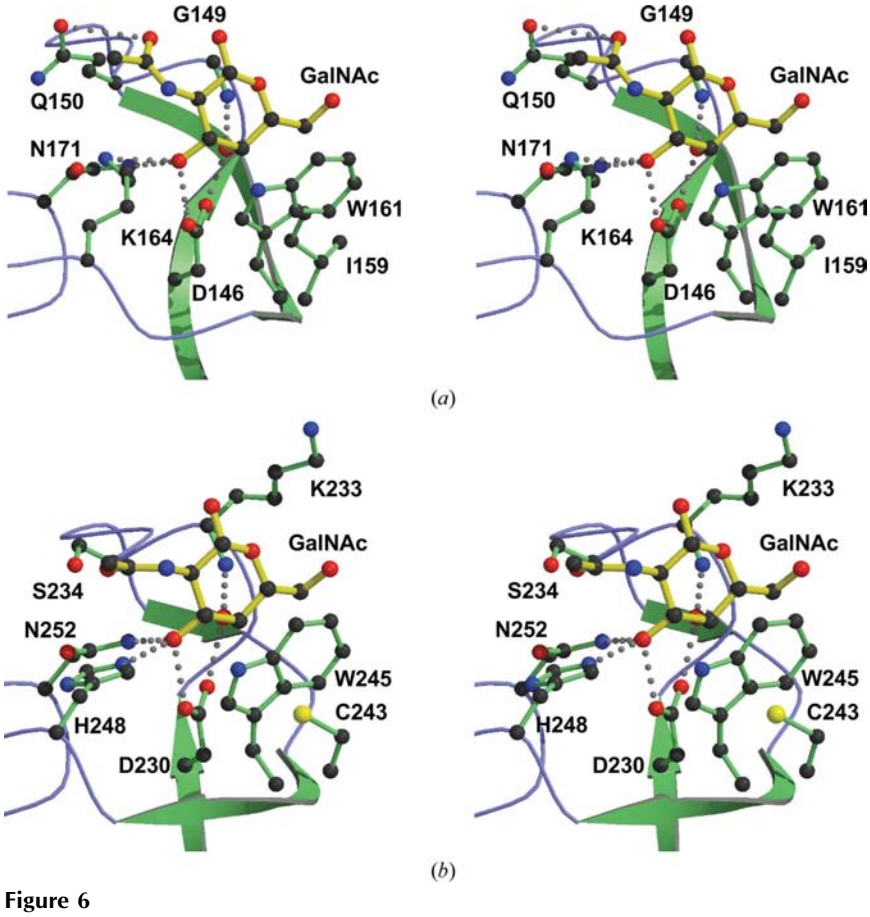

Figure 6
Stereoview of the sugar-binding sites of rC-half complexed with GalNAc (yellow) in (*a*) subdomain α and (*b*) subdomain γ. Labelled residues participate in GalNAc binding. Hydrogen bonds are indicated by broken lines.

3.5. Ligand specificity

R-type lectins were previously considered to be Gal/GalNAc-binding lectins. In order to investigate the sugar-binding specificity of GalNAc, GalNAc-binding models for RTB, abrin, ebulin and *So*CBM13 were constructed based on their crystal structures. The constructed GalNAc-binding models suggested that all the proteins could accommodate the GalNAc molecule as well as galactose, similar to rC-half, and may show almost the same sugar-binding affinity for galactose and GalNAc.

Biochemical experiments indicated that rC-half could bind to *N*-acetylglucosaminyl sugar-containing disaccharides such as LacNAc (Gal-β1,4-GlcNAc; a type II lactosamine) or lacto-*N*-biose I (Gal-β1,3-GlcNAc; a type I lactosamine) more tightly than to lactose or GalNAc (Hirabayashi *et al.*, 1998). Attempts to crystallize the rC-half–LacNAc complex failed to produce any crystals. Instead, type I and type II lactosamine docking models of rC-half were constructed manually based on the structure of the rC-half–Lac complex (Fig. 7). According to the docking model, the O1 atom of the bound galactose faces the solvent region and there is sufficient space to link the adjacent sugar moiety through a β-1,3- or β-1,4-glycoside bond. In the rC-half–Lac complex structure the glucose moiety did not make direct contact with the binding pocket in subdomain α, but the GlcNAc moiety of the type I lactosamine molecule seems to make a hydrogen bond to the protein between the O7 atom of the *N*-acetyl group and the Gln150 $N^{\varepsilon 2}$ atom

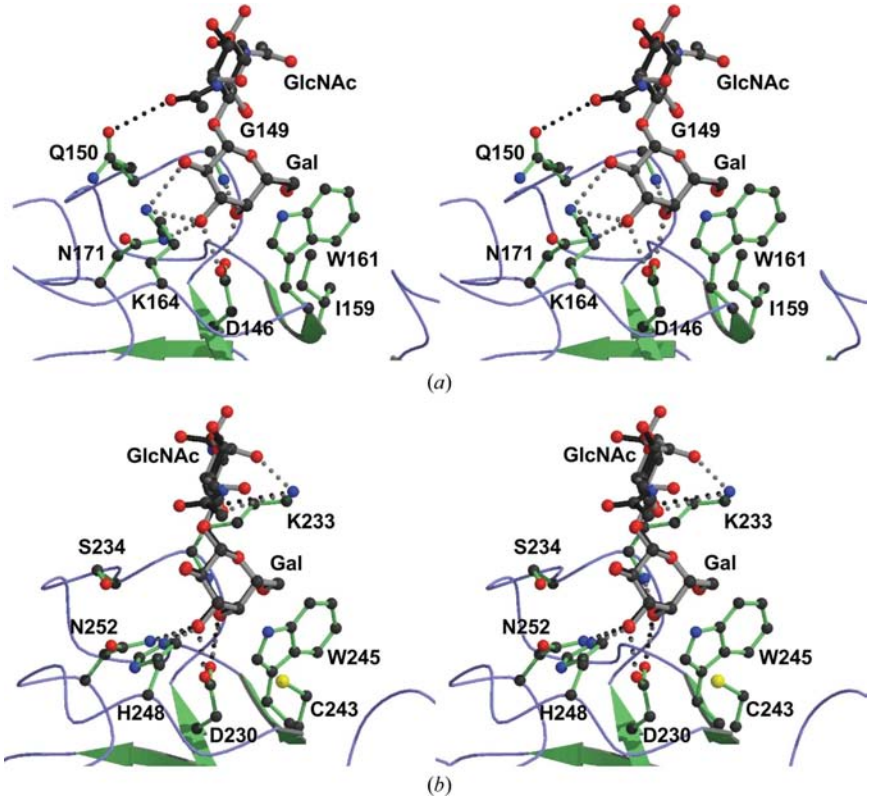

Figure 7
Stereoview of putative Gal-β1,3-GlcNAc (a type I lactosamine; black) and Gal-β1,4-GlcNAc (a type II lactosamine; grey) binding (*a*) in subdomain α of rC-half and (*b*) in subdomain γ.

(Fig. 7*a*). This bond might be caused by the flexibility around the β-1,3-glycoside linkage. Conversely, subdomain γ of rC-half lacks a hydrogen bond to the *N*-acetyl group of type I lactosamine (Fig. 7*b*). In contrast, a new hydrogen bond between the O7 atom of the *N*-acetyl group and the Lys233 N^{ζ} atom is found in the model for type II lactosamine and a hydrogen bond between the Lys233 N^{ζ} atom and the O3 atom of the glucose moiety was also observed in the rC-half–Lac complex. Additional hydrogen bonds were found in subdomains α or γ and in this sense rC-half appears to have advantages in recognizing both type I and type II lactosamines compared with the rC-half–Lac complex, which is compatible with the biochemical data.

In general, the biosynthesis of *N*-glycan in Golgi apparatus is initiated by addition of a GlcNAc unit to an Asn residue; the glycan is then branched at the trimannosyl core and the biosynthesis is terminated at the peripheral galactose. The *N*-glycans are occasionally decorated with Neu5Ac and L-fucose. It is considered that the glycan-synthesis system co-evolved with glycan-recognition systems such as lectins based on a rule of 'terminal addition', in which the protein recognizes GlcNAc and mannose at an earlier stage than Neu5Ac, L-fucose or galactose. This has been proposed as the 'galactose latecomer' hypothesis (Hirabayashi, 2004). R-type lectins are the most widely distributed family in nature and have the potential to bind various ligands. For instance, ppGalNAc-T has an R-type lectin domain at the C-terminus and binds GalNAc (Kubota *et al.*, 2006). SSA binds Neu5Ac-α2,6-Gal or Neu5Ac-α2,6-GalNAc 1000-fold more strongly than galactose or GalNAc (Kaku *et al.*, 1996). *So*Xyn10A binds not only xylo-oligosaccharides but also galactose or lactose with rather high binding affinities (Fujimoto *et al.*, 2002; Ito *et al.*, 2004; Kuno *et al.*, 2000). Recently, Kaku and coworkers have pointed out that Neu5Ac-specific R-type lectins evolved from ancestral Gal/GalNAc-binding lectins (Kaku *et al.*, 2007). In this sense, R-type lectins are good targets for investigating evolutionary changes in ligand specificity. Recently, we reported the first example of the modification of ligand specificity of an R-type lectin from galactose specificity to Neu5Ac specificity by employing natural evolution mimicry (Yabe *et al.*, 2007). This might be a good example of artificial reproduction of evolution in glycan-recognition systems. The crystal structures presented here would provide the structural basis for further molecular design of R-type lectins that are specific for various sugars.

In conclusion, sugar-bound crystal structures of rC-half were determined. The sugar-binding manner of rC-half was consistent with that of other lectins from the same family such as RTB. Two subdomains, α and γ, of rC-half were found to bind the non-reducing-end galactose unit of glycans such as lactose. We showed that rC-half has two sugar-binding sites, resulting in haemagglutination of erythrocytes. Although the biological function of EW29 is unknown because it only has a lectin domain and not an enzyme module, it is of interest to study why EW29 exists in the earthworm. In the case of ricin, RTB plays a role in binding to the cell surface as a lectin, while the A chain plays a crucial role in degrading RNA, resulting in a highly toxic activiy. In the case of *So*CBM13 and *Sl*CBM13, a role is played in binding to the insoluble substrate through any of the three xylose-binding sites for efficient xylanase activity. The biological role of EW29 would therefore seem to be something like the aggregation of galactose-containing oligosaccharides or polysaccharides. However, in EW29, since the C-half has a lectin activity by itself and the N-half shows little sugar-binding ability, it is important to clarify the role of the N-half in relation to the C-half in order to elucidate the biological function of EW29. Further structural analysis of whole EW29 containing the N-half is required.

Synchrotron radiation was used for this work and we would like to thank the scientists and staff at the Photon Factory.

References

Arndt, J. W., Gu, J., Jaroszewski, L., Schwarzenbacher, R., Hanson, M. A., Lebeda, F. J. & Stevens, R. C. (2005). *J. Mol. Biol.* **346**, 1083–1093.

Bohne, A., Lang, E. & von der Lieth, C. W. (1999). *Bioinformatics*, **15**, 767–768.
Brünger, A. T. (1992). *Nature (London)*, **355**, 472–475.
Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J.-S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). *Acta Cryst.* D**54**, 905–921.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Endo, Y. & Tsurugi, K. (1987). *J. Biol. Chem.* **262**, 8128–8130.
Frankel, A. E., Burbage, C., Fu, T., Tagge, E., Chandler, J. & Willingham, M. C. (1996). *Biochemistry*, **35**, 14749–14756.
Fritz, T. A., Hurley, J. H., Trinh, L. B., Shiloach, J. & Tabak, L. A. (2004). *Proc. Natl Acad. Sci. USA*, **101**, 15307–15312.
Fujimoto, Z., Kuno, A., Kaneko, S., Kobayashi, H., Kusakabe, I. & Mizuno, H. (2002). *J. Mol. Biol.* **316**, 65–78.
Fujimoto, Z., Kuno, A., Kaneko, S., Yoshida, S., Kobayashi, H., Kusakabe, I. & Mizuno, H. (2000). *J. Mol. Biol.* **300**, 575–585.
Hartley, M. R. & Lord, J. M. (2004). *Biochim. Biophys. Acta*, **1701**, 1–14.
Hatakeyama, T., Unno, H., Kouzuma, Y., Uchida, T., Eto, S., Hidemura, H., Kato, N., Yonekura, M. & Kusunoki, M. (2007). *J. Biol. Chem.* **282**, 37826–37835.
Hazes, B. (1996). *Protein Sci.* **5**, 1490–1501.
Hemmi, H., Kuno, A., Ito, S., Suzuki, R., Kaneko, S., Hasegawa, T., Hirabayashi, J. & Kasai, K. (2004). *J. Biomol. NMR*, **30**, 377–378.
Hirabayashi, J. (2004). *Trends Glycosci. Glycotechnol.* **16**, 63–85.
Hirabayashi, J., Dutta, S. K. & Kasai, K. (1998). *J. Biol. Chem.* **273**, 14450–14460.
Inoue, K., Sobhany, M., Transue, T. R., Oguma, K., Pedersen, L. C. & Negishi, M. (2003). *Microbiology*, **149**, 3361–3370.
Ito, S., Kuno, A., Suzuki, R., Kaneko, S., Kawabata, Y., Kusakabe, I. & Hasegawa, T. (2004). *J. Biotechnol.* **110**, 137–142.
Kaku, H., Kaneko, H., Minamihara, N., Iwata, K., Jordan, E. T., Rojo, M. A., Minami-Ishii, N., Minami, E., Hisajima, S. & Shibuya, N. (2007). *J. Biochem.* **142**, 393–401.
Kaku, H., Tanaka, Y., Tazaki, K., Minami, E., Mizuno, H. & Shibuya, N. (1996). *J. Biol. Chem.* **271**, 1480–1485.
Kraulis, P. J. (1991). *J. Appl. Cryst.* **24**, 946–950.
Kubota, T., Shiba, T., Sugioka, S., Furukawa, S., Sawaki, H., Kato, R., Wakatsuki, S. & Narimatsu, H. (2006). *J. Mol. Biol.* **359**, 708–727.
Kuno, A., Kaneko, S., Ohtsuki, H., Ito, S., Fujimoto, Z., Mizuno, H., Hasegawa, T., Taira, K., Kusakabe, I. & Hayashi, K. (2000). *FEBS Lett.* **482**, 231–236.
Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.
Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
McRee, D. E. (1999). *J. Struct. Biol.* **125**, 156–165.
Merritt, E. A. & Bacon, D. J. (1997). *Methods Enzymol.* **277**, 505–524.
Montfort, W., Villafranca, J. E., Monzingo, A. F., Ernst, S. R., Katzin, B., Rutenber, E., Xuong, N.-H., Hamlin, R. & Robertus, J. D. (1987). *J. Mol. Biol.* **262**, 5398–5403.
Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). *Acta Cryst.* D**53**, 240–255.
Murshudov, G. N., Vagin, A. A., Lebedev, A., Wilson, K. S. & Dodson, E. J. (1999). *Acta Cryst.* D**55**, 247–255.
Murzin, A. G., Lesk, A. M. & Chothia, C. (1992). *J. Mol. Biol.* **223**, 531–543.
Nešić, D., Hsu, Y. & Stebbins, C. E. (2004). *Nature (London)*, **429**, 429–433.
Notenboom, V., Boraston, A. B., Williams, S. J., Kilburn, D. G. & Rose, D. R. (2002). *Biochemistry*, **41**, 4246–4254.
Olsnes, S. & Pihl, A. (1973). *Biochemistry*, **12**, 3121–3126.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Pascal, J. M., Day, P. J., Monzingo, A. F., Ernst, S. R., Robertus, J. D., Iglesias, R., Perez, Y., Ferreras, J. M., Citores, L. & Girbes, T. (2001). *Proteins*, **43**, 319–326.
Perrakis, A., Morris, R. & Lamzin, V. S. (1999). *Nature Struct. Biol.* **6**, 458–463.
Rutenber, E., Ready, M. & Robertus, J. D. (1987). *Nature (London)*, **326**, 624–626.
Rutenber, E. & Robertus, J. D. (1991). *Proteins*, **10**, 260–269.
Rossmann, M. G. (1990). *Acta Cryst.* A**46**, 73–82.
Suzuki, R., Fujimoto, Z., Kuno, A., Hirabayashi, J., Kasai, K. & Hasegawa, T. (2004). *Acta Cryst.* D**60**, 1895–1896.
Tahirov, T. H., Lu, T. H., Liaw, Y. C., Chen, Y. L. & Lin, J. Y. (1995). *J. Mol. Biol.* **250**, 354–367.
Uchida, T., Yamasaki, T., Eto, S., Sugawara, H., Kurisu, G., Nakagawa, A., Kusunoki, M. & Hatakeyama, T. (2004). *J. Biol. Chem.* **279**, 37133–37141.
Vagin, A. & Teplyakov, A. (1997). *J. Appl. Cryst.* **30**, 1022–1025.
Villafranca, J. E. & Robertus, J. D. (1981). *J. Biol. Chem.* **256**, 554–556.
Yabe, R., Suzuki, R., Kuno, A., Fujimoto, Z., Jigami, Y. & Hirabayashi, J. (2007). *J. Biochem.* **141**, 389–399.