BIOLOGICAL CRYSTALLOGRAPHY

ISSN 1399-0047

Three-dimensional structure and ligand-binding site of carp fishelectin (FEL)

Stefano Capaldi,[a] Beniamino Faggion,[a] Maria E. Carrizo,[b] Laura Destefanis,[a] Maria C. Gonzalez,[a] Massimiliano Perduca,[a] Michele Bovi,[a] Monica Galliano[c] and Hugo L. Monaco[a]*

[a]Biocrystallography Laboratory, Department of Biotechnology, University of Verona, Ca Vignal 1, Strada Le Grazie 15, 37134 Verona, Italy, [b]Departamento de Química Biológica, Facultad de Ciencias Químicas, Universidad Nacional de Córdoba, CP 5016, Córdoba, Argentina, and [c]Department of Molecular Medicine, University of Pavia, Via Taramelli 3b, 27100 Pavia, Italy. *Correspondence e-mail: hugo.monaco@univr.it

Received 1 December 2014
Accepted 27 February 2015

Edited by K. Miki, Kyoto University, Japan

Keywords: fishelectin; fish-egg lectin; β-propeller; tachylectins.

PDB references: carp fishelectin, apo, 4ruq; holo, 4rus

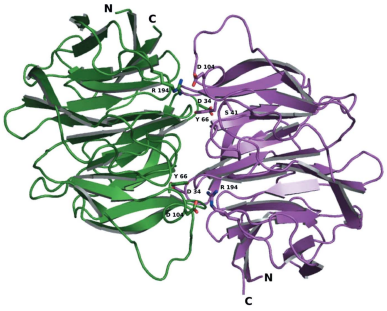
Supporting information: this article has supporting information at journals.iucr.org/d

Carp FEL (fishelectin or fish-egg lectin) is a 238-amino-acid lectin that can be purified from fish eggs by exploiting its selective binding to Sepharose followed by elution with *N*-acetylglucosamine. Its amino-acid sequence and other biochemical properties have previously been reported. The glycoprotein has four disulfide bridges and the structure of the oligosaccharides linked to Asn27 has been described. Here, the three-dimensional structures of apo carp FEL (cFEL) and of its complex with *N*-acetylglucosamine determined by X-ray crystallography at resolutions of 1.35 and 1.70 Å, respectively, are reported. The molecule folds as a six-bladed β-propeller and internal short consensus amino-acid sequences have been identified in all of the blades. A calcium atom binds at the bottom of the funnel-shaped tunnel located in the centre of the propeller. Two ligand-binding sites, α and β, are present in each of the two protomers in the dimer. The first site, α, is closer to the N-terminus of the chain and is located in the crevice between the second and the third blades, while the second site, β, is located between the fourth and the fifth blades. The amino acids that participate in the contacts have been identified, as well as the conserved water molecules in all of the sites. Both sites can bind the two anomers, α and β, of *N*-acetylglucosamine, as is clearly recognizable in the electron-density maps. The lectin presents sequence homology to members of the tachylectin family, which are known to have a function in the innate immune system of arthropods, and homologous genes are present in the genomes of other fish and amphibians. This structure is the first of a protein of this group and, given the degree of homology with other members of the family, it is expected that it will be useful to experimentally determine other crystal structures using the coordinates of cFEL as a search probe in molecular replacement.

1. Introduction

Lectins are proteins of non-immune origin that are devoid of catalytic activity; they reversibly bind carbohydrates with high specificity and are involved, through sugar binding, in many fundamental biological processes such as cell–cell interactions and immunity (Sharon, 2007). They are present in all kinds of organisms from viruses to man, were initially identified in the plant kingdom for their haemagglutination activity and are now being widely used in basic and applied research. Lectins are found either in soluble form or membrane-bound and are usually structurally complex molecules with one or more carbohydrate-recognition domains (Taylor & Drickamer, 2003).

Animal lectins can have a variety of physiological functions, among which is a crucial role in the immune response, in particular in the innate immune system of vertebrates and invertebrates (Rabinovich *et al.*, 2012).



During the last three decades, the molecular structures and functions of several lectin components of the innate immune system of invertebrates have been established. A model species is the Japanese horseshoe crab, *Tachypleus tridentatus*, which possesses several lectins, present in both the haemolymph and haemocytes, that have been shown to play a role in the defence of the arthropod against invading bacterial, fungal and viral pathogens (Kawabata & Iwanaga, 1999; Iwanaga, 2002; Iwanaga & Lee, 2005).

Teleost fish have a very sophisticated repertoire of proteins involved in innate immunity that includes lectins, some of which have been identified and characterized in the eggs, where they are generally believed to have a protective function against pathogens during embryo development (Russell & Lumsden, 2005; Vasta *et al.*, 2011; Zhu *et al.*, 2013). The fact that most fish-egg lectins selectively bind rhamnose, a carbohydrate that is absent in animal cells but is very widely distributed in the bacterial cell wall, coupled to the observation that their expression can be up-regulated in the presence of challenging bacteria, has supported the proposal of a role in the defence mechanism against infection (Kim *et al.*, 2011).

Some years ago, we isolated and sequenced a novel glycoprotein present in the eggs of the carp (*Cyprinus carpio*; Galliano *et al.*, 2003). The protein, which binds to a Sepharose 4B matrix column and can be eluted with 0.4 *M* *N*-acetylglucosamine, behaves like a lectin with a molecular mass of 26 686 Da. We also determined its 238-amino-acid sequence, the positions of its four disulfide bridges and the structure of its single N-linked carbohydrate chain. The lectin does not bind rhamnose, shows a very low agglutinating activity for human A-type erythrocytes and interacts very weakly with both Gram-positive and Gram-negative bacteria, with these latter interactions being inhibited by *N*-acetylglucosamine. A database search showed its amino-acid sequence to be significantly similar to those of members of the invertebrate lectin family, which includes tachylectin-1, which is present in the amoebocytes of the horseshoe crab *T. tridentatus* and is known to participate in the innate defence system of this species (Saito *et al.*, 1995; Chen *et al.*, 2001), and two other lectins, tectonins I and II, that were characterized in the plasmodium *Physarum polycephalum* and are located on the external surface of the plasma membrane (Huh *et al.*, 1998). We proposed the name fishelectins (by analogy with tachylectins) or fish-egg lectins for this new vertebrate protein family.

More recently, a different 24 kDa molecular-weight lectin was purified from salmon eggs and crystallized in a form suitable for X-ray diffraction studies (Murata *et al.*, 2007). This lectin was found to be specific for galactose/rhamnose and is believed not to be involved in the innate immunity of the fish but to block polyspermy during fertilization. Its amino-acid sequence does not show any similarity to that of carp fish-electin (cFEL) and its three-dimensional structure has not yet been determined.

Here, we report the three-dimensional structure of cFEL determined by X-ray diffraction analysis of single crystals of the apo form and of the complex with *N*-acetylglucosamine (GlcNAc). The molecule folds as a six-bladed β-propeller and possesses two binding sites for GlcNAc that can bind both carbohydrate anomers. This is the first reported structure of a member of this novel protein family.

2. Materials and methods

2.1. Protein purification

Carp FEL was purified as described previously (Galliano *et al.*, 2003). Briefly, 500 g of frozen carp eggs were thawed overnight in 2 l 20 m*M* Tris–HCl pH 7.4, 50 m*M* KCl, 5 m*M* $CaCl_2$, 0.02% NaN_3 containing phenylmethylsulfonyl fluoride (PMSF) as a protease inhibitor and homogenized in a blender for 3–4 min at 4°C. The mixture was centrifuged at 8000 rev min^{-1} for 1 h at 4°C, the pellet was discarded and the supernatant was filtered through a 10 µm nylon mesh and then applied onto a Sepharose 4B column (2.6 × 15 cm) previously equilibrated with the same buffer. The column was extensively washed until the absorbance monitored at 280 nm was negligible. The bound lectin was then eluted with 0.4 *M* *N*-acetylglucosamine in the same Tris–HCl buffer followed by extensive dialysis at 4°C against 20 m*M* Tris–HCl pH 7.4, 0.02% NaN_3 to remove the excess sugar.

The protein concentration was determined from the specific extinction coefficient at 280 nm using ε = 1.5 ml mg^{-1} cm^{-1}, and the purity and molecular weight were assessed by SDS–PAGE with standard molecular-weight markers.

2.2. Crystallization

Crystallization experiments with apo cFEL were performed by the hanging-drop vapour-diffusion method using 96 different conditions (Crystal Screen and Crystal Screen 2, Hampton Research). Each drop was prepared by mixing 1 µl of a 20 mg ml^{-1} protein solution (in 20 m*M* Tris–HCl pH 7.4, 0.02% NaN_3) with the same volume of precipitant solution and was equilibrated against a volume of 0.3 ml in the reservoir. The crystallization experiments were kept at 4°C and periodically checked. Clusters of small plate-shaped crystals were obtained after about two weeks in a drop containing 10% 2-propanol, 20% PEG 4000, 0.1 *M* HEPES pH 7.5 as the precipitant. After refinement of the initial conditions, diffraction-quality crystals could be grown using the sitting-drop method (5 µl protein solution and 5 µl precipitant solution) with 15% 2-propanol, 20% PEG 6000, 0.1 *M* HEPES pH 7.5 as the precipitant.

Heavy-atom derivatives for MIR (multiple isomorphous replacement) phasing were prepared by soaking apo cFEL crystals in mother liquor containing an excess of the heavy-atom compound and leaving the crystals at 4°C for 12–48 h. Co-crystals of cFEL with *N*-acetylglucosamine were obtained by the sitting-drop method by mixing equal volumes of a 100 m*M* solution of the saccharide containing 20 mg ml^{-1} cFEL with 19% PEG 3350, 0.2 *M* magnesium formate, 50 m*M* Tris–HCl pH 8.0 as the precipitant.

2.3. Data collection and processing

X-ray data for the native apo cFEL crystals, as well as for the heavy-atom derivatives necessary for phasing, were

Table 1
Data-collection, phasing and refinement statistics for carp FEL.

Values in parentheses are for the highest resolution shell.

FEL data set	Native apo (used for phasing)	K_2PtCl_4 derivative	$K_3UO_2F_5$ derivative	Mercury(II) acetate derivative	High-resolution apo form (refined)	Co-crystals with GlcNAc
Space group	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$C2_1$
Unit-cell parameters						
a (Å)	44.7	44.7	44.7	44.5	44.3	204.4
b (Å)	72.3	71.8	71.7	70.5	71.0	120.8
c (Å)	167.6	167.8	166.4	171.9	163.0	69.3
α (°)	90.0	90.0	90.0	90.0	90.0	90.0
β (°)	90.0	90.0	90.0	90.0	90.0	91.4
γ (°)	90.0	90.0	90.0	90.0	90.0	90.0
Molecules in asymmetric unit	2	2	2	2	2	6
Resolution range (Å)	19.9–2.50	19.9–2.50	25.0–2.50	24.5–2.50	40.8–1.35	29.7–1.70
Observed reflections	77678	73521	68950	55616	918399	409533
Independent reflections	18173	18124	19967	18339	113338	171485
Multiplicity	4.3	4.1	3.5	3.0	8.1	2.4
R_{merge}† (%)	6.1	6.3	9.0	12.1	8.9 (37.9)	9.8 (35.1)
$\langle I/\sigma(I)\rangle$	11.1	11.2	8.5	4.8	14.2 (5.4)	6.6 (2.7)
Completeness (%)	93.2	93.5	92.1	94.2	99.6 (99.0)	93.2 (95.1)
MFID (%)	—	23.6	19.6	9.8	—	—
No. of sites	—	9	2	4	—	—
Phasing power (acentric/centric)	—	1.19/1.40	1.38/1.19	0.96/0.81	—	—
Reflections in refinement	—	—	—	—	107565	162866
R_{cryst}‡ (%)	—	—	—	—	17.3 (19.8)	19.5 (25.0)
R_{free}§ (%)	—	—	—	—	18.5 (20.8)	21.4 (27.5)
No. of atoms						
Protein atoms	—	—	—	—	3605	10814
Ligand atoms						
IPA	—	—	—	—	20	—
GOL	—	—	—	—	18	—
NAG–NDG–PE1	—	—	—	—	—	167
Water molecules	—	—	—	—	385	783
R.m.s.d.¶						
Bond lengths (Å)	—	—	—	—	0.005	0.007
Bond angles (°)	—	—	—	—	1.103	1.151
Planar groups (Å)	—	—	—	—	0.004	0.004
Chiral volumes (Å^3)	—	—	—	—	0.076	0.080
Average B factors (Å^2)						
Overall	—	—	—	—	7.46	11.65
Protein atoms	—	—	—	—	6.41	11.05
Ligand atoms						
IPA	—	—	—	—	12.8	—
GOL	—	—	—	—	12.6	—
NAG–NDG–PE1	—	—	—	—	—	19.89
Solvent atoms	—	—	—	—	15.98	19.97
Ramachandran plot						
Most favourable (%)	—	—	—	—	89	88.1
Allowed (%)	—	—	—	—	11	11.9

† $R_{merge} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle|/\sum_{hkl}\sum_i I_i(hkl)$, where $\langle I(hkl)\rangle$ is the mean intensity of the i observations of reflection hkl. ‡ $R_{cryst} = \sum_{hkl} ||F_{obs}| - |F_{calc}||/\sum_{hkl}|F_{obs}|$, where $|F_{obs}|$ and $|F_{calc}|$ are the observed and calculated structure-factor amplitudes, respectively. Summation includes all reflections used in the refinement. § $R_{free} = \sum_{hkl} ||F_{obs}| - |F_{calc}||/\sum_{hkl}|F_{obs}|$ evaluated for a randomly chosen subset of 5% of the diffraction data not included in the refinement. ¶ Root-mean-square deviation from ideal values.

collected at room temperature in our laboratory using the oscillation method from crystals mounted in glass capillaries on a Rigaku RU-300 rotating-anode X-ray source using a Rigaku R-AXIS II imaging-plate area detector. The generator was operated at 50 kV and 100 mA using a focal spot size of 0.3 × 3 mm. Monochromatic Cu $K\alpha$ radiation (λ = 1.5418 Å) was obtained using platinum Yale mirror optics.

The final data used for refinement of both the apo protein and the co-crystals were collected on the ID29 beamline at the European Synchrotron Radiation Facility (ESRF) in Grenoble. The diffraction data were collected from crystals cooled to 100 K after brief immersion into a mixture of 70% mother liquor and 30% glycerol. The data were indexed, integrated and reduced using *MOSFLM* (Leslie & Powell, 2007) and *SCALA* (Evans, 2006). The processed data were converted to structure factors using *TRUNCATE* from the *CCP*4 suite (Winn *et al.*, 2011).

A summary of the data-collection statistics is given in Table 1.

2.4. Structure determination and refinement

Diffraction data for the heavy-atom derivatives were collected as described in the previous section and the mean

fractional isomorphous differences (MFIDs) against the native data set were calculated with the *CCP*4 suite (Winn *et al.*, 2011). For initial location of the heavy-atom sites, difference Patterson maps of the three derivatives (see Table 1) were calculated and interpreted using *SHELX*-97 (Sheldrick, 2008). The platinum derivative was identified as the best of the three and the coordinates of its major sites were used to identify the locations of the principal sites in the other two derivatives in difference Fourier synthesis maps calculated with single isomorphous replacement (SIR) phases using *MLPHARE* and *FFT* from the *CCP*4 package. The location of additional minor sites, phasing and density modification were performed with *autoSHARP* (Bricogne *et al.*, 2003). Model building and modification were carried out with *O* (Jones *et al.*, 1991) and *Coot* (Emsley *et al.*, 2010) and refinement was performed with *REFMAC*5 (Murshudov *et al.*, 2011). During the process of refinement and model building, the quality of the model was checked with *PROCHECK* (Laskowski *et al.*, 1993).

The structure of the cFEL–GlcNAc complex, space group $C2_1$, was solved using the molecular replacement method as implemented in *MOLREP* (Vagin & Teplyakov, 2010). The search probe used was the dimer of the apo form and the solution had an *R* factor of 46.5% for data in the 29.7–1.85 Å resolution interval.

Residual electron density for GlcNAc bound to cFEL was located and modelled in $2F_o - F_c$ and $F_o - F_c$ maps obtained from the co-crystal data refined against the unliganded protein model.

The models were finally subjected to rounds of TLS refinement. Solvent molecules were added to the models in the final stages of refinement according to hydrogen-bond criteria and only if their *B* factors refined to reasonable values and if they improved the R_{free}.

The diffraction data and refinement statistics of the two models are summarized in Table 1.

Superposition of the models matching the secondary structure was performed using the SSM Superposition subroutine of *Coot* (Krissinel & Henrick, 2004). The distances between the ligand and protein atoms were calculated with the *CCP*4 program *CONTACT* (v.1.12.88; T. Skarzynski, Imperial College, London). Figures showing the structures were prepared and rendered with *PyMOL* (http://www.pymol.org).

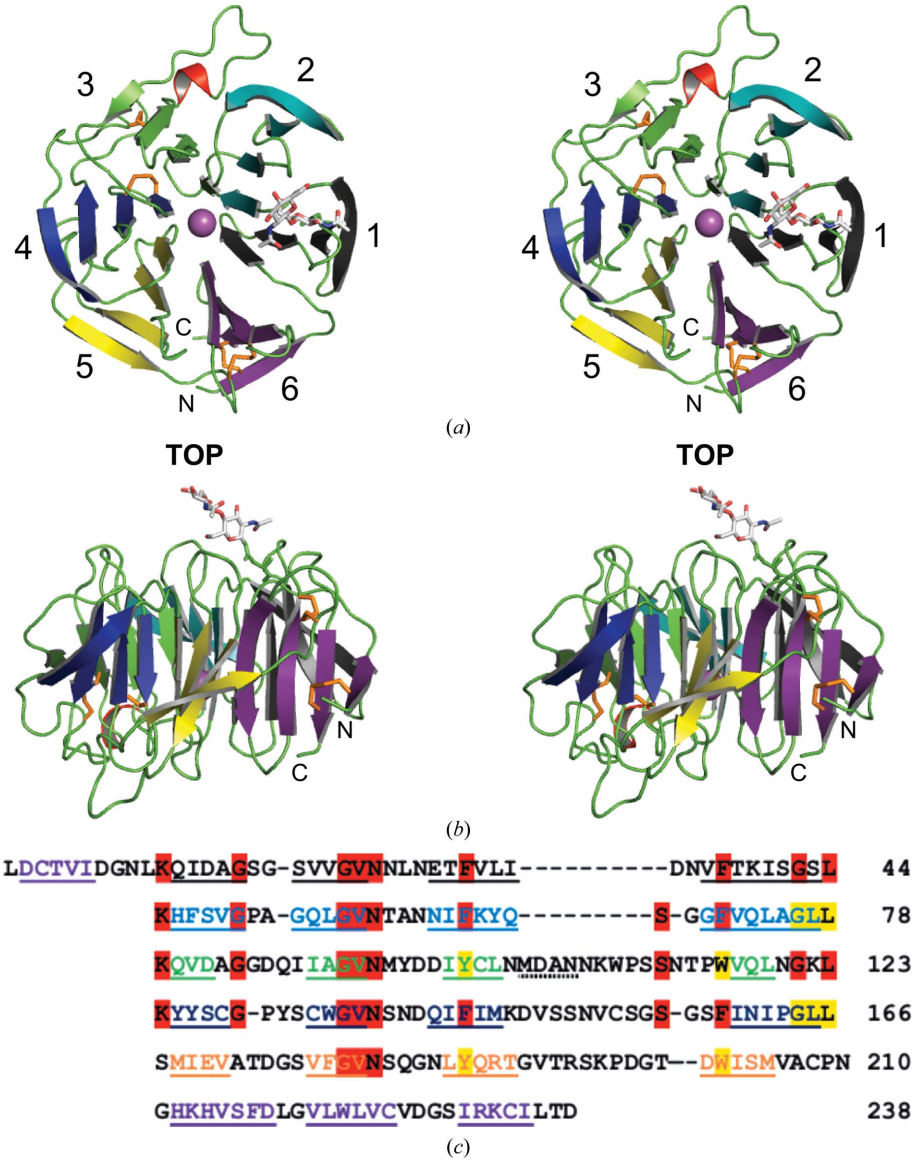


Figure 1
Structure of carp FEL. (*a*) Stereo diagram representing the overall structure of a protomer of carp FEL. The figure was prepared with the coordinates of molecule *A* of the crystals of the apo form and the view is looking down the pseudo-sixfold axis. The blades are labelled from 1 to 6, each consisting of four β-strands, and are represented in different colours. The disulfide bridges are shown in orange and the position of the glycosylated Asn27 is indicated with a stick model of the carbohydrate moiety linked to the amino acid. The calcium ion located in the central water-filled tunnel is represented as a magenta sphere. (*b*) Stereo side view of the carp FEL monomer. The β-strands belonging to the same blade are colour-coded as in (*a*). The pseudo-sixfold axis is in the plane of the figure and is vertical. The N- and C-termini are at the bottom of the figure. (*c*) Amino-acid sequence alignment of the six β-propeller blades of the carp FEL molecule. The amino acids of each of the blades are represented in different colours and the β-strands are underlined using the same colour for each blade. The amino acids in the helix are underlined with a dotted line. The residues that are conserved in at least three blades are in red boxes and those participating in the interactions with the ligand are in yellow boxes. Note the conservation of the GVN motif in the first five blades.

2.5. Fluorometric titrations

Fluorescence measurements were conducted with an FP-8200 spectrofluorometer (Jasco, Easton, Maryland,

USA) with the samples at room temperature. 1 ml of 1 μ*M* cFEL in 20 m*M* Tris–HCl pH 7.5 was titrated with small aliquots of 1 m*M* ligand solution in the same buffer. After each addition, the solution was mixed and left to equilibrate and the emission spectra (λ_{ex} = 295 nm) were recorded in the 310–400 nm range with 2.5 nm width excitation and 5 nm width emission slits. Tryptophan fluorescence quenching was calculated as the percentage decrease of the initial fluorescence. Appropriate dilution corrections were applied to both the titrant and the protein concentrations. The reported values are the average of three independent experiments. Data were analyzed using *SigmaPlot* v.9.0 (Systat Software Inc., San Jose, California, USA) and *Origin*8 (OriginLab Corporation).

3. Results

3.1. Overall architecture of the protomer

We will define and describe the structural elements of the monomer using the orthorhombic crystal form of the apoprotein, which was the first structure that we solved and was determined to 1.35 Å resolution. The final model of the apoprotein comprises 236 amino-acid residues for monomer *A* and 237 for monomer *B* in the asymmetric unit of this crystal form (see Table 1). The maps do not show electron density for one amino acid at the C-terminus of chain *B*, Asp238, and two at the C-terminus of chain *A*, Thr237 and Asp238. The two molecules in the asymmetric unit are related by an approximate noncrystallographic dyad, with an r.m.s.d. of 0.427 Å calculated over 236 C^{α} pairs of equivalent residues. A monomer of carp FEL fits into a cylindrical box of approximately 50 Å in diameter and 30 Å in height. The maps show clear electron density for two GlcNAc moieties in the only glycosylation site present, at Asn27 of both of the two polypeptide chains of the asymmetric unit. The four disulfide bridges, 3–234, 100–153, 128–133 and 208–226, are also very clearly defined in the electron-density maps. Supplementary Fig. S1(*a*) shows a stereoview of the first bridge, which links the N-terminal end of the polypeptide chain to the C-terminal end.

The carp FEL protomer is a single-domain structure organized as a six-bladed β-propeller with a pseudo-sixfold axis and a central water-filled tunnel that runs approximately in the direction of the axis (Fig. 1*a*). Each of the blades consists of a twisted four-stranded antiparallel β-sheet with strands spanning a minimum of three and a maximum of eight amino acids. There is also a short, four-amino-acid α-helix in the chain bridging strands *C* and *D* of the third blade. In our description, we follow the standard convention of labelling the four β-strands in each of the first five blades *A*–*B*–*C*–*D* starting from the N-terminus. Strands *A* are closest to the pseudo-symmetry axis and strands *D* are on the surface of the protein molecule. Supplementary Table S1 summarizes the secondary-structure assignments and Supplementary Fig. S1(*b*) shows a topology diagram of the model. The sixth blade, which closes the ring, is made up of three β-strands that are contiguous in the final portion of the protein chain (*A*–*C*) and a fourth strand (*D*) located on the surface of the molecule, which is the first element of secondary structure in the initial portion of the chain and is connected to the last strand through a disulfide bridge. These N- and C-terminal areas are well ordered in the maps owing to the presence of this bridge linking residues 3 and 234 of the polypeptide chain (Supplementary Fig. S1*a*). Closure of the last blade of a β-propeller by the first strand at the N-terminus is often referred to as a 'velcro closure'. By convention, the top of the molecule is defined as the face in which the *D* strands of each blade connect to the *A* strands of the following blade. The C-terminus of the chain is located at the bottom of the model. A side view of the model is shown in the stereo diagram in Fig. 1(*b*), which uses the same colour code as Fig. 1(*a*) and in which the pseudo-sixfold axis is vertical and in the plane of the figure.

As in many other β-propeller structures, a consensus motif is found in the sequence of the propeller blades of carp FEL. Fig. 1(*c*) represents the sequences of the six blades aligned on the basis of the equivalent β-strands. In the figure, the amino acids that are conserved in at least three blades are shown in red boxes and those that participate in ligand binding are in yellow boxes. Note that the GVN motif in β-strand *B* is conserved in the first five blades.

3.2. Metal binding in the central cavity

During the process of model building and refinement of the apoprotein, a relatively strong peak of electron density was identified in the central channel of the two β-propellers present in the asymmetric unit. The density was too strong to be a solvent molecule, and a more likely candidate was a Ca^{2+} ion, which was present at a concentration of 5 m*M* in the buffer used to homogenize the fish eggs for protein purification. Supplementary Fig. S2(*a*) shows the electron density for the metal and the nearby protein side chains present in monomer *A* of the orthorhombic crystals of cFEL. A standard procedure used to identify trace elements in protein crystals is X-ray fluorescence spectroscopy (Jones *et al.*, 1988), and we therefore recorded the spectrum shown in Supplementary

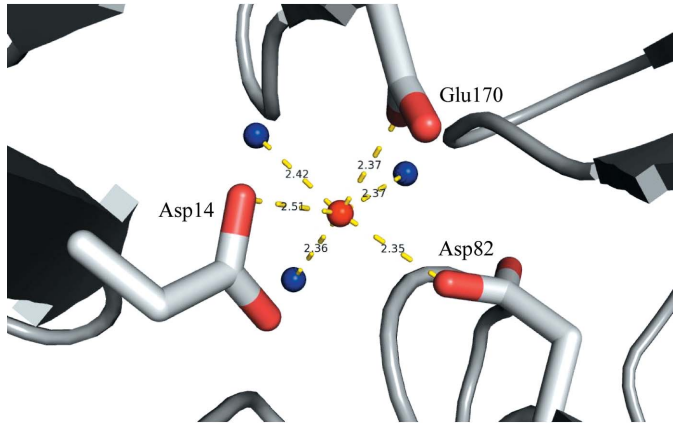


Figure 2
A calcium ion in the central channel: details of the Ca^{2+}-ion coordination. The sixfold axis is approximately perpendicular to the plane of the figure. The Ca^{2+} ion is shown in red and the solvent molecules are shown in blue. The distances indicated are in Å and correspond to those listed in Supplementary Table S2.

Fig. S2(*b*), exciting the sample with monochromatic synchrotron radiation of 12.7 keV. In addition to the expected sulfur, three other elements were detected in the sample: Ca, Fe and Ti. The unexpected signal for the last two has also been observed by others and was attributed to the interaction of the beam with beamline or sample-holder components (Sanz-Gaitero *et al.*, 2014).

The Ca^{2+} ion is located in the funnel-shaped cFEL channel closer to the bottom rather than the top, as defined above, and is coordinated by three acidic residues and three water molecules. The three residues are Asp14 in the middle of strand *A* of the first blade, Asp82 at the end of strand *A* of blade 3 and Glu170 towards the end of strand *A* of blade 5. The three coordinated residues are thus located in strands *A* of the odd-numbered blades and the geometry of the coordinated ion is thus very symmetrical. Supplementary Table S2 lists the shortest distances between the Ca^{2+} ion and amino-acid side chains and water molecules for the two sites present in the asymmetric unit of the orthorhombic form and Fig. 2 represents the Ca^{2+} ion and its ligands. Note that the equivalent distances in the two molecules are very similar and agree reasonably well with the experimentally determined values for calcium(II) coordination (Harding, 2006; Harding & Hsin, 2014; http://mespeus.bch.ed.ac.uk/MESPEUS_10/). Equally well defined density is observed in the six protomers present in the monoclinic crystal form of the co-crystals with GlcNAc.

The occupancy of the Ca^{2+} site appears to be complete and the temperature factors of the eight ions that we have examined in the two crystal forms refine to values ranging between 4 and 5 Å^2 in the orthorhombic form and 6 and 11 Å^2 in the monoclinic crystals, which diffracted to lower resolution.

3.3. Carp FEL is a dimer

Dynamic light-scattering experiments yield a hydrodynamic diameter consistent with the presence of a dimer in solution. This result is at variance with our previous report that cFEL was a monomer, which was based on using a calibrated gel-filtration column to estimate protein size (Galliano *et al.*, 2003). However, our DLS and crystallographic data give a different and unambiguous answer, *i.e.* the lectin is a dimer under normal conditions. Fig. 3(*a*) shows a scattered intensity plot of two different concentrations of carp FEL. The highest peak corresponds to a hydrodynamic diameter that agrees very well with the value measured for the dimer present in the asymmetric unit of the orthorhombic crystal form. The second peak, explained by the presence of high-molecular-weight aggregates, becomes insignificant in percentage volume and number of particles plots.

Since the monoclinic crystal form of the co-crystals of the protein, which contains six protomers in the asymmetric unit, was solved using the asymmetric unit of the unliganded protein, *i.e.* a dimer, as a search probe, we can be confident that the relationship between protomers in the two crystal forms is the same. In the PDB file of this second crystal form the dimers are formed by protomers *AB*, *CD* and *EF*.

We have also calculated the solvent-accessible area of the protomers and the dimer in the orthorhombic crystals. The *A* monomer of carp FEL has a solvent-accessible area of 8774 Å^2 and the *B* monomer 8555 Å^2. The contact areas are 2574 and 2510 Å^2, *i.e.* about 29% of the total surface of the monomers and a value that is comparable with those reported for other physiologically relevant dimers (Winn *et al.*, 2011; Jones & Thornton, 1995; Ponstingl *et al.*, 2000).

The contacts between two protomers in the dimer related by an approximate twofold axis are established through the first three blades and the sixth blade. A key residue participating in the interactions is Asp34, the N atom of which interacts with the OH of Tyr66 and the side chain of which is in contact with the N and the OG atoms of Ser41 of the opposite protomer. The same OH of Tyr66 is also hydrogen-bonded to the O atom of Leu32 of the other member of the dimer.

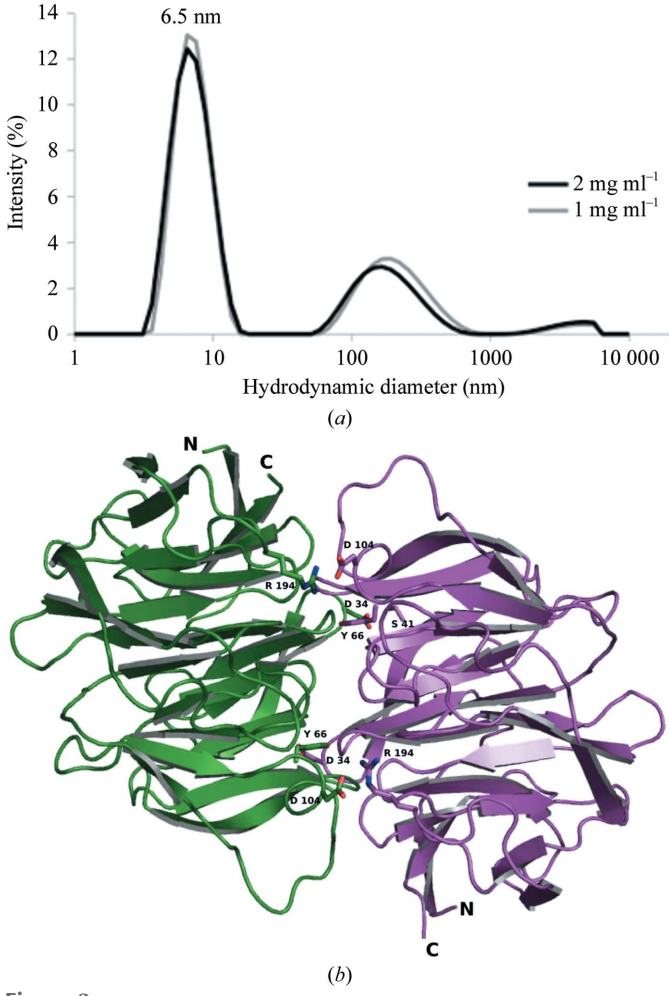


Figure 3
The carp FEL dimer. (*a*) Dynamic light-scattering experiment on carp FEL. The size distribution is by intensity and two curves corresponding to two different concentrations were recorded. The small high-molecular-weight peak corresponds to the presence of a low percentage of aggregates. The hydrodynamic diameter of the highest peak is consistent with the presence of a dimer in solution. (*b*) Ribbon representation of the carp FEL dimer. The approximate twofold axis present in the dimer is roughly perpendicular to the plane of the figure. Note the positions of the N- and C-termini and the areas of the protomers where contacts are established.

Another important contact is between the two side chains of Asp86, one in each monomer, which are also in contact with the side chains of the next residue in the sequence, Gln87, of the other chain in the dimer. Two salt bridges are established between the NH1 and NH2 atoms of Arg194 of one protomer and the OD1 and OD2 atoms of Asp104 of the other protomer in the dimer. Supplementary Table S3 lists the main distances shorter than about 3.5 Å measured between atoms in the interacting protomers, and Fig. 3(*b*) shows a ribbon representation of the dimeric molecule with the approximate twofold axis normal to the plane of the figure and stick models of some of the side chains of the interacting amino acids listed in Supplementary Table S3.

3.4. Ligand binding to carp FEL

Two ligand-binding sites were identified in each of the six monomers of carp FEL present in the asymmetric unit of the monoclinic crystals. They will be labelled α and β; the former, which is closer to the N-terminus and is located in the cleft between the second and the third blades, engages amino acids mostly from the third blade, while the latter, which is closer to the C-terminus and is located in the cleft between the fourth and the fifth blades, requires the participation of residues mostly from the fifth blade. The two sites can bind the two anomeric forms of *N*-acetylglucosamine and site α binds polyethylene glycol, a precipitant present at high concentration in the mother liquor, which is found in the crystals bridging two sites in monomers related by crystallographic symmetry.

Fig. 4(*a*) shows the electron density of the ligands that bind to the four sites of the dimer made up of protomers *E* and *F*. Protomer *E*, represented in green in the figure, binds one molecule of *N*-acetyl-β-D-glucosamine (βGlcNAc, NAG) at site β and one molecule of polyethylene glycol (PEG) at site α. The other protomer in the dimer, *F*, represented in magenta in

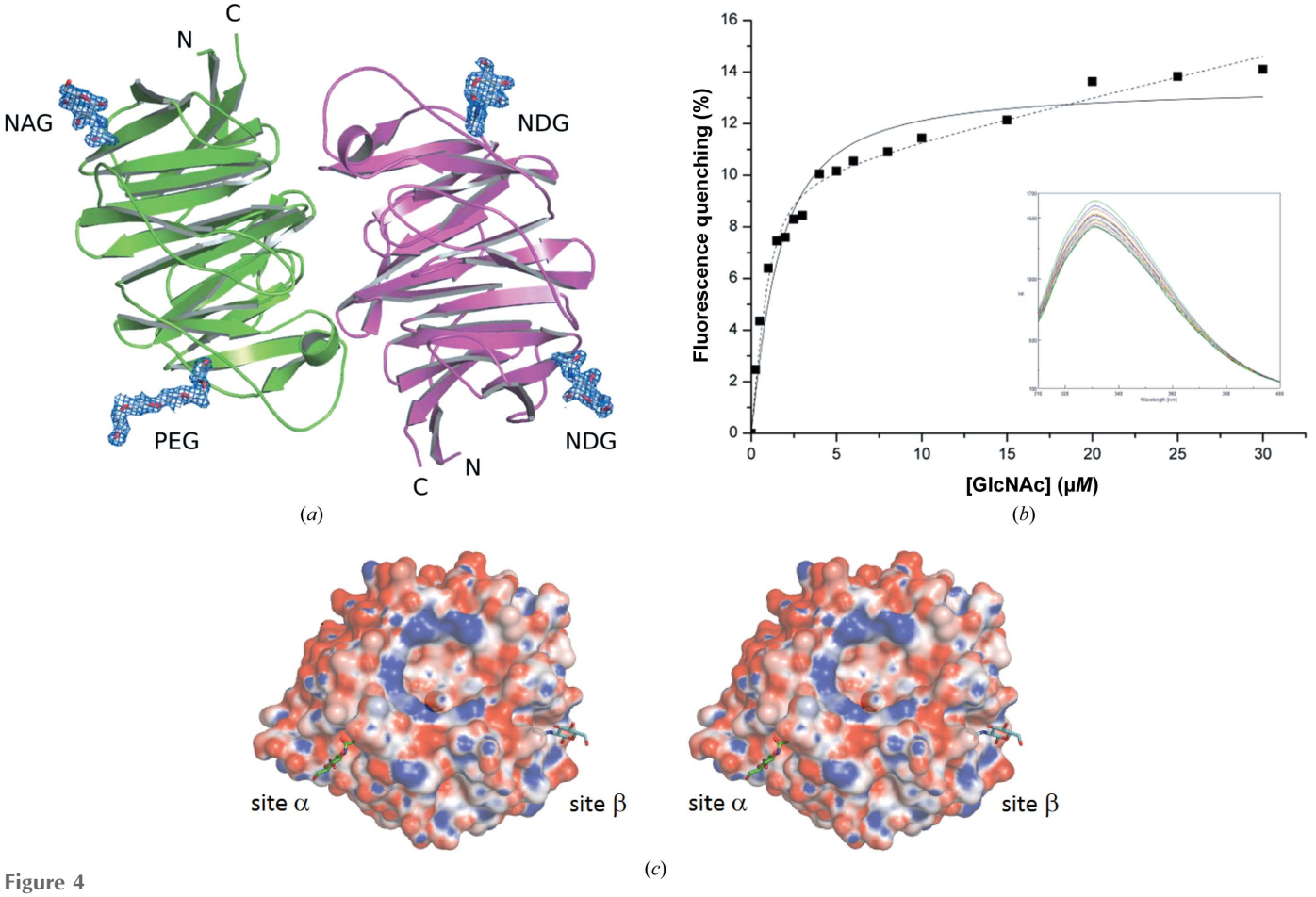


Figure 4
Ligand binding to carp FEL. (*a*) Ribbon representation of the carp FEL *EF* dimer showing the four ligand-binding sites occupied by the three ligands. Protomer *E* (green) binds one molecule of *N*-acetyl-β-D-glucosamine (βGlcNAc) in site β and one molecule of PEG in site α, whereas protomer *F* (magenta) binds two molecules of the α-anomer, *N*-acetyl-α-D-glucosamine (αGlcNAc), in both sites. The electron density of the $2F_o - F_c$ map was contoured at the 1.0σ level. (*b*) Fluorometric titration of carp FEL with *N*-acetylglucosamine. The titration curve was obtained by adding 0.5 µl injections of 1 m*M* *N*-acetylglucosamine solution to a 1 µ*M* sample of carp FEL in 20 m*M* Tris–HCl pH 7.5. The solid line was drawn by fitting the data with a single binding-site model (*i.e.* two identical sites per cFEL protomer) and yields a K_d value of 1.0 ± 0.19 m*M*. Fitting the data with a modified equation that also takes into account the linear contribution of nonspecific quenching owing to the presence of unbound ligand yields the dotted curve. In this case the calculated K_d value is 0.3 ± 0.09 m*M*. The inset shows the variation of the spectra upon ligand addition. (*c*) Stereo drawing of an electrostatic surface representation of the *C* protomer of the monoclinic crystal form showing the relative position of the two ligand-binding sites and the areas involved in the interaction with βGlcNAc. Negative potential is shown in red and positive potential in blue. The red sphere in the channel is the Ca^{2+} ion.

the figure, binds a molecule of *N*-acetyl-α-D-glucosamine (αGlcNAc, NDG) at both binding sites.

The two anomers of *N*-acetylglucosamine present in our crystals appear to have no preference for one site or the other, and we have observed that they can bind at both sites in different protomers. Supplementary Table S4 summarizes the occupancies of the six protomers. Note that the sites occupied by PEG do not show clear density for solvent molecules, whereas those occupied by the two anomers of *N*-acetylglucosamine do and one of them is conserved in all of the sites examined.

Fig. 4(*b*) shows a representative fluorometric titration of a carp FEL sample with GlcNAc. The figure represents the percentage fluorescence quenching obtained on titrating 1 ml carp FEL (1 μM in 20 m*M* Tris–HCl pH 7.5) with small aliquots of 1 m*M* *N*-acetylglucosamine in the same buffer. Fitting of the data assuming a single binding-site model (*i.e.* two identical sites per cFEL protomer) gave a K_d value of 1.0 $\pm$ 0.19 m*M* (the continuous line in the figure). A slightly better fit can be obtained using a modified equation that also takes into account the linear contribution from nonspecific quenching owing to the presence of unbound ligand (the dashed line in the figure). In this case, the calculated K_d value is 0.3 $\pm$ 0.09 m*M*. Given the high degree of structural similarity between the two binding sites present in each cFEL monomer, it is reasonable to assume that the two sites have an equal or a very similar dissociation constant.

Carp FEL can also bind other monosaccharides and disaccharides, although with lower affinity. Supplementary Fig. S3(*a*) shows fluorometric titrations of the protein with *N*-acetylglucosamine (black squares), *N*-acetylgalactosamine (GalNAc; blue triangles), *N*,*N*-diacetylchitobiose (red circles), glucose (cyan triangles) and mannose (magenta triangles). In a previous report (Galliano *et al.*, 2003), we have shown that D-GlcNAc and D-GalNAc can, to some extent, inhibit the binding of carp FEL to *Escherichia coli* and *Staphylococcus aureus* cells, while non-acetylated sugars have very little or no effect in this competition assay. The tryptophan fluorescence-quenching titrations that we present here show that D-GlcNAc, *N*,*N*-diacetylchitobiose (a dimer of D-GlcNAc) and also D-GalNAc can bind to cFEL with almost the same affinity. This fact can be explained by examining the binding mode of the acetylated sugars to cFEL. The main contacts involve O1 and O3, in addition to the acetylated N2, while the epimeric hydroxyl (O4) that distinguishes D-GlcNAc from D-GalNAc points out into the solvent and is not engaged in any interactions with the protein. Glucose and mannose instead give a quenching curve that is more similar to a nonspecific effect (*i.e.* they do not bind or bind with a very low affinity).

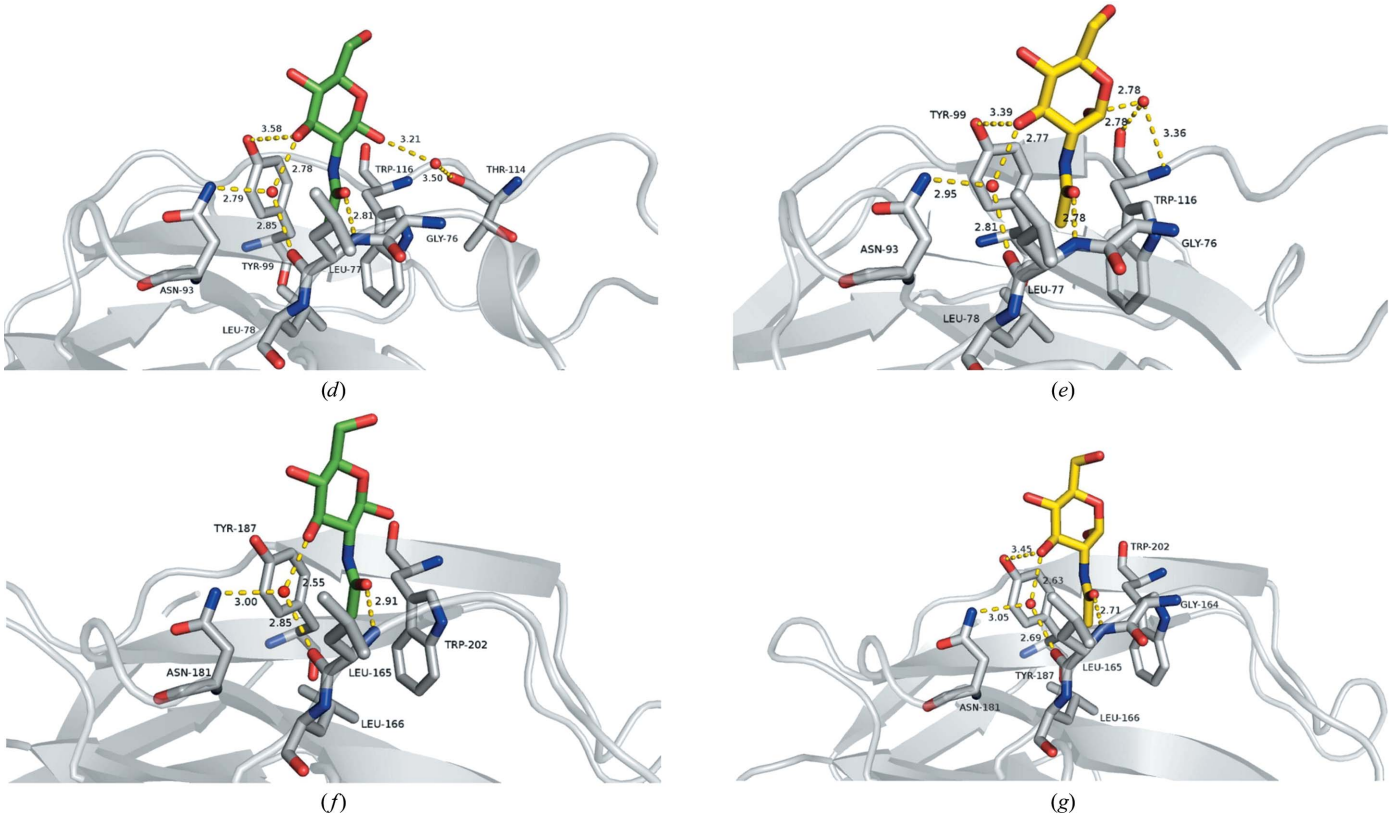


Figure 4 (continued)
(*d*) Details of the interaction of carp FEL with βGlcNAc in site α. The most relevant distances are indicated in Å. In (*d*)–(*g*) the β-anomer of *N*-acetylglucosamine is represented in green and the α-anomer in yellow. Protomer *C* was used to represent the binding of βGlcNAc to sites α and β and protomer *B* was used to represent the binding of αGlcNAc. (*e*) Details of the interaction of carp FEL with αGlcNAc in site α. The most relevant distances are indicated in Å. (*f*) Details of the interaction of carp FEL with βGlcNAc in site β. (*g*) Details of the interaction of carp FEL with αGlcNAc in site β.

Fig. 4(*c*) shows a stereo electrostatic surface representation of the *C* protomer of the monoclinic crystal form, showing the relative position of the two ligand-binding sites and the areas involved in the interaction with *N*-acetyl-β-D-glucosamine. Supplementary Fig. S3(*b*) shows a stereo line drawing of the C^{α} chain trace of the same protomer instead showing the main residues involved in interaction with the carbohydrate. The magenta sphere is the Ca^{2+} ion in the central cavity, which does not play any role in the interactions with the ligands but probably has a structural function. A list of selected distances between ligand and cFEL residues is given in Table 2. The interaction of the acetylated amino group of the carbohydrate appears to be very important since glucose and glucosamine do not bind significantly to the lectin. Note the approximate equivalence of Leu77, Tyr99 and Trp116 in site α to Leu165, Tyr187 and Trp202 in site β. The N atoms of the two leucines makes a hydrogen bond to O7 of the monosaccharides. It is also worth noticing that the presence of the aromatic amino acids (in particular Trp) at the binding sites justifies the use of the fluorescence-quenching percentage plots in the fluorometric titrations. Fig. 4(*d*) represents the side chains of the residues involved in the interactions of the β-anomer of *N*-acetylglucosamine in site α as stick models. Note the role played by a water molecule, which is conserved in all of the binding sites, at approximately the same distance from the ND2 atom of Asn93, the O atom of Leu77 and O3 of NGA. A second water molecule, found only interacting with the β-anomer of the ligand, bridges the O atom of Thr114 and O1 of the carbohydrate.

The binding of the α-anomer of *N*-acetylglucosamine in site α is represented in Fig. 4(*e*) and the most significant ligand–protein distances are also given in Table 2. It is evident that the two anomers bind in a very similar way and the distances are almost the same, with the only important difference being in O1 of the monosaccharide, which, as it points in a different direction, cannot be connected to Thr114 through a water molecule. The connection of O1 to the binding site is instead established through another water molecule to the O and N atoms of Trp116 as shown in Fig. 4(*e*). The interactions of the two anomers in site β are represented in Figs. 4(*f*) and 4(*g*).

A glycan-array screening performed with fluorescently labelled carp FEL by the Consortium for Functional Glycomics (https://www.functionalglycomics.org) did not identify any complex carbohydrates that bind effectively to the lectin.

Table 2
Selected distances between the closest carp FEL residues and the βGlcNAc (*N*-acetyl-β-D-glucosamine) and αGlcNAc (*N*-acetyl-α-D-glucosamine) molecules in the two different binding sites.

The distances refer to protomers *C* (for βGlcNAc) and *B* (for αGlcNAc) of the monoclinic crystals. Site α is closer to the N-terminus (Leu77, Tyr99 and Trp116) and site β is closer to the C-terminus (Leu165, Tyr187 and Trp202). Only one distance per residue has been included for each carbohydrate atom in the table.

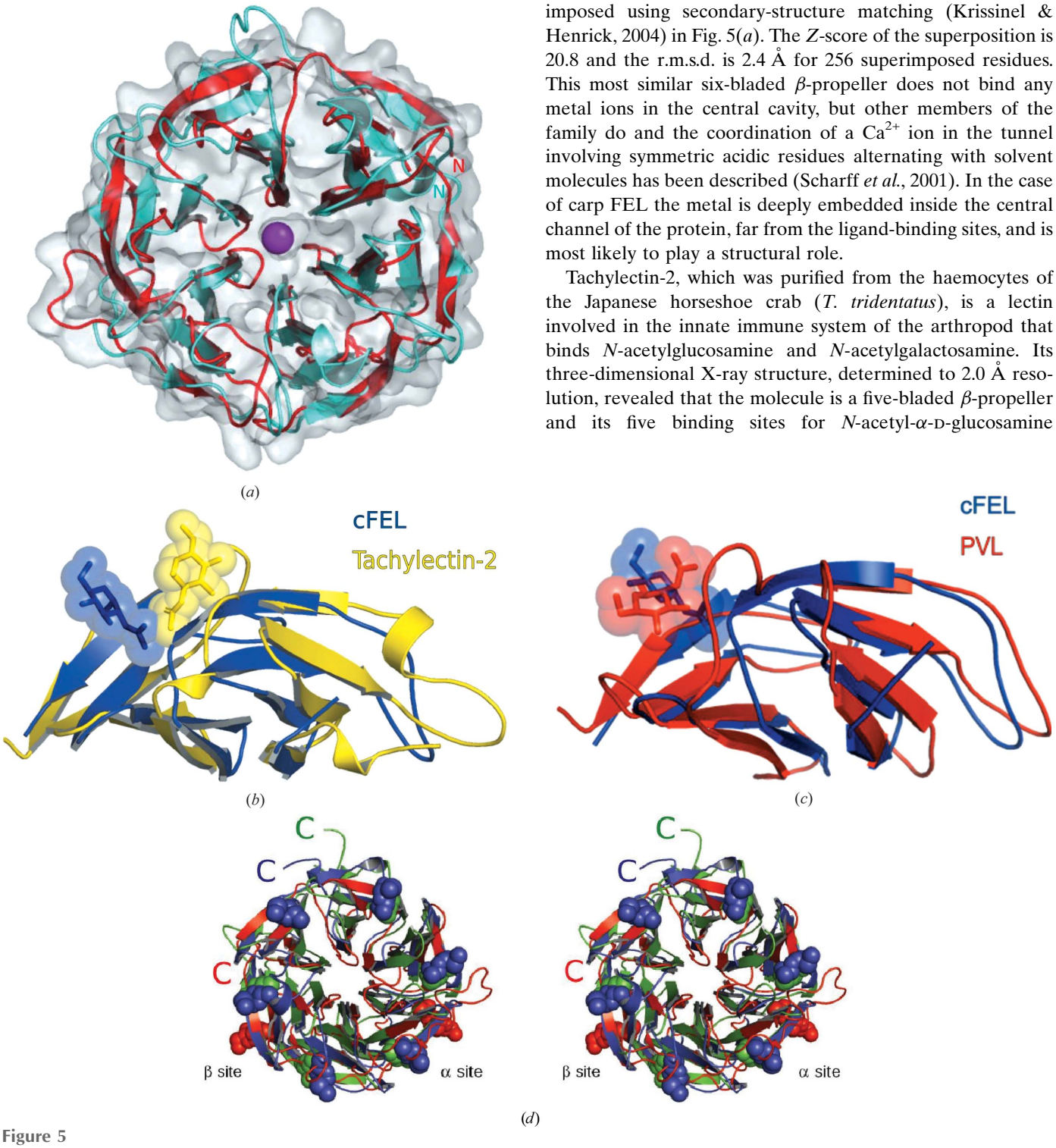
Carp FEL residue	Atom	β-GlcNAc atom	Distance (Å)
βGlcNAc			
Leu77	N	O7	2.81 (α)
Tyr99	CE1	N2	3.60 (α)
Tyr99	CZ	C3	3.81 (α)
Tyr99	OH	O3	3.58 (α)
Trp116	CD1	O1	3.56 (α)
Trp116	CD1	C7	3.78 (α)
Trp116	CD1	C8	3.65 (α)
Leu165	N	O7	2.91 (β)
Tyr187	CE1	N2	3.48 (β)
Tyr187	CZ	C3	3.86 (β)
Tyr187	OH	O3	3.66 (β)
Trp202	CB	O1	3.62 (β)
Trp202	CD1	C7	3.92 (β)
Trp202	CD1	C8	3.83 (β)
Carp FEL residue	**Atom**	**αGlcNAc atom**	**Distance (Å)**
Leu77	N	O7	2.78 (α)
Tyr99	CE1	N2	3.44 (α)
Tyr99	CZ	C3	3.78 (α)
Tyr99	OH	O3	3.39 (α)
Trp116	O	O1L	3.95 (α)
Trp116	CD1	C7	3.85 (α)
Trp116	CD1	C8	3.77 (α)
Leu165	N	O7	2.86 (β)
Tyr187	CE1	N2	3.60 (β)
Tyr187	CZ	C3	3.92 (β)
Tyr187	OH	O3	3.67 (β)
Trp202	O	O1L	3.76 (β)
Trp202	CD1	C7	3.86 (β)
Trp202	CD1	C8	3.80 (β)

4. Discussion

Although it was anticipated that carp FEL was likely to belong to the family of β-propeller proteins, no structure was available of a member of the family with significant amino-acid sequence identity. Structures of proteins with sequence homology, members of the arthropod innate immune system or molecules with a probable function in fish embryo development, have not yet been determined experimentally. Therefore, the structure of cFEL was solved using the standard MIR method and a first important result of this work is that there now exists a prototype of the two protein families with similar amino-acid sequences to cFEL, the crystal structures of which can probably be phased by molecular replacement using our cFEL model as search probe.

Carp FEL folds as a β-propeller, a versatile structure of cylindrically shaped molecules with a number of similar blades ranging in number from four to ten (Chen *et al.*, 2011). A feature found in many β-propeller proteins is the existence of short consensus sequences present in all of the blades; for six-bladed β-propellers the consensus sequence YWTD has been shown to occur in groups of six contiguous repeats (Springer, 1998). This motif is not present in carp FEL, but the sequence GVN is easily identified in five blades and GV in all six blades of the molecule (see Fig. 1*c*).

Six-bladed β-propellers are very diffuse in nature and show a very high degree of functional variability. In fact, when the model of cFEL is compared with the structures available in the Protein Data Bank using the *DALI* server (Holm & Sander, 1999), the molecule identified as structurally most similar to cFEL is the sensor domain of the *Mycobacterium tuberculosis* receptor Ser/Thr protein kinase, an extracellular domain of the enzyme that does not seem to belong to the lectin family (Good *et al.*, 2004). The two models are represented super-



imposed using secondary-structure matching (Krissinel & Henrick, 2004) in Fig. 5(*a*). The *Z*-score of the superposition is 20.8 and the r.m.s.d. is 2.4 Å for 256 superimposed residues. This most similar six-bladed β-propeller does not bind any metal ions in the central cavity, but other members of the family do and the coordination of a Ca^{2+} ion in the tunnel involving symmetric acidic residues alternating with solvent molecules has been described (Scharff *et al.*, 2001). In the case of carp FEL the metal is deeply embedded inside the central channel of the protein, far from the ligand-binding sites, and is most likely to play a structural role.

Tachylectin-2, which was purified from the haemocytes of the Japanese horseshoe crab (*T. tridentatus*), is a lectin involved in the innate immune system of the arthropod that binds *N*-acetylglucosamine and *N*-acetylgalactosamine. Its three-dimensional X-ray structure, determined to 2.0 Å resolution, revealed that the molecule is a five-bladed β-propeller and its five binding sites for *N*-acetyl-α-D-glucosamine

Figure 5
Comparison of the structure and ligand binding of carp FEL and other proteins. (*a*) Comparison of the model of carp FEL (molecule *A* of the orthorhombic crystals; red) and the model of the sensor domain of the *M. tuberculosis* receptor Ser/Thr protein kinase (molecule *A*, PDB entry 1rwi; Good *et al.*, 2004; blue). The models were superimposed using *LSQKAB* (Kabsch, 1976). The transparent grey surface represents the approximate volume of cFEL. The magenta sphere is the Ca^{2+} ion bound to carp FEL. (*b*) Superposition of the fourth and fifth blades of carp FEL (residues 122–206; molecule *A* of the orthorhombic crystals; blue) and the second and third blades of tachylectin-2 (PDB entry 1tl2; residues 85–178; Beisel *et al.*, 1999; yellow). The two space-filling and stick models of αGlcNAc are represented in the same colour as the lectin that they are bound to. (*c*) Superposition of the fourth and fifth blades of carp FEL (residues 122–206; molecule *A* of the orthorhombic crystals; blue) and the second and third blades of *P. velutina* lectin (PDB entry 2c4d; residues 59–158; Cioci *et al.*, 2006; red). The two space-filling and stick models of αGlcNAc are represented in the same colour as the lectin that they are bound to. (*d*) Stereo diagram superimposing the model of the *C* protomer of carp FEL (red) on AAL (*A. aurantia* lectin; molecule *A*, PDB entry 1ofz; Wimmerova *et al.*, 2003; green) and AFL (*A. fumigatus* lectin; molecule *A*, PDB entry 4agi; Houser *et al.*, 2013; blue). The space-filling models of the carbohydrates are represented in the same colour as the lectin that they are bound to. Note the similarity in the mode of binding of the saccharide to the two fucose-specific lectins and the different positions of the α and β ligand-binding sites of cFEL.

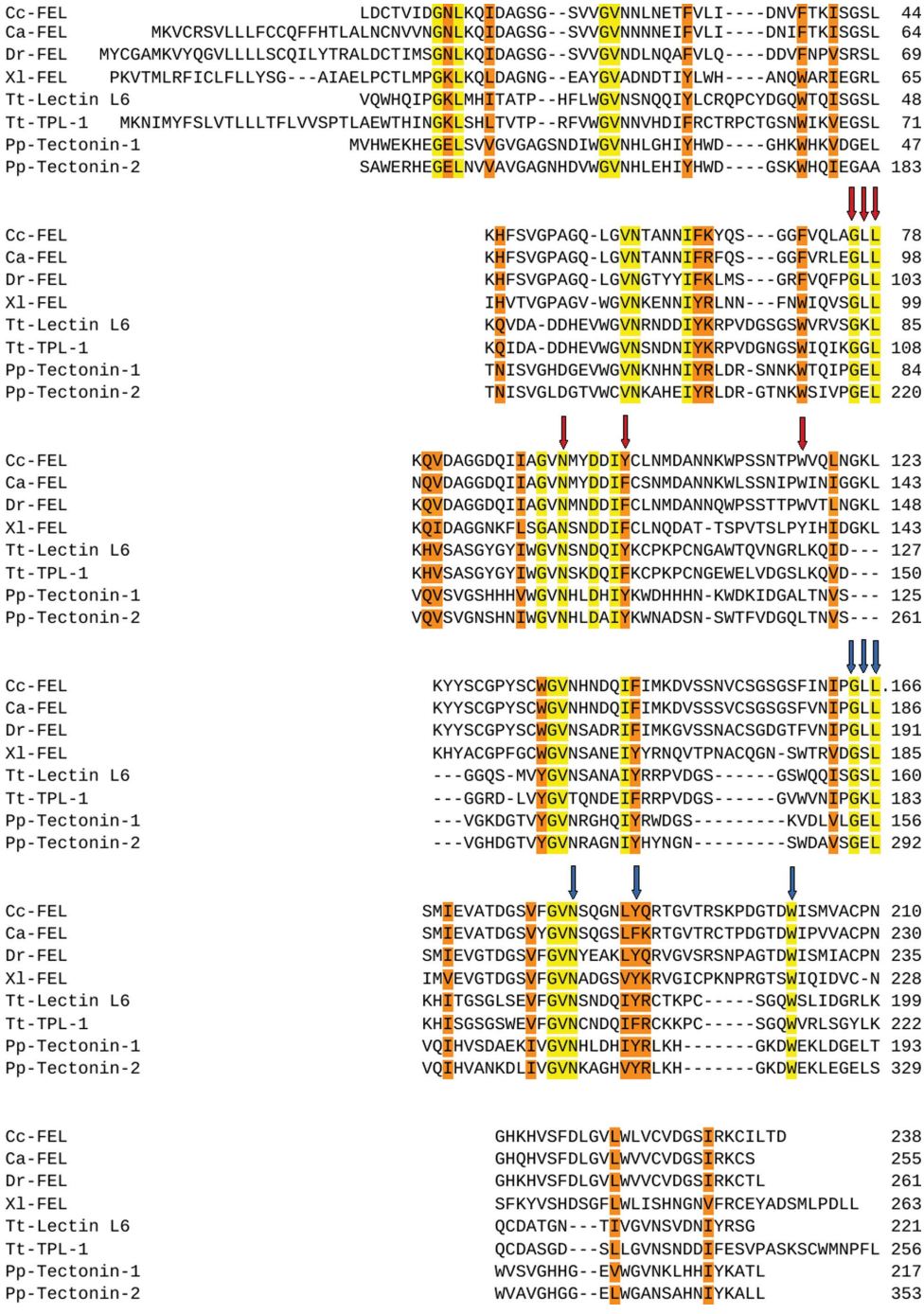


Figure 6
Amino-acid sequence homology of carp FEL: comparison with proteins encoded by homologous genes in fish and amphibians and with members of the invertebrate lectin family. The sequences were aligned using *ClustalW* (Thompson *et al.*, 1994; Larkin *et al.*, 2007) and correspond to the following lectins (UniProt codes are given in parentheses): Cc-FEL (P68512), carp (*Cyprinus carpio*) FEL; Ca-FEL (Q78BR1), goldfish (*Carassius auratus*) FEL; Dr-FEL (A8E5B4), zebrafish (*Danio rerio*) FEL; Xl-FEL (A8E602), African clawed frog (*Xenopus laevis*) FEL; Tt-Lectin L6 (P82151), Japanese horseshoe crab (*Tachypleus tridentatus*) lectin L6; Tt-TPL-1 (Q9NB63), galactose-binding protein from Japanese horseshoe crab (*T. tridentatus*); Pp-Tectonin-1 (O61063), tectonin-1 from the slime mould *Physarum polycephalum*; Pp-Tectonin-2 (O61064, amino acids 138–353), tectonin-2 from the slime mould *P. polycephalum*. The six blocks correspond to the six blades of carp FEL. The residues that are conserved in all of the sequences are contained in yellow boxes, while those that have strongly similar properties (scoring >0.5 in the Gonnet PAM 250 matrix; Gaston *et al.*, 1992) are contained in orange boxes. Red arrows identify the residues of cFEL involved in ligand binding in site α and blue arrows those in site β.

(αGlcNAc) have been described in detail (Beisel *et al.*, 1999).

In order to compare the binding mode of the same carbohydrate (αGlcNAc) to tachylectin-2 and cFEL, we superimposed the fourth and fifth blades of carp FEL (residues 122–206) on the second and third blades of tachylectin-2 (residues 85–178). The models are shown in Fig. 5(*b*), with cFEL represented in blue and tachylectin-2 in yellow. Although the two blades superimpose rather well, it is evident that the mode of binding of αGlcNAc to the two proteins is very different. In the figure, the two stick and space-filling models of αGlcNAc are represented in the same colour as the lectin that they are bound to. We have also compared the binding mode of GlcNAc with that of a seven-bladed β-propeller, the fungal lectin PVL (*Psathyrella velutina* lectin), which is specific for *N*-acetylglucosamine and *N*-acetylneuraminic acid (Cioci *et al.*, 2006). Fig. 5(*c*) represents the fourth and fifth blades of carp FEL superimposed on the second and third blades of PVL (residues 59–158). Again it is evident that the mode of binding of the carbohydrate to the two lectins is quite different.

There are, however, lectins with a six-bladed β-propeller fold, and those that have been structurally characterized in detail are specific for fucose. The first to be studied by X-ray diffraction was a fungal lectin from the orange-peel mushroom *Aleuria aurantia* (AAL) which, because of its specificity of binding to fucose-containing saccharides, is widely used in the fractionation of glycoproteins (Wimmerova *et al.*, 2003; Fujihashi *et al.*, 2003). Structural information on a second fucose-specific lectin is more recent and concerns the lectin from the opportunistic pathogen *Aspergillus fumigatus*

(AFL; Houser *et al.*, 2013). Interestingly, a six-bladed β-propeller with an affinity for fucose can also be formed by the association of three shorter chains each containing two blades and two carbohydrate-binding sites (Kostlánová *et al.*, 2005; Audfray *et al.*, 2012).

Fig. 5(*d*) shows a stereo diagram superimposing the model of the *C* protomer of carp FEL (red) on AAL (green) and AFL (blue), with space-filling models of the carbohydrates represented in the same colour as the lectin that they are bound to. In spite of the different number of occupied sites in AAL (five) and AFL (six), the similarities in the positions of the sites as well as the mode of binding of fucose is evident in the figure, which shows how the carbohydrates occupy almost the same volumes in the two lectins. In the figure, the bottom of the molecule is closer to the observer and the very different position of the α and β sites of cFEL occupied by *N*-acetyl-β-D-glucosamine (red) highlights the fact that the carbohydrate binds on the other side of the β-propeller in cFEL.

Carp FEL shows sequence similarities to lectins believed to either play a role in innate immunity (Saito *et al.*, 1995) or in embryonic development (Xie *et al.*, 2001).

Fig. 6 compares the sequence of cFEL with those of proteins encoded by homologous genes in fish and amphibians and with the members of the invertebrate lectin family, the tachylectins. The sequences were aligned using *ClustalW* (Thompson *et al.*, 1994; Larkin *et al.*, 2007) and correspond to the following lectins (the abbreviations used in Fig. 6 and the UniProt codes are given in parentheses): carp (*C. carpio*) FEL (Cc-FEL; P68512), goldfish (*Carassius auratus*) FEL (Ca-FEL; Q78BR1), zebrafish (*Danio rerio*) FEL (Dr-FEL; A8E5B4), African clawed frog (*Xenopus laevis*) FEL (Xl-FEL; A8E602), Japanese horseshoe crab (*T. tridentatus*) lectin L6 (Tt-Lectin L6; P82151), Japanese horseshoe crab (*T. tridentatus*) galactose-binding protein (Tt-TPL-1; Q9NB63), tectonin-1 from the slime mould *P. polycephalum* (Pp-Tectonin-1; O61063) and tectonin-2 from the slime mould *P. polycephalum* (Pp-Tectonin-2; O61064, amino acids 138–353).

The residues that are conserved in all of the sequences are shown in yellow boxes in Fig. 6, while those that have strongly similar properties (with a score of >0.5 in the Gonnet PAM 250 matrix) are shown in orange boxes. Red arrows identify the residues of cFEL involved in ligand binding in site α and blue arrows those in site β.

Lectin L6 from the Japanese horseshoe crab (*T. tridentatus*; UniProt code P82151) shows a sequence identity with cFEL of about 36% over a 193-amino-acid overlap, while caFEL, which was identified in goldfish (*C. auratus*) as a 'putative membrane protein' (UniProt code Q78BR1), has a sequence identity with cFEL of about 86% over an overlap of 234 amino acids. This higher degree of sequence similarity is not surprising given the closer proximity of the two species in evolution. While there is no question that the two families represented by these two proteins will also reveal six-bladed β-propeller structures, the very high level of sequence identity to the second protein makes it likely that cFEL may have the same or a similar function.

The goldfish (*C. auratus*) is a triploid bisexual species that can reproduce both sexually and by gynogenesis. In the latter case, the sperm of a related species triggers embryogenesis but does not make any genetic contribution to the egg. This species thus has two types of eggs and the gene coding for the protein with a sequence similar to cFEL is expressed differently in the two types of egg (Xie *et al.*, 2001). In addition, the protein similar to cFEL is totally specific to oocytes and thus it is expected to be likely to play a role in the development of the embryo, a function that is anticipated to be mediated by its interactions with carbohydrates. The probable role of cFEL in embryonic development undoubtedly deserves further exploration.

Acknowledgements

We thank the staff of the ESRF in Grenoble (Proposal MX 1640) for assistance during data collection and the Consortium for Functional Glycomics (grant No. R24 GM098791) for the glycan-array screening of carp FEL. This work was supported by Fondazione Cassa di Risparmio di Verona, Vicenza, Belluno e Ancona.

References

Audfray, A., Claudinon, J., Abounit, S., Ruvoën-Clouet, N., Larson, G., Smith, D. F., Wimmerová, M., Le Pendu, J., Römer, W., Varrot, A. & Imberty, A. (2012). *J. Biol. Chem.* **287**, 4335–4347.
Beisel, H. G., Kawabata, S., Iwanaga, S., Huber, R. & Bode, W. (1999). *EMBO J.* **18**, 2313–2322.
Bricogne, G., Vonrhein, C., Flensburg, C., Schiltz, M. & Paciorek, W. (2003). *Acta Cryst.* D**59**, 2023–2030.
Chen, C. K.-M., Chan, N.-L. & Wang, A. H.-J. (2011). *Trends Biochem. Sci.* **36**, 553–561.
Chen, S.-C., Yen, C.-H., Yeh, M.-S., Huang, C.-J. & Liu, T.-Y. (2001). *J. Biol. Chem.* **276**, 9631–9639.
Cioci, G., Mitchell, E. P., Chazalet, V., Debray, H., Oscarson, S., Lahmann, M., Gautier, C., Breton, C., Perez, S. & Imberty, A. (2006). *J. Mol. Biol.* **357**, 1575–1591.
Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). *Acta Cryst.* D**66**, 486–501.
Evans, P. (2006). *Acta Cryst.* D**62**, 72–82.
Fujihashi, M., Peapus, D. H., Kamiya, N., Nagata, Y. & Miki, K. (2003). *Biochemistry*, **42**, 11093–11099.
Galliano, M., Minchiotti, L., Campagnoli, M., Sala, A., Visai, L., Amoresano, A., Pucci, P., Casbarra, A., Cauci, M., Perduca, M. & Monaco, H. L. (2003). *Biochem. J.* **376**, 433–440.
Gaston, H., Gonnet, M., Cohen, A. & Benner, S. (1992). *Science*, **256**, 1443–1445.
Good, M. C., Greenstein, A. E., Young, T. A., Ng, H.-L. & Alber, T. (2004). *J. Mol. Biol.* **339**, 459–469.
Harding, M. M. (2006). *Acta Cryst.* D**62**, 678–682.
Harding, M. M. & Hsin, K.-Y. (2014). *Methods Mol. Biol.* **1091**, 333–342.
Holm, L. & Sander, C. (1999). *Nucleic Acids Res.* **27**, 244–247.
Houser, J., Komarek, J., Kostlanova, N., Cioci, G., Varrot, A., Kerr, S. C., Lahmann, M., Balloy, V., Fahy, J. V., Chignard, M., Imberty, A. & Wimmerova, M. (2013). *PLoS One*, **8**, e83077.
Huh, C.-G., Aldrich, J., Mottahedeh, J., Kwon, H., Johnson, C. & Marsh, R. (1998). *J. Biol. Chem.* **273**, 6565–6574.
Iwanaga, S. (2002). *Curr. Opin. Immunol.* **14**, 87–95.
Iwanaga, S. & Lee, B.-L. (2005). *J. Biochem. Mol. Biol.* **38**, 128–150.
Jones, K. W., Gordon, B. M., Hanson, A. L., Kwiatek, W. M. & Pounds, J. G. (1988). *Ultramicroscopy*, **24**, 313–328.

Jones, S. & Thornton, J. M. (1995). *Prog. Biophys. Mol. Biol.* **63**, 31–65.
Jones, T. A., Zou, J.-Y., Cowan, S. W. & Kjeldgaard, M. (1991). *Acta Cryst.* A**47**, 110–119.
Kabsch, W. (1976). *Acta Cryst.* A**32**, 922–923.
Kawabata, S. & Iwanaga, S. (1999). *Dev. Comp. Immunol.* **23**, 391–400.
Kim, B.-S., Nam, B.-H., Kim, J.-W., Park, H.-J., Song, J.-H. & Park, C.-I. (2011). *Fish Shellfish Immunol.* **31**, 1201–1207.
Kostlánová, N., Mitchell, E. P., Lortat-Jacob, H., Oscarson, S., Lahmann, M., Gilboa-Garber, N., Chambat, G., Wimmerová, M. & Imberty, A. (2005). *J. Biol. Chem.* **280**, 27839–27849.
Krissinel, E. & Henrick, K. (2004). *Acta Cryst.* D**60**, 2256–2268.
Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A., Lopez, R., Thompson, J. D., Gibson, T. J. & Higgins, D. G. (2007). *Bioinformatics*, **23**, 2947–2948.
Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.
Leslie, A. G. W. & Powell, H. R. (2007). *Evolving Methods for Macromolecular Crystallography*, edited by R. J. Read & J. L. Sussman, pp. 41–51. Dordrecht: Springer.
Murata, K., Fisher, A. J. & Hedrick, J. L. (2007). *Acta Cryst.* F**63**, 396–398.
Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F. & Vagin, A. A. (2011). *Acta Cryst.* D**67**, 355–367.
Ponstingl, H., Henrick, K. & Thornton, J. M. (2000). *Proteins*, **41**, 47–57.
Rabinovich, G. A., van Kooyk, Y. & Cobb, B. A. (2012). *Ann. N. Y. Acad. Sci.* **1253**, 1–15.
Russell, S. & Lumsden, J. S. (2005). *Vet. Immunol. Immunopathol.* **108**, 111–120.
Saito, T., Kawabata, S., Hirata, M. & Iwanaga, S. (1995). *J. Biol. Chem.* **270**, 14493–14499.
Sanz-Gaitero, M., Keary, R., Garcia-Doval, C., Coffey, A. & van Raaij, M. J. (2014). *Virol. J.* **11**, 133.
Scharff, E. I., Koepke, J., Fritzsch, G., Lücke, C. & Rüterjans, H. (2001). *Structure*, **9**, 493–502.
Sharon, N. (2007). *J. Biol. Chem.* **282**, 2753–2764.
Sheldrick, G. M. (2008). *Acta Cryst.* A**64**, 112–122.
Springer, T. A. (1998). *J. Mol. Biol.* **283**, 837–862.
Taylor, M. E. & Drickamer, K. (2003). *Introduction to Glycobiology*. Oxford University Press.
Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994). *Nucleic Acids Res.* **22**, 4673–4680.
Vagin, A. & Teplyakov, A. (2010). *Acta Cryst.* D**66**, 22–25.
Vasta, G. R., Nita-Lazar, M., Giomarelli, B., Ahmed, H., Du, S., Cammarata, M., Parrinello, N., Bianchet, M. A. & Amzel, L. M. (2011). *Dev. Comp. Immunol.* **35**, 1388–1399.
Wimmerova, M., Mitchell, E., Sanchez, J. F., Gautier, C. & Imberty, A. (2003). *J. Biol. Chem.* **278**, 27059–27067.
Winn, M. D. *et al.* (2011). *Acta Cryst.* D**67**, 235–242.
Xie, J., Wen, J.-J., Chen, B. & Gui, J.-F. (2001). *Gene*, **271**, 109–116.
Zhu, L.-Y., Nie, L., Zhu, G., Xiang, L.-X. & Shao, J.-Z. (2013). *Dev. Comp. Immunol.* **39**, 39–62.