Fine-tuning of Proto-type Chicken Galectins: Crystal Structure of CG-2 and Structure-Activity Correlations

Federico M. Ruiz[1,#], Israel S. Fernández[1,#], Lara López-Merino[2,3,#], Laura Lagartera[3], Herbert Kaltner[4], Margarita Menéndez[2,3], Sabine André[4], Dolores Solís[2,3,*], Hans-Joachim Gabius[4], Antonio Romero[1*]

[1] Departamento de Biología Físico-Química, Centro de Investigaciones Biológicas - CSIC, Ramiro de Maeztu 9, E-28040 Madrid, Spain

[2] Departamento de Química-Física Biológica, Instituto de Química Física Rocasolano, CSIC, Serrano 119, E-28006 Madrid, Spain

[3] Centro de Investigación Biomédica en Red de Enfermedades Respiratorias (CIBERES), E-07110 Bunyola, Mallorca, Illes Balears, Spain

[4] Institut für Physiologische Chemie, Tierärztliche Fakultät, Ludwig-Maximilians-Universität, Veterinärstrasse 13, D-80539 München, Germany

Running Title: 3D Structure of CG-2

# These authors contributed equally to the work

* To whom correspondence should be addressed: Antonio Romero, Departamento de Biología Físico-Química, Centro de Investigaciones Biológicas - CSIC, Ramiro de Maeztu 9, E-28040 Madrid, Spain; Dolores Solís, Departamento de Química-Física Biológica, Instituto de Química Física Rocasolano - CSIC, Serrano 119, E-28006 Madrid, Spain.

E-mail addresses: romero@cib.csic.es; d.solis@iqfr.csic.es

ABSTRACT

The comparatively small number of members of the family of adhesion/growth-regulatory galectins in chicken predestines this system as an attractive model to study divergence of these lectins after gene duplications. Expression profiling of the three homodimeric (prototype) chicken galectins (CG-1A, CG-1B and CG-2) has raised evidence for distinct functionalities, explaining the interest in a detailed crystallographic analysis of CG-2. As revealed here, marked differences are found in the ligand-binding site and in the contact pattern within the homodimer interface, underlying a characteristic orientation of the two subunits. Notably, a distinctive trimer of dimers that is unique in all galectin crystal structures reported to date forms the core unit of the crystallographic assembly. Combination with spectroscopic and thermodynamic measurements, and comparisons with CG-1A and CG-1B, identifies differential changes in circular-dichroism spectra in the presence of lactose, reflecting the far-reaching impact of the ligand on hydrodynamic behavior, and inter-galectin differences in both the entropy and the enthalpy of binding. This structural information is a salient step to complete the analysis of the full set of galectins from this model organism.

Keywords: β-sandwich; homology; lectin; oligomerisation; phylogeny

Introduction

The central role of lectins in initiating the translation of sugar-encoded information into physiological responses explains the interest in defining their structures (Gabius, 2009; Gabius *et al.*, 2011). The members of the family of adhesion/growth-regulatory galectins share a distinct sequence signature, a β-sandwich fold with a central Trp residue in the binding site and affinity for lactose (Barondes *et al.*, 1994, Kasai & Hirabayashi, 1996, Gabius, 1997, Villalobo *et al.*, 2006, Klyosov *et al.*, 2008, Schwartz-Albiez, 2009). They are divided into three groups based on the mode of presentation of their carbohydrate recognition domain(s) (CRD). Monomeric and homodimeric proteins belong to the proto-type category, while the covalent association of the CRD with two other modules (collagen-like repeats and a peptide terminus with sites for Ser phosphorylation) leads to the chimera-type design, and the chain-like display of up to four different CRDs covalently connected by linker peptide(s) is characteristic of the tandem-repeat-type proteins (Kasai & Hirabayashi, 1996, Cooper, 2002, Tasumi & Vasta, 2007). Respective genes for these three groups are invariably found in vertebrates (Cooper, 2002, Houzelstein *et al.*, 2004). Gene duplications and the ensuing sequence divergence have given rise to different levels of complexity in the overall organization of the species-specific galectin profile. Obviously, an organism with a rather low extent of diversification is best suited for performing a comprehensive comparative analysis of the network comprising the three described classes. Based on this criterion, the chicken genome is an attractive model candidate. It harbors a total of only five genes for functional galectins (Cooper, 2002, Houzelstein *et al.*, 2004, Kaltner & Gabius, 2012).

We had previously reported the crystal structures of two proto-type proteins, i.e. the paralogue pair CG-1A and CG-1B (Varela *et al.*, 1999; López-Lucendo *et al.*, 2009). The third proto-type CG, i.e. CG-2, is assumed to have arisen from a duplication event prior to the emergence of CG-1A/B genes (the orthologues of human galectin-1; hGal-1) and, as its

indicates by its designation, CG-2 is suggested to be orthologous to human galectin-2 (hGal-2; for sequence comparisons, please see Fig. 1) (Hirabayashi *et al.*, 1987; Sakakura *et al.*, 1990; Houzelstein *et al.*, 2004; Kaltner *et al.*, 2008).

CG-2 had first been purified from adult chicken intestine and was subsequently detected to be strongly expressed in embryonic kidney (Beyer *et al.*, 1980; Beyer & Barondes, 1982; Stierstorfer *et al.*, 2000). Its immunohistochemical expression profile, with a presence in the epithelial lining of villi and intestinal glands as well as the collecting ducts in adult kidney, is mostly different from those of the other two proto-type CGs (Kaltner *et al.*, 2008; Kaltner & Gabius, 2012). Its presence in the digestive tract is shared by mammalian galectin-2, as revealed by analysis of human, mouse, pig and rat tissues (Oka *et al.*, 1999; Saal *et al.*, 2005; Lohr *et al.*, 2007; Thomsen *et al.*, 2009; Nio-Kobayashi *et al.*, 2009), and delineation of the profile of caspase involvement in the induction of apoptosis of activated T cells revealed functional divergence among human prototype galectins (Sturm *et al.*, 2004). In this respect, CG-1A was very active, while CG-2 showed selective B cell-binding capacity among avian immune cells (Schneller *et al.*, 1995; Kaltner *et al.*, 2008). On examination of the lectin sites of CG-2 and CG-1A in solution, chemical mapping detected initial evidence for quantitative differences in the capacity of CG-2 to bind β-lactoside derivatives relative to CG-1A (Solís *et al.*, 1996).

Here, we report the structure of CG-2. Owing to the availability of corresponding data for hGal-1 and hGal-2 (López-Lucendo *et al.*, 2004a; Lobsanov *et al.*, 1993), the proposed relationships between the avian and mammalian proteins could then likewise be tested. On the grounds of models of the binding of disaccharides to pentasaccharides to CG-1A. CG-1B, hGal-1 and hGal-2, along with glycan-array data for the human lectins (Siebert *et al.*, 2003; André *et al.*, 2005; Wu *et al.*, 2007; Stowell *et al.*, 2008), making assumptions on how affinity and specificity are altered during the course of sequence divergence appears to be possible. Flanking the crystallographic work, we carried out analytical experiments in solution with the

three proto-type CGs. Since circular dichroism (CD) spectroscopy has proven its value as a sensor for fine-structural alterations upon ligand binding in the case of human galectins (Nesmelova *et al.*, 2010; Solís *et al.*, 2010), we extended our comparison accordingly, adding determination of thermal stability in the absence and presence of lactose. The thermodynamics of lactose binding was also measured in order to answer the question on enthalpic/entropic contributions to binding in each case.

Materials and Methods

Protein expression and purification

The complete cDNA for CG-2 was cloned from total RNA of embryonic kidney (developmental day 15) and was inserted into the vector pQE60 (Qiagen, Hilden, Germany). *Escherichia coli* strain M15 (pREP4) was used for recombinant protein production of the product, which is identical to the physiological form, as described (Kaltner *et al.*, 2008). As given therein, the sequence of the full-length cDNA including the first exon (two amino acids: methionine and alanine) is identical to the respective GenBank entries XM_00123499.2 (mRNA) and XM_00123499.2 (protein), also completely in line with the genomic sequence (GenID: 425107). Following purification with affinity chromatography on lactosylated Sepharose 4B as a crucial step, controls for purity and activity assays were performed as described (Gabius *et al.*, 1984; Kaltner *et al.*, 2008; Amano *et al.*, 2012; André *et al.*, 2012). Production and quality controls of CG-1A, CG-1B and its C7S mutant were also carried out as described (Wu *et al.*, 2007; López-Lucendo *et al.*, 2009), and mass spectrometric fingerprint analyses of native and thermally denatured CG-1B and its mutant were performed after alkylation with 80 mM iodoacetamide in the presence of 9 M urea, during 1 h at 20 ºC and in the dark, without prior reduction ((López-Lucendo *et al.*, 2009).

Crystallization, data collection and processing

For crystallization experiments, CG-2-containing solution was dialyzed against a buffer containing 5 mM K,Na-phosphate, 150 mM NaCl, 10 mM lactose, and 4 mM ß-mercaptoethanol (pH 7.2), thereafter concentrated to 5 mg/ml using a Centricon-10 (Millipore) ultrafiltration unit. Initial crystallization trials were performed at room temperature using the sitting-drop vapour-diffusion method with commercial screening solutions including Crystal Screen I and II (Hampton Research), JBScreen Basic 1-4 (Jena Bioscience) and Wizard III (Emerald Biosciences). The sitting drops were prepared by mixing 0.2 µl of the protein solution and 0.2 µl of the reservoir solution in 96-well MRC plates (Swissci) using the Synquad (CARTESIAN) nanodispenser robot. Based on this initial screening, crystals were obtained using conditions 32 and 47 of Crystal Screen I and condition 24 of Wizard III, all of which contained 2.0 M ammonium sulphate. These conditions were further optimized and single crystals, reaching a maximum size of 0.1 x 0.1 x 0.1 mm, were obtained over the course of a week. The final condition was: 2 M ammonium sulphate, 0.1 M sodium acetate (pH 4.8) and 1% (v/v) β-mercaptoethanol. For data collection, crystals were rapidly transferred to a cryosolution containing reservoir solution supplemented with 20% (v/v) glycerol and flash-frozen in liquid nitrogen.

X-ray diffraction data to 1.75 Å resolution were collected on the ID14-2 beamline at the ESRF (Grenoble, France) using a single frozen crystal (100 K). A total of 180º was collected with a 1º oscillation angle and exposure time of 1s. The diffraction images were processed with MOSFLM (Leslie, 2006). Data scaling, merging and reduction were carried out with programs of the CCP4 suite (Collaborative Computational Project, 1994). The CG-2 crystals belong to space group $P2_12_12_1$, with unit-cell parameters of a = 65.85, b = 90.7, c = 151.66 Å. I/σ(I) drops below 2.0 at 1.7 Å resolution. However, to determine the highest -resolution cutoff for our data we used the correlation coefficient $CC_{1/2}$ (Karplus & Diederichs, 2012).

Values of 72.6 % for the highest resolution shell supported our high resolution limit at 1.75 Å. Data collection statistics are listed in Table 1.

Structure determination and refinement

An estimate of the solvent content of the CG-2 crystals (Matthews, 1968) suggested four (V_M = 3.79 Å^3/Da) to six (V_M = 2.52 Å^3/Da) molecules in the asymmetric unit. Self-rotation analysis showed peaks corresponding to 3-fold (κ=120º) (parallel to the y-axis), and 2-fold axes (κ=180º) (in the x-z plane) indicating the probable presence of six CG-2 molecules in the asymmetric unit.

The structure of CG-2 was determined using molecular replacement and the programs MOLREP (Vagin & Teplyakov, 2010) and PHASER (McCoy *et al.*, 2007) from the CCP4 suite (Collaborative Computational Project, 1994). The structure of CG-1A (40% sequence identity; PDB accession number 1QMJ; Varela *et al.*, 1999) was used as the search probe with data up to 3.5 Å. The first rotational and translational searches using MOLREP only identified four monomers of the hexamer in the asymmetric unit. The best solution had an Rfactor and a MOLREP score of 0.615 and 0.188, respectively. This partial resolution was then refined by REFMAC (Murshudov *et al.*, 2011). Subsequently, searches for the two remaining units were done using PHASER (McCoy *et al.*, 2007), keeping fixed the already detected tetramer and using the model of one of the CG-2 units as a probe in a new query. A solution comprising six monomers (TFZ = 16.5, LLG = 479) was obtained and then refined by several rounds of simulated annealing, followed by further cycles of B-factor improvement with non-crystallographic symmetry restraints. The final R_{work} and R_{free} factors were 18.9% and 22.7%, respectively. Manual model building/refinement cycles were done using COOT (Emsley *et al.*, 2010) and REFMAC (Murshudov *et al.*, 2011). Water molecules were gradually added during further conjugate-gradient refinement using COOT (Emsley *et al.*, 2010). The stereochemistry of the final model was validated using Molprobity (Davis *et al.*, 2007) with

no residues within prohibited regions of the Ramachandran plot. Protein-protein interactions were analysed using the PISA web server (Krissinel & Henrick, 2007). Refinement statistics are summarized in Table 1.

Circular dichroism

CD spectra were acquired in a J-810 spectropolarimeter, equipped with a Peltier temperature control system, using a band width of 1 nm and a response time of 4 s. Far UV spectra were recorded in 0.1 cm path-length quartz cells at a protein concentration of 0.2 mg/ml in 5 mM sodium phosphate buffer, pH 7.2, containing 0.2 M NaCl and 4 mM β-mercaptoethanol (PBS_β), while near UV spectra were registered at 1.0 mg/ml protein concentration in 1-cm path length cells. Buffer contribution was subtracted from protein raw data. Thermal denaturation experiments were carried out by increasing the temperature from 30 to 90 ºC at a scanning rate of 0.66 ºC min^{-1}. Spectra were recorded every 10 ºC and variations in ellipticity at a selected wavelength were then monitored every 0.2 ºC. Thermal denaturation profiles were described in terms of one or two sigmodial functions (Campanero-Rhodes *et al.*, 2006), depending on the number of transitions observed, using the equation:

$$\Theta(T) = \Theta_N + \sum_{i=1}^{n} \Delta\Theta_i \{\exp[-HD_i(T_{1/2i} - T)/R \cdot T_{1/2i} \cdot T]\} / \{1 + \exp[-HD_i(T_{1/2i} - T)/R \cdot T_{1/2i} \cdot T]\}$$

where $\Theta(T)$ is the ellipticity at absolute temperature T; Θ_N, the ellipticity of the native state; n, the number of transitions; $\Delta\Theta_i$, the variation in ellipticity associated with transition i; $T_{1/2i}$ and HD_i, the half-transition temperature and the parameter accounting for the cooperativity of the respective transition; and R, the gas constant.

Isothermal titration calorimetry

ITC measurements were carried out at 25 °C in PBS_β buffer using a VP-ITC microcalorimeter (MicroCal) as described (Martín-Santamaría *et al.*, 2011). Alternatively, 5 mM HEPES, pH 7.2, containing 0.2 M NaCl and 4 mM β-mercaptoethanol ($HEPES_\beta$), a

buffer with an ionization enthalpy quite different from that of phosphate, was used to evaluate a possible contribution of protein/buffer proton exchange resulting from variations in the pKa values of protein groups upon lactose binding. Before measurements, protein samples were exhaustively dialyzed against the corresponding buffer and lactose solutions for titrating galectin samples were prepared using the last dialysis buffer. The heat developed by lactose dilution was determined in separate runs and subtracted from the apparent heat of reaction. The thermodynamic parameters were calculated by analysing the binding isotherms using the MicroCal Origin software. The monomer concentration of the lectins was used as input in the fitting procedures.

Results and Discussion

Overall description of the structure

The crystal structure of CG-2 was determined at 1.75 Å resolution by molecular replacement, using previously published results for CG-1A as a search probe (PDB entry: 1QMJ; (Varela *et al.*, 1999). Statistics of data processing and refinement parameters of the structure are presented in Table 1. The final model comprises a total of six monomers with 130 residues each, two lactose molecules and 507 water molecules. The asymmetric unit in the crystal contains three CG-2 dimers, in which the two subunits of each dimer are related by a non-crystallographic 2-fold axis. Thus, the crystal organization can be viewed as a non-crystallographic trimer of dimers (Fig. 2a). Initial sedimentation equilibrium analyses at loading concentrations between 0.1 mg/ml and 1 mg/ml had shown CG-2 to be completely dimeric (Kaltner *et al.*, 2008), as gel filtration had also so far not provided evidence for oligomerization (Beyer *et al.*, 1980; Kaltner *et al.*, 2008). However, the concentration could markedly matter. To test this assumption, further sedimentation velocity analyses were carried out at a higher protein concentration (4 mg/ml). They revealed the presence of a predominant

peak (70% of the total protein) with a sedimentation coefficient of 2.3 S ($s_{20,w}$), corresponding to the CG-2 dimer, together with species with $s_{20,w}$ values of 3.4 S and 5.7 S, accounting for 14.9% and 14.6% of the total protein, respectively. Of note, a mass value of 86.3 kDa was estimated for the 5.7 S-species, fitting the mass of a hexamer. Thus, sedimentation velocity data are in favor of the formation of discrete oligomers by self-association of dimers at the protein concentration used for crystallization. Under similar conditions (5% (v/v) isopropanol in 2 M ammonium sulphate solution at pH 5.6), CG-1A is a dimer (Varela *et al.*, 1999), as is also the case for CG-1B (López-Lucendo *et al.*, 2009) and hGal-2 (Lobsanov *et al.*, 1993). However, several cases are known with evidence for oligomerization, the lectin 2 from the basidomycete *Coprinus cinereus* (inky cap mushroom) (Cooper *et al.*, 1997; López-Lucendo *et al.*, 2004b; Walser *et al.*, 2004), the N-domain of mouse galectin-4 (Krejčiřiková et al., 2011), and the galectin from the marine sponge *Cinachyrella* sp. (Freymann *et al.*, 2012), all of them crystallizing as tetramers. Furthermore, hGal-1, a dimer in crystals (López-Lucendo *et al.*, 2004a), builds a dimer of dimers of cylindrical shape in solution in dimethyl sulfoxide (He *et al.*, 2003). It is thus noteworthy that the mentioned galectins, given appropriate conditions, can self-associate to a quaternary structure beyond dimers. Still, the formation of a trimer of dimers is a unique feature of CG-2.

Monomer architecture

Looking at the CG-2 monomer, it adopts the typical galectin fold in which two anti-parallel β-sheets of six (S1-S6) and five (Fl-F5) β-strands, connected by several loops, formed a β-sandwich structure (Fig. 2b). Topological alignment of the monomers revealed no significant conformational differences, the r.m.s. deviations on the level of C_α atom positions being low (0.30 Å for 130 atoms). Only two of the six binding sites in the crystallographic asymmetric unit host a ligand (Fig. 2b). Crystal packing leads to blocking of the other four

binding sites, at which a lactose molecule cannot be accommodated because of steric clashes (Fig. S1).

With these data in hand, it was possible to address the issue whether CG-2 will present structural traits more akin to hGal-2 (interspecies comparison) than to CG-1A/B (intraspecies comparison) and hGal-1. Examining the superposed structures of these five galectins (Fig. 2c), there indeed is a difference in the S4-S5 loop, which connects the antiparallel β-strands S4-S5. This loop is four residues shorter in CG-2 and hGal-2 compared with hGal-1/CG-1A/B, in which the four additional amino acids protrude into the solvent (Ala-His-Gly-Asp in hGal-1; Cys-His-Gly-Asp in CG-1A and Ala-His-Gly-Asp in CG-1B). This special feature furnishes CG-2 and hGal-2 with an open cavity in this region, unlike the relatively narrow space in the vicinity of the carbohydrate-binding site of galectin-1 proteins, with potential consequences for ligand binding (please see below). The inspection of the structures also tracked down a minor local conformational change in the F5 β-strand of CG-2/hGal-2 vs hGal-1 and CG-1A/B (Fig. 2c), despite the high conservation of residues in this region (Fig. 1a).

Dimer interface

The canonical dimer interface is formed by the S1 and F1 β-strands from the N- and C-termini of each subunit (Fig. 3a). They face each other to set up a large antiparallel β-sheet. Sequence deviations shown in Fig. 1 let architecture changes become likely. Indeed, the hydrogen-bond pattern between S1 β-strands harbours a distinct number of interactions, with four pairs of backbone hydrogen bonds between Met4, Glu6 and Phe8, whereas only two residues are involved in the other listed proteins (Glu6-Lys8 in hGal-2; Val6-Ser8 in hGal-1; Val6-Thr8 in CG1-A/B). On the other side of the interface, the two antiparallel F1 strands establish four pairs of backbone hydrogen bonds, involving Thr126, Phe128 and Val130, as similarly observed for hGal-1/2 and CG1-A/B.

Apart from this polar network, hydrophobic interactions also contribute significantly to the stability of the dimer. The distribution of non-polar side chains from both the S1 (Met4, Phe5 and Met7) and F1 strands (Val125, Phe128 and Val130) shapes a hydrophobic surface that partakes in dimer assembly. The most striking characteristic of the CG-2 dimer interface is the way the side chains of Phe5 and Phe128 become key constituents of the hydrophobic pocket (Fig. 3b). hGal-2 shares only one of these two Phe residues in the sequence (Fig. 1a). Furthermore, CG-2 extends the hydrophobic core through recruiting additional methionine side chains, in positions 4 and 7. These two methionines are characteristic of CG-2, disclosing that sequence diversification of galectin-2 proteins after the ancestral gene duplication in these two organisms resulted in more than conservative differences.

In addition, the relative orientation of the monomers sets the CG-2 dimer apart from the canonical dimer displayed by the proto-type galectin-1 family (hGal-1, CG-1A/B) and also by hGal-2. Using superposition, the relatively large change in the arrangement of subunits in the CG-2 dimer is clearly seen in Fig. 4, with hGal-1 as control. A large rotation (approx. 25º) must be applied to bring the second subunit of the CG-2 dimer to its position in hGal-1. As a result, the extended β-sheet is less distorted in CG-2 than in CG-1A/B and the two human galectins. These disparities, given the similar overall fold of the monomers, are unlikely to be caused by crystal packing forces involving other regions of the protein (Fig. 2a). In detail, the three dimers present in the crystal asymmetric unit can readily be placed onto each other with r.m.s. deviations for all C_α atoms of 0.8 Å. Obviously, the fine structure of the interface and the shape distinguish CG-2 from CG-1A/B and hGal-1 and -2. These features can account for the distinct hydrodynamic behaviour of CG-2 reported previously, based on gel filtration, ultracentrifugation and measurement of the diffusion constant (Beyer *et al.*, 1980; Kaltner *et al.*, 2008; Göhler *et al.*, 2012).

Structure and stability in solution

In the first step, the far-UV CD spectrum of CG-2 was determined. The presence of a positive signal at 200 nm, together with a negative band centered at 219 nm (Fig. 5a), is indicative of β-sheet structure. Spectra obtained in the absence and presence of 0.1 M lactose indicated that there were no significant changes in secondary structure upon ligand binding in solution, in agreement with the crystallographic results. Comparison with the far-UV CD spectra of CG-1A and CG-1B (Fig. 5a) revealed smaller ellipticity signals for the two hGal-1 orthologues and a shift of the minimum towards 217 nm. Of further note, the spectra of CG-1A/B were significantly different from each other in the 220-234 nm region. These differences are compatible with presentation of distinctive structural features by each protein such as length and/or relative orientation of β-strands, sheets, and/or loops.

The near-UV CD spectrum of CG-2 was characterized by the presence of a broad positive signal, from 254 to 305 nm, with defined bands at 277 and 284 nm, in the tyrosine region, and a faint shoulder at 290 nm, attributable to tryptophan (Fig. 5b). The intensity of these signals increased perceptibly in the presence of lactose, pointing to changes in the environment of tyrosine and tryptophan residues upon ligand binding while the phenylalanine region (254-270 nm) was not affected. Moving on to the comparison of the three CGs, the overall shape of the near-UV CD spectrum of CG-1A was similar, although signal intensity was reduced. Moreover, the influence of lactose binding was rather small and restricted to the region of 280-295 nm, probably due to changes in the environment of the tryptophan residue located at the binding site. On the other hand, the spectrum of CG-1B presented a negative band centered at 288 nm, and the binding of lactose substantially affected the whole spectrum, including the region attributable to phenylalanine residues.

That the three proto-type CGs are sensitive to lactose binding to different degrees was also reflected by their responses to lactose presence when measuring the diffusion constant: its value increased by 5.6% for CG-1B, as that of hGal-1 (He *et al.*, 2003; Göhler *et al.*,

2010), whereas it decreased by 3.8% for CG-2 and remained unchanged for CG-1A (Göhler *et al.*, 2012), in line with lactose-induced changes observed in the CD spectra.

The stepwise heating of CG-2 resulted in the loss of tertiary and secondary structure, as evidenced by a progressive decrease of ellipticity signals both in the near- and far-UV CD spectra (Fig. 5c). CG-1B underwent similar changes, while for CG-1A changes were only apparent in the near-UV region. In contrast, the far-UV spectrum remained basically unaltered, with just a slight increase in negative ellipticity and significant noise below 210 nm at temperatures of 70 ºC and above (Fig. 5d). The denaturation process was monitored by measuring the decrease of ellipticity at an appropriate wavelength, that is, 217 nm for CG-2/CG-1B and 270 nm for CG-1A, as a function of the temperature (Fig. 6). Denaturation was irreversible and the experimental curves were analyzed phenomenologically using a sigmoidal function (please see methods), yielding the $T_{1/2}$ values compiled in Table 2. Both CG-2 and CG-1A denatured in a single cooperative process (Fig. 6a,b), as similarly observed for hGal-1 (Nesmelova *et al.,* 2010). Intriguingly, the denaturation profile of CG-1B showed two transitions (Fig. 6c) with a difference in $T_{1/2}$ of 11.6 ºC. Owing to previous experiments, unveiling the formation of intra- (Cys2-Cys7) and intermolecular (Cys7-Cys7) disulfide bonds in CG-1B associated to protein oxidation (López-Lucendo *et al.*, 2009), presence of these linkages in the thermally denatured protein was checked. Gel electrophoresis revealed the presence of a minor population of covalently-linked dimers, while mass spectrometric fingerprint analyses showed the formation of the intramolecular Cys2-Cys7 disulfide, the intensity of the respective ion being about 60% that of species without disulfide bridging. This result prompted the additional analysis of the C7S mutant of CG-1B. Its denaturation profile had a single transition (Fig. 6d), indicating that the two transitions observed for the wild-type protein were linked to the formation of disulfide-bonded species. Equally important in quantitative terms, the $T_{1/2}$ obtained for the mutant was in the range of that found for CG-2,

clearly below that of CG-1A. Thus, despite the detected disparities in the interface, thermal stabilities of CG-2 and CG-1B appear to be similar.

The impact of ligand binding on the thermal stability of the three CGs was also assessed. In all three cases, the $T_{1/2}$ increased with lactose binding, with the only exception of the second transition observed in the denaturation profile of CG-1B (Table 2). The increase in stability, however, differed markedly between the CGs, CG-2 being the frontrunner with a $T_{1/2}$ difference of 11.6 °C. The corresponding value was 2.5 °C for CG-1A (Table 2). The apparently protein-specific effect of lactose binding could derive from distinct ligand-induced structural changes but it may also reflect non-identical modes of binding and affinities. Thus, we next present details on the architecture of the binding site.

Carbohydrate-binding site

The concave surface of the β-sheet, comprising residues from the S4-S6 β-strands, constitutes the contact site for β-galactosides (Fig. 7a). Reflecting the common sequence signature of CGs (Kaltner *et al.*, 2008), a series of conserved amino acid residues is present in this region, i.e. His45, Asn47, Arg49, Asn58, Trp65, Glu68, and Arg70. As noted above, two of the six CG-2 monomers in the crystallographic asymmetric unit were occupied by lactose. The bound conformer corresponds to the low-energy (*syn*) state, extending the data basis for conformer selection by CG-1A/hGal-1 in solution (Siebert *et al.*, 1996; Asensio *et al.*, 1999; Alonso-Plaza *et al.*, 2001; Siebert *et al.*, 2003; García-Aparicio *et al.*, 2007; Solís *et al.*, 2009). When bound, the hydroxyl groups at 4'- and 6'-positions of galactose are the key contact sites for hydrogen bonding: the axial OH-4' of galactose establishes these interactions with the functional groups of the well-conserved residues His45, Asn47 and Arg49, while the hydroxyl group at the 6'-position is hydrogen bonded to Asn58 and Glu68. The glucose ring, through its OH-3 group, engages in hydrogen bonds with Arg49, Glu68 and Arg70, the latter two residues being also connected by a hydrogen bond. The invariant impact of lactose

presence on Trp signals in CD spectroscopy is structurally explained by the C-H/π-interaction with Trp65, as also seen in hGal-1/-2 (López-Lucendo *et al.*, 2004a; Lobsanov *et al.*, 1993).

Comparison of the architecture of the binding sites reveals that both CG-2 and hGal-2 share the contact of the glucose ring with Arg70 via a hydrogen bond. This addition to the common hydrogen-bond network in the recognition of lactose by galectins is favoured by the presence of the conserved acidic residue Glu68. It is appropriately located for a salt-bridge interaction with Arg70, thereby positioning its guanidinium group close to the equatorial OH-3 of glucose (3.1/2.8 Å N^E Arg70 – O-3 in CG-2/h-Gal2, respectively) (Fig. 7b). In hGal-1, as well as in CG-1A/B, the corresponding Arg residue is too far from the OH-3 for making contact with this group. Due to the presence of the short S4-S5 loop in CG-2 and hGal-2, the open space in the region above position 3 of glucose enables the accommodation of an axial hydroxyl group, although with a concomitant decrease in binding affinity due to the loss of contacts of the equatorial OH-3, as in fact observed for the binding of the 3-epi derivative of methyl β-lactoside to CG-2 (Solís *et al.*, 1996). In contrast, in hGal-1, as also in CG-1A/B, the longer S4-S5 loop places the conserved His52 at 4.8 Å over C3 of glucose. Whereas this type of placement for His52 maintains contact building with the galactose moiety, it may hinder the accommodation of axial substituents at position 3 of glucose.

Thermodynamic parameters of lactose binding

The thermodynamic parameters of the binding of lactose to CG-2 were assessed using isothermal titration calorimetry (ITC). The heat produced per mole of ligand injected as a function of the [lactose]/[CG] ratio is shown in Fig. 8. Experimental data, obtained at 25 ºC in PBS_β, could be fitted using a one-set-of-sites model, compatible with the presence of 1.08 ± 0.1 binding sites per CG-2 subunit with an association constant of 6000 ± 300 M^{-1} and ΔH and ΔS values of -11.1 ± 0.9 kcal·mol^{-1} and -20 ± 3 cal·mol^{-1}·K^{-1}, respectively (Table 3). Comparable parameters were obtained for CG-1A under similar conditions, this galectin

exhibiting a perceptibly higher affinity derived from a larger enthalpic contribution to the binding. By contrast, the enthalpy change observed for CG-1B was noticeably smaller. It resulted in an association constant 3- to 5-fold lower than those obtained for the other two proto-type CGs, despite the rather atypically lower ΔS value (close to zero).

In order to investigate the origin of these unusual parameters, other possible events that could contribute to the overall energetics of the process were explored. Titration of CG1-B with lactose in $HEPES_{\beta}$ yielded very similar thermodynamic parameters (Table 3), thus excluding a sizeable contribution of buffer ionization due to variations in the pKa values of protein induced by ligand binding (please see Materials and Methods). To evaluate a potential impact of disulfide bridge formation in CG-1B during titration, the thermodynamic parameters of lactose binding to the C7S mutant were measured. As shown in Table 3, ΔH and ΔS values calculated from the analysis approached those obtained for CG-2 and CG-1A, although the binding affinity remained in the same range, as the gain in favourable enthalpy was counteracted by a loss of entropy. The observed variations in enthalpy and entropy of binding between CG-1B and its C7S mutant point at disulfide bridging being a main factor to be reckoned with.

Overall, the results indicate that there is not a clear-cut correlation between the different architectures of the carbohydrate-binding sites of CG-2 vs CG-1A/B and the thermodynamic parameters of lactose binding. CG-1A and CG-1B already differ in the contributions to the Gibbs' free energy. Interestingly, CG-1B, the protein undergoing the largest ligand-induced changes in the near-UV CD spectrum, has the lowest binding affinity, whereas CG-1A, which displays rather small alterations in the spectrum, shows the highest affinity. These observations hint at the likely bearing of conformational rearrangements at sites other than the contact region for the ligand on the overall energetics of the binding process, what may vary with the nature of the ligand. Of interest, the conformational entropy

of the CRD of hGal-3 has been reported to be increased upon lactose binding without a major structural change, leading to the suggestion that affecting this parameter could conversely impact on the affinity of ligand binding (Diehl *et al.*, 2009). As this example attests, altering conformational entropy will not necessarily engender an effect on the shape, because the diffusion constant of hGal-3 is not affected by ligand loading (Gohler *et al.*, 2010).

Conclusions

The presented information of CG-2 completes the crystallographic analysis of homodimeric galectins in the chicken model system. It detects a shortened loop between strands S4 and S5, regional differences in the area for ligand contact, and unique features in the interface. Since counterreceptor selection critically depends on carbohydrate specificity and topological features (Murphy *et al.*, 2013), the potential of CG-2 dimers to self-associate at high local concentration could explain its significantly higher abundance at saturation on the surface of avian B cells compared to CG-1A, despite a rather similar pattern of reactive glycoproteins for the two lectins (Schneller *et al.*, 1995). This potential, which can be favoured by topologically suited counter-receptor presentation in microdomains (Stechly *et al.*, 2009; Kopitz *et al.*, 2010; Velasco *et al.*, 2013), may be physiologically relevant. It is the case for induction of pentamerization of the chimera-type Gal-3 by polyvalent ligands, in model systems with synthetic ligands and *in vitro* (Ahmad *et al.*, 2004; Kopitz *et al.*, 2010). Hereby, the monomeric lectin turns into a competitive inhibitor of binding of homodimeric (cross-linking) proteins. The case of the tumor suppressor $p16^{INK4a}$, which recruits hGal-1 to drive tumor cells into anoikis and at the same time downregulates Gal-3 availability, provides an example for functional divergence in a physiological context (André *et al.*, 2007; Sanchez-Ruderisch *et al.*, 2010).

Together with the tight regulation of expression of the three proto-type CGs with few cases of overlaps in the respective profiles (Kaltner *et al.*, 2008; Kaltner & Gabius, 2012),

these structural insights intimate a low degree of functional redundancy. An excellent example for this concept is the preferential expression of CG-1A in the zeugopod region of 5-day leg buds and its role in cell self-organizing dynamics to generate condensations and let cartilage mature in chicken limb skeletal morphogenesis (Bhat *et al.*, 2011). The interplay with the tandem-repeat-type CG-8 in this process cascade certainly provides incentive to characterize this family member to the same extent. Of medical interest, a genetic polymorphism in the coding region of the N-terminal CRD of human galectin-8 (here the F19Y variant) was associated with autoimmune diseases (Pal *et al.*, 2012). Recalling CG-2's peculiar hydrodynamic behavior, explaining how the individual pattern of structural dynamics and flexibility is encoded in the sequence of the CGs is an attractive challenge, as it is to complete the crystallographic analysis for all five CGs, along with their spectroscopic and thermodynamic characterization.

Data bank accession code

The coordinates of CG-2 have been deposited in the RCSB under accession code 2YMZ.

Acknowledgements

This work has been kindly supported by grants BFU2009-10052, BFU2011-24615, BFU2012-36825 and *CSD2009-00088* from the Spanish Ministry of Science and Innovation, the CIBER of Respiratory Diseases (CIBERES), an initiative from ISCIII, the Regional Government of Madrid (S2010/BMD-2353), and funding from the European Union's Seventh Framework Program FP7/2007-2013 under REA grant agreements no. 260600 ("GlycoHIT") and 317297 ("GLYCOPHARM"). F.M.R. and L.L.-M. were supported by JAE-DOC and JAE_PRE Grants, respectively, of CSIC (Spain). Finally, the authors cordially thank the staff of beamline ID14-2 at ESRF (Grenoble) for generous support.

References

Ahmad, N., Gabius, H.-J., André, S., Kaltner, H., Sabesan, S., Roy, R., Liu, B., Macaluso, F. & Brewer, C. F. (2004). *J. Biol. Chem.* **279**, 10841-10847.

Alonso-Plaza, J. M., Canales, M. A., Jiménez, M., Roldán, J. L., García-Herrero, A., Iturrino, L., Asensio, J. L., Cañada, F. J., Romero, A., Siebert, H. C., André, S., Solís, D., Gabius, H.-J. & Jiménez-Barbero, J. (2001). *Biochim. Biophys. Acta* **1568**, 225-236.

Amano, M., Eriksson, H., Manning, J. C., Detjen, K. M., André, S., Nishimura, S., Lehtio, J. & Gabius, H.-J. (2012). *Febs J.* **279**, 4062-4080.

André, S., Kaltner, H., Lensch, M., Russwurm, R., Siebert, H. C., Fallsehr, C., Tajkhorshid, E., Heck, A. J., von Knebel Doeberitz, M., Gabius, H.-J. & Kopitz, J. (2005). *Int. J. Cancer* **114**, 46-57.

André, S., Sanchez-Ruderisch, H., Nakagawa, H., Buchholz, M., Kopitz, J., Forberich, P., Kemmner, W., Bock, C., Deguchi, K., Detjen, K. M., Wiedenmann, B., von Knebel Doeberitz, M., Gress, T. M., Nishimura, S., Rosewicz, S. & Gabius, H.-J. (2007). *Febs J.* **274**, 3233-3256.

André, S., Jarikote, D. V., Yan, D., Vincenz, L., Wang, G. N., Kaltner, H., Murphy, P. V. & Gabius, H.-J. (2012). *Bioorg. Med. Chem. Lett.* **22**, 313-318.

Asensio, J. L., Espinosa, J. F., Dietrich, H., Cañada, F. J., Schmidt, R. R., Martín-Lomas, M., André, S., Gabius, H.-J. & Jiménez-Barbero, J. (1999). *J. Am. Chem. Soc.* **121**, 8995–9000.

Barondes, S. H., Cooper, D. N., Gitt, M. A. & Leffler, H. (1994). *J. Biol. Chem.* **269**, 20807-20810.

Beyer, E. C., Zweig, S. E. & Barondes, S. H. (1980). *J. Biol. Chem.* **255**, 4236-4239.

Beyer, E. C. & Barondes, S. H. (1982). *J. Cell. Biol.* **92**, 23-27.

Bhat, R., Lerea, K. M., Peng, H., Kaltner, H., Gabius, H.-J. & Newman, S. A. (2011). *BMC Dev. Biol.* **11**, 6.

Campanero-Rhodes, M. A., Menéndez, M., Saiz, J. L., Sanz, L., Calvete, J. J. & Solís, D. (2006). *Biochemistry* **45**, 8227-8235.

Collaborative Computational Project, N. (1994). *Acta Crystallogr. D Biol. Crystallogr.* **50**, 760-763.

Cooper, D. N., Boulianne, R. P., Charlton, S., Farrell, E. M., Sucher, A. & Lu, B. C. (1997). *J. Biol. Chem.* **272**, 1514-1521.

Cooper, D. N. (2002). *Biochim. Biophys. Acta* **1572**, 209-231.

Davis, I. W., Leaver-Fay, A., Chen, V. B., Block, J. N., Kapral, G. J., Wang, X., Murray, L. W., Arendall, W. B., 3rd, Snoeyink, J., Richardson, J. S. & Richardson, D. C. (2007). *Nucleic Acids Res.* **35**, W375-383.

DeLano, W. L. (2002). *The PyMOL Molecular Graphics System.*

Diehl, C., Genheden, S., Modig, K., Ryde, U. & Akke, M. (2009). *J. Biomol. NMR* **45**, 157-169.

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

Freymann, D. M., Nakamura, Y., Focia, P. J., Sakai, R. & Swanson, G. T. (2012). *Acta Crystallogr. D Biol. Crystallogr.* **68**, 1163-1174.

Gabius, H.-J., Engelhardt, R., Rehm, S. & Cramer, F. (1984). *J. Natl. Cancer Inst.* **73**, 1349-1357.

Gabius, H. J. (1997). *Eur. J. Biochem.* **243**, 543-576.

Gabius, H.-J. (Ed.) (2009). *The Sugar Code. Fundamentals of glycosciences*. Wiley-VCH, Weinheim.

Gabius, H.-J., André, S., Jiménez-Barbero, J., Romero, A. & Solís, D. (2011). *Trends Biochem. Sci.* **36**, 298-313.
García-Aparicio, V., Sollogoub, M., Bleriot, Y., Colliou, V., André, S., Asensio, J. L., Cañada, F. J., Gabius, H.-J., Sinay, P. & Jiménez-Barbero, J. (2007). *Carbohydr. Res.* **342**, 1918-1928.
Göhler, A., André, S., Kaltner, H., Sauer, M., Gabius, H.-J. & Doose, S. (2010). *Biophys J* **98**, 3044-3053.
Göhler, A., Buchner, C., André, S., Soren, D., Kaltner, H. & Gabius, H.-J. (2012). *Biochimie* **94**, 2649-2655.
He, L., André, S., Siebert, H. C., Helmholz, H., Niemeyer, B. & Gabius, H.-J. (2003). *Biophys. J.* **85**, 511-524.
Hirabayashi, J., Kawasaki, H., Suzuki, K. & Kasai, K. (1987). *J. Biochem.* **101**, 775-787.
Houzelstein, D., Goncalves, I. R., Fadden, A. J., Sidhu, S. S., Cooper, D. N., Drickamer, K., Leffler, H. & Poirier, F. (2004). *Mol. Biol. Evol.* **21**, 1177-1187.
Kaltner, H., Solís, D., Kopitz, J., Lensch, M., Lohr, M., Manning, J. C., Murnseer, M., Schnolzer, M., André, S., Saiz, J. L. & Gabius, H.-J. (2008). *Biochem. J.* **409**, 591-599.
Kaltner, H. & Gabius, H.-J. (2012). *Histol. Histopathol.* **27**, 397-416.
Karplus, P.A. & Diederichs, K. (2012). *Science* **336**, 1030–1033.
Kasai, K. & Hirabayashi, J. (1996). *J. Biochem.* **119**, 1-8.
Klyosov, A. A., Witczak, Z. J. & Platt, D. (2008). *Galectins*. Hoboken: J. Wiley & Sons.
Kopitz, J., Bergmann, M. & Gabius, H.-J. (2010). *IUBMB Life* **62**, 624-628.
Krejčiřiková, V., Pachl, P., Fabry, M., Maly, P., Rezacova, P. & Brynda, J. (2011). *Acta Crystallogr. D Biol. Crystallogr.* **67**, 204-211.
Krissinel, E. & Henrick, K. (2007). *J. Mol. Biol.* **372**, 774-797.
Leslie, A. G. (2006). *Acta Crystallogr. D Biol. Crystallogr.* **62**, 48-57.
Lobsanov, Y. D., Gitt, M. A., Leffler, H., Barondes, S. H. & Rini, J. M. (1993). *J. Biol. Chem.* **268**, 27034-27038.
Lohr, M., Lensch, M., André, S., Kaltner, H., Siebert, H. C., Smetana, K., Jr., Sinowatz, F. & Gabius, H.-J. (2007). *Folia Biol. (Praha)* **53**, 109-128.
López-Lucendo, M. F., Solís, D., André, S., Hirabayashi, J., Kasai, K., Kaltner, H., Gabius, H.-J. & Romero, A. (2004a). *J. Mol. Biol.* **343**, 957-970.
López-Lucendo, M. F., Giménez-Gallego, G., Cooper, D. N., Gabius, H.-J. & Romero, A. (2004b). *Acta Crystallogr. D Biol. Crystallogr.* **60**, 721-724.
López-Lucendo, M. F., Solís, D., Saiz, J. L., Kaltner, H., Russwurm, R., André, S., Gabius, H.- J. & Romero, A. (2009). *J. Mol. Biol.* **386**, 366-378.
Martín-Santamaría, S., André, S., Buzamet, E., Caraballo, R., Fernandez-Cureses, G., Morando, M., Ribeiro, J. P., Ramirez-Gualito, K., de Pascual-Teresa, B., Cañada, F. J., Menéndez, M., Ramstrom, O., Jiménez-Barbero, J., Solís, D. & Gabius, H.-J. (2011). *Org. Biomol. Chem.* **9**, 5445-5455.
Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491-497.
McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. (2007). *J. Appl. Crystallogr.* **40**, 658-674.
Murphy, P. V., André, S. & Gabius, H.-J. (2013) *Molecules* **18**, 4026-4053.
Murshudov, G. N., Skubak, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F. & Vagin, A. A. (2011). *Acta Crystallogr. D Biol. Crystallogr.* **67**, 355-367.
Nesmelova, I. V., Ermakova, E., Daragan, V. A., Pang, M., Menéndez, M., Lagartera, L., Solís, D., Baum, L. G. & Mayo, K. H. (2010). *J. Mol. Biol.* **397**, 1209-1230.

Nio-Kobayashi, J., Takahashi-Iwanaga, H. & Iwanaga, T. (2009). *J. Histochem. Cytochem.* **57**, 41-50.
Oka, T., Murakami, S., Arata, Y., Hirabayashi, J., Kasai, K., Wada, Y. & Futai, M. (1999). *Arch. Biochem. Biophys.* **361**, 195-201.
Pal, Z., Antal, P., Srivastava, S. K., Hullam, G., Semsei, A. F., Gal, J., Svebis, M., Soos, G., Szalai, C., André, S., Gordeeva, E., Nagy, G., Kaltner, H., Bovin, N. V., Molnar, M. J., Falus, A., Gabius, H.-J. & Buzas, E. I. (2012). *Biochim. Biophys. Acta* **1820**, 1512-1518.
Saal, I., Nagy, N., Lensch, M., Lohr, M., Manning, J. C., Decaestecker, C., André, S., Kiss, R., Salmon, I. & Gabius, H.-J. (2005). *Histol. Histopathol.* **20**, 1191-1208.
Sakakura, Y., Hirabayashi, J., Oda, Y., Ohyama, Y. & Kasai, K. (1990). *J. Biol. Chem.* **265**, 21573-21579.
Sanchez-Ruderisch, H., Fischer, C., Detjen, K. M., Welzel, M., Wimmel, A., Manning, J. C., André, S. & Gabius, H.-J. (2010). *Febs J.* **277**, 3552-3563.
Schneller, M., André, S., Cihak, J., Kaltner, H., Merkle, H., Rademaker, G. J., Haverkamp, J., Thomas-Oates, J. E., Losch, U. & Gabius, H.-J. (1995). *Cell Immunol.* **166**, 35-43.
Schwartz-Albiez, R. (2009). Inflammation and glycosciences, The Sugar Code. Fundamentals of glycosciences, edited by H.-J. Gabius, pp. 447-467. Weinheim: Wiley-VCH.
Siebert, H. C., Gilleron, M., Kaltner, H., von der Lieth, C. W., Kozar, T., Bovin, N., Korchagina, E. Y., Vliegenthart, J. F. & Gabius, H.-J. (1996). *Biochem. Biophys. Res. Commun.* **219**, 205-212.
Siebert, H. C., André, S., Lu, S. Y., Frank, M., Kaltner, H., van Kuik, J. A., Korchagina, E. Y., Bovin, N., Tajkhorshid, E., Kaptein, R., Vliegenthart, J. F., von der Lieth, C. W., Jiménez-Barbero, J., Kopitz, J. & Gabius, H.-J. (2003). *Biochemistry* **42**, 14762-14773.
Smetana, K. Jr., André, S., Kaltner, H., Kopitz, J. & Gabius, H.-J. (2013). *Exp. Opin. Ther. Targ.* **17**, 379-392.
Solís, D., Romero, A., Kaltner, H., Gabius, H.-J. & Díaz-Mauriño, T. (1996). *J. Biol. Chem.* **271**, 12744-12748.
Solís, D., Romero, A., Menéndez, M. & Jiménez-Barbero, J. (2009). *Protein-carbohydrate interactions: basic concepts and methods for analysis*. Weinheim: Wiley-VCH.
Solís, D., Mate, M. J., Lohr, M., Ribeiro, J. P., López-Merino, L., André, S., Buzamet, E., Cañada, F. J., Kaltner, H., Lensch, M., Ruiz, F. M., Haroske, G., Wollina, U., Kloor, M., Kopitz, J., Saiz, J. L., Menéndez, M., Jiménez-Barbero, J., Romero, A. & Gabius, H.-J. (2010). *Int. J. Biochem. Cell Biol.* **42**, 1019-1029.
Stechly, L., Morelle, W., Dessein, A. F., André, S., Grard, G., Trinel, D., Dejonghe, M. J., Leteurtre, E., Drobecq, H., Trugnan, G., Gabius, H.-J. & Huet, G. (2009). *Traffic* **10**, 438-450.
Stierstorfer, B., Kaltner, H., Neumuller, C., Sinowatz, F. & Gabius, H.-J. (2000). *Histochem. J.* **32**, 325-336.
Stowell, S. R., Arthur, C. M., Mehta, P., Slanina, K. A., Blixt, O., Leffler, H., Smith, D. F. & Cummings, R. D. (2008). *J. Biol. Chem.* **283**, 10109-10123.
Sturm, A., Lensch, M., André, S., Kaltner, H., Wiedenmann, B., Rosewicz, S., Dignass, A. U. & Gabius, H.-J. (2004). *J. Immunol.* **173**, 3825-3837.
Tasumi, S. & Vasta, G. R. (2007). *J. Immunol.* **179**, 3086-3098.
Thomsen, M. K., Hansen, G. H. & Danielsen, E. M. (2009). *Mol. Membr. Biol.* **26**, 347-355.
Vagin, A. & Teplyakov, A. (2010). *Acta Crystallogr. D Biol. Crystallogr.* **66**, 22-25.
Varela, P. F., Solís, D., Díaz-Mauriño, T., Kaltner, H., Gabius, H.-J. & Romero, A. (1999). *J. Mol. Biol.* **294**, 537-549.
Velasco, S., Díez-Revuelta, N., Hernández-Iglesias, T., Kaltner, H., André, S., Gabius, H.-J. & Abad-Rodriquez, J. (2013) *J. Neurochem.* **125**, 49-62.

Villalobo, A., Nogales-González, A. & Gabius, H.-J. (2006). *Trends Glycosci. Glycotechnol.* **18**, 1-37.
Walser, P. J., Haebel, P. W., Kunzler, M., Sargent, D., Kues, U., Aebi, M. & Ban, N. (2004). *Structure* **12**, 689-702.
Wu, A. M., Singh, T., Liu, J. H., Krzeminski, M., Russwurm, R., Siebert, H. C., Bonvin, A. M., André, S. & Gabius, H.-J. (2007). *Glycobiology* **17**, 165-184.

Figure Legends

Figure 1. Sequence alignment of CG-2 with those of (a) human galectins-1 and -2 (hGal-1/-2) and (b) with those of the other two proto-type CGs (CG-1A, CG-1B). Strictly conserved residues (red boxes) and homologous residues (> 70% conservation, white boxes with capital letters) are shown.

Figure 2. Overall structure of CG-2. (a) Two different views of the asymmetric unit containing a total of six CG-2 units arranged into three dimers. The two subunits of each dimer (coloured in yellow, cyan and green) are related by a non-crystallographic 2-fold axis. (b) The CG-2 monomer adopts the typical β-sandwich fold formed by two anti-parallel β-sheets of six (S1-S6) and five (Fl-F5) β-strands. The lactose molecule in the binding site of the right subunit is shown in the stick mode. (c) Superimposition of the C_α atoms of members of the proto-type group, i.e. CG-2 (yellow), CG1-A (PDB code: 1QMJ) (Varela *et al.*, 1999) (magenta), CG1-B (PDB code: 3DUI) (López-Lucendo *et al.*, 2009) (cyan), hGal-1 (PDB code: 1GZW) (López-Lucendo *et al.*, 2004a) (black) and hGal-2 (PDB code: 1HLC) (Lobsanov *et al.*, 1993) (green) (for sequences, please see Fig. 1). The region displaying the largest deviation (S4-S5 loop) is remarked. Figures were prepared using the program PYMOL (DeLano, 2002).

Figure 3. Overview of the canonical interface of the CG-2 dimer. (a) S1 and F1 β-strands from the N- and C-termini are in the contact region, facilitating both polar and hydrophobic interactions. (b) The special distribution of non-polar side chains in both S1 and F1 strands establishes a hydrophobic surface. Most notable is the involvement of the side chains of Phe5 and Phe128 within the core of the hydrophobic pocket (yellow). In addition, CG-2 extends the hydrophobic core through additional methionine side chains, in positions 4 and 7.

Figure 4. Comparison of the dimer arrangement (left-side) in CG-2 (grey) and hGal-1 (yellow). The superposition is based on one subunit of CG-2 and yielded a r.m.s.d. deviation of 0.4 Å for 130 residues. On the right-side a separated view of the superposition, from the bottom (lower panel) and from the top (upper panel). As can be seen on this last panel, a large rotation must be applied to bring the second subunit of the CG-2 dimer to the equivalent position in hGal-1.

Figure 5. CD spectra of the three proto-type CGs. The far- (a) and near-UV (b) CD spectra of CG-2 (□,■), CG-1A (○,●) and CG-1B (Δ,▲) were obtained at 20 ºC for 0.2 and 1 mg/ml solutions, respectively, in PBS_{β}, in the absence (open symbols) and presence (solid symbols) of 0.1 M lactose. Far-UV spectra of CG-2 (c) and CG-1A (d) recorded at 30 ºC (solid line), 50 ºC (dash-dot line), 60 ºC (CG-2)/70 ºC (CG-1A) (dash line) and 90 ºC (dot line).

Figure 6. Thermal denaturation profiles of CG-2 (a), CG-1A (b), CG-1B (c) and the C7S mutant of CG-1B (d). The variation of ellipticity with temperature was measured either at 217 nm (a, c, d) or at 270 nm (b) in the absence (open symbols) and in the presence of 0.1 M lactose (solid symbols) for 0.2 (a, c, d) and 1 mg/ml (b) protein solutions in PBS_{β}. The continuous lines correspond to the fit of sigmoidal functions to experimental data.

Figure 7. Carbohydrate-binding site of CG-2. (a) The lactose moiety and the amino acids of the hydrogen-bond network with the sugar are shown in the stick mode. Observed electron density map of lactose in a Fo-Fc omit map contoured at 2.0 σ. (b) Superposition of this site of CG-2 in ligand-loaded (grey) and ligand-free (cyan) forms.

Figure 8. Representative calorimetric titrations of CG-2, CG-1A, CG-1B and the C7S mutant

of CG-1B with lactose. Symbols represent the heat released per mole of lactose injected as a function of the molar ratio between lactose and CG, the solid lines correspond to the best fit of the experimental data based on a one-set-of-sites model. Titrations were carried out at 25 ºC by two successive series of injections of 11-13 mM lactose aliquots into the sample cell containing solutions with monomer concentrations of 192 μM CG-2 (□), 150 μM CG-1A (○), 196 μM CG-1B (Δ) or 80 μM of the C7S mutant of CG-1B (▽) in PBS_{β}.

Table 1. Data collection and refinement statistics

Data collection	
Space group	$P2_12_12_1$
Cell dimensions (Å)	a = 65.85, b = 90.7, c = 151.66
Resolution (Å) (outermost shell)	43.6 – 1.75 (1.84 – 1.75)*
R_{merge}	0.065 (0.479)
$I / \sigma I$	5.7 (1.6)
Completeness (%)	98.0 (88.6)
Multiplicity	6.5 (3.8)
Reflections measured	587,745 (44,993)
Unique Reflections	90,417 (11,740)
B-Factor (Wilson) ($Å^2$)	25.05
Mosaicity (º)	0.2
Refinement	
Resolution (Å)	43.6 – 1.75
No. reflections	85,820 (4,511)
R_{work} / $R_{free\ (\%)}$	18.9 / 22.7
No. atoms	
Protein	6354
Ligand/ion	46
Water	507
B-factors	
Protein	30.05
Ligand/ion	53.2
Water	40.5
R.m.s. deviations	
Bond lengths (Å)	0.019
Bond angles (°)	2.171

*Values in parentheses are for outer-resolution shell.

$R_{merge}=\Sigma|I-\langle I\rangle|/\Sigma I$, where I is the measurement of the intensity for each reflection and $\langle I\rangle$ is the mean intensity of that reflection.

$R_{work} = \sum|F_o - F_c|/\sum F_o$, where F_o and F_c are the observed and calculated structure factor amplitudes, respectively. R_{free} is calculated as for R_{work} but for a subset of randomly chosen reflections (5%), which were not used for structure refinement.

Table 2. Transition temperatures of thermal denaturation of proto-type chicken galectins monitored by circular dichroism

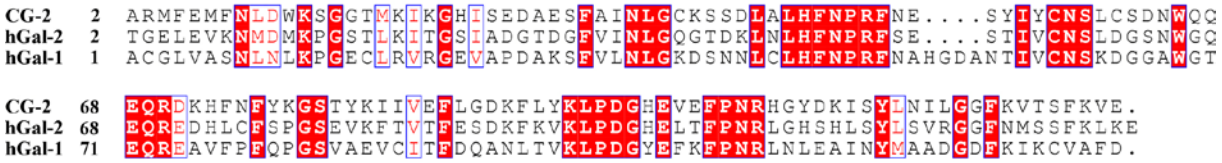
Galectin	$T_{1/2}$ (ºC)	
	-	+ Lac 0.1 M
CG-2	57.9 ± 0.15	69.5 ± 0.1
CG-1A	71.5 ± 0.25	74.0 ± 0.2
CG-1B		
wild-type	65.0 ± 0.2 76.6 ± 0.2	72.4 ± 0.1
C7S mutant	60.5 ± 0.1	67.4 ± 0.1

Table 3. Thermodynamic parameters for lactose binding to the proto-type chicken galectins obtained by ITC

Protein	Buffer	K_a (M^{-1})	ΔG (kcal·mol^{-1})	ΔH (kcal·mol^{-1})	ΔS (cal·mol^{-1}·K^{-1})
CG-2	PBS_β	6000 ± 300	-5.14 ± 0.03	-11.1 ± 0.9	-20 ± 3
CG-1A	PBS_β	9000 ± 700	-5.38 ± 0.06	-12.1± 0.7	-23 ± 2
CG-1B					
wild type	PBS_β	1790 ± 200	-4.42 ± 0.07	-4.1 ± 0.3	1 ± 0.7
	$HEPES_\beta$	2230 ± 20	-4.55 ± 0.01	-4.46 ± 0.02	0.4 ± 0.09
C7S mutant	PBS_β	1600 ± 200	-4.36 ± 0.07	-9.0 ± 0.4	-16 ± 1

Figure 1.

a


CG-2 2 ARMFEMFNLDWKSGGTMKIKGHISEDAESFAINLGCKSSDLALHFNPRFNE....SYIYCNSLCSDNWQQ
hGal-2 2 TGELEVKNMDMKPGSTLKITGSIADGTDGFVINLGQGTDKLNLHFNPRFSE....STIVCNSLDGSNWGQ
hGal-1 1 ACGLVASNLNLKPGECLRVRGEVAPDAKSFVLNLGKDSNNLCLHFNPRFNAHGDANTIVCNSKDGGAWGT
CG-2 68 EQRDKHFNFYKGSTYKIIVEFLGDKFLYKLPDGHEVEFPNRHGYDKISYLNILGGFKVTSFKVE.
hGal-2 68 EQREDHLCFSPGSEVKFTVTFESDKFKVKLPDGHELTFPNRLGHSHLSYLSVRGGFNMSSFKLKE
hGal-1 71 EQREAVFPFQPGSVAEVCITFDQANLTVKLPDGYEFKFPNRLNLEAINYMAADGDFKIKCVAFD.

b

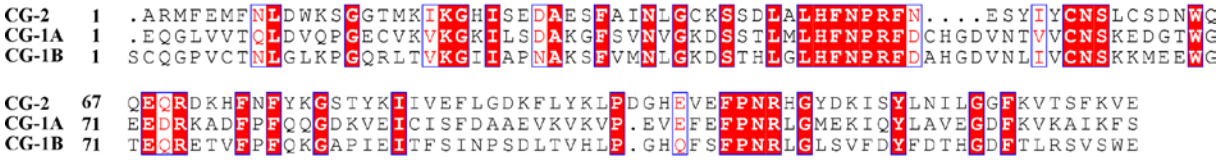

CG-2 1 .ARMFEMFNLDWKSGGTMKIKGHISEDAESFAINLGCKSSDLALHFNPRFN....ESYIYCNSLCSDNWQ
CG-1A 1 .EQGLVVTQLDVQPGECVKVKGKILSDAKGFSVNVGKDSSTLMLHFNPRFDCHGDVNTVVCNSKEDGTWG
CG-1B 1 SCQGPVCTNLGLKPGQRLTVKGIIAPNAKSFVMNLGKDSTHLGLHFNPRFDAHGDVNLIVCNSKKMEEWG
CG-2 67 QEQRDKHFNFYKGSTYKIIVEFLGDKFLYKLPDGHEVEFPNRHGYDKISYLNILGGFKVTSFKVE
CG-1A 71 EEDRKADFPFQQGDKVEICISFDAAEVKVKVP.EVEFEFPNRLGMEKIQYLAVEGDFKVKAIKFS
CG-1B 71 TEQRETVFPFQKGAPIEITFSINPSDLTVHLP.GHQFSFPNRLGLSVFDYFDTHGDFTLRSVSWE

Figure 2.

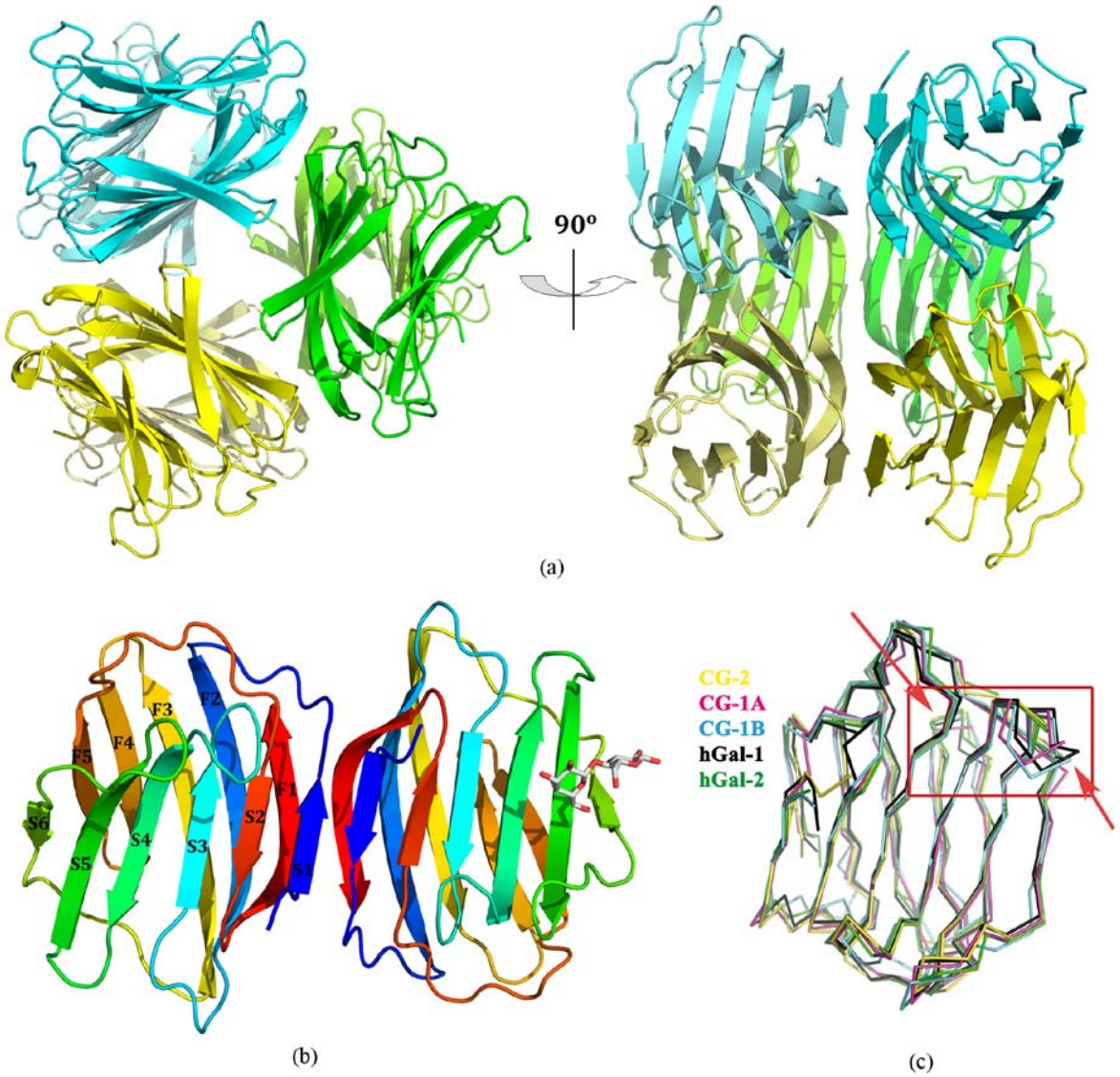


Figure 3

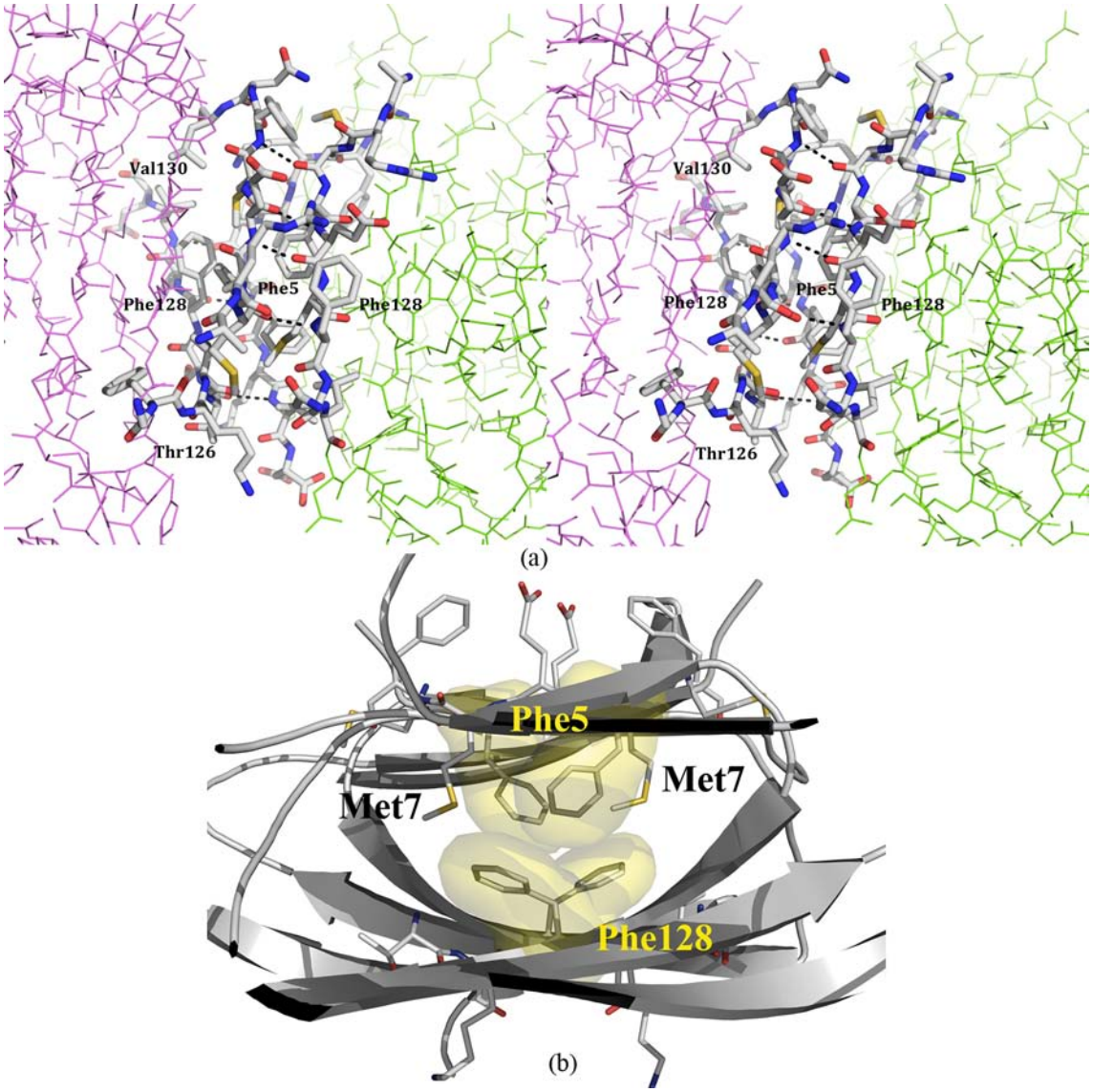

Val130
Phe5
Phe128
Phe128
Thr126
Val130
Phe5
Phe128
Phe128
Thr126
(a)
Phe5
Met7
Met7
Phe128
(b)

Figure 4

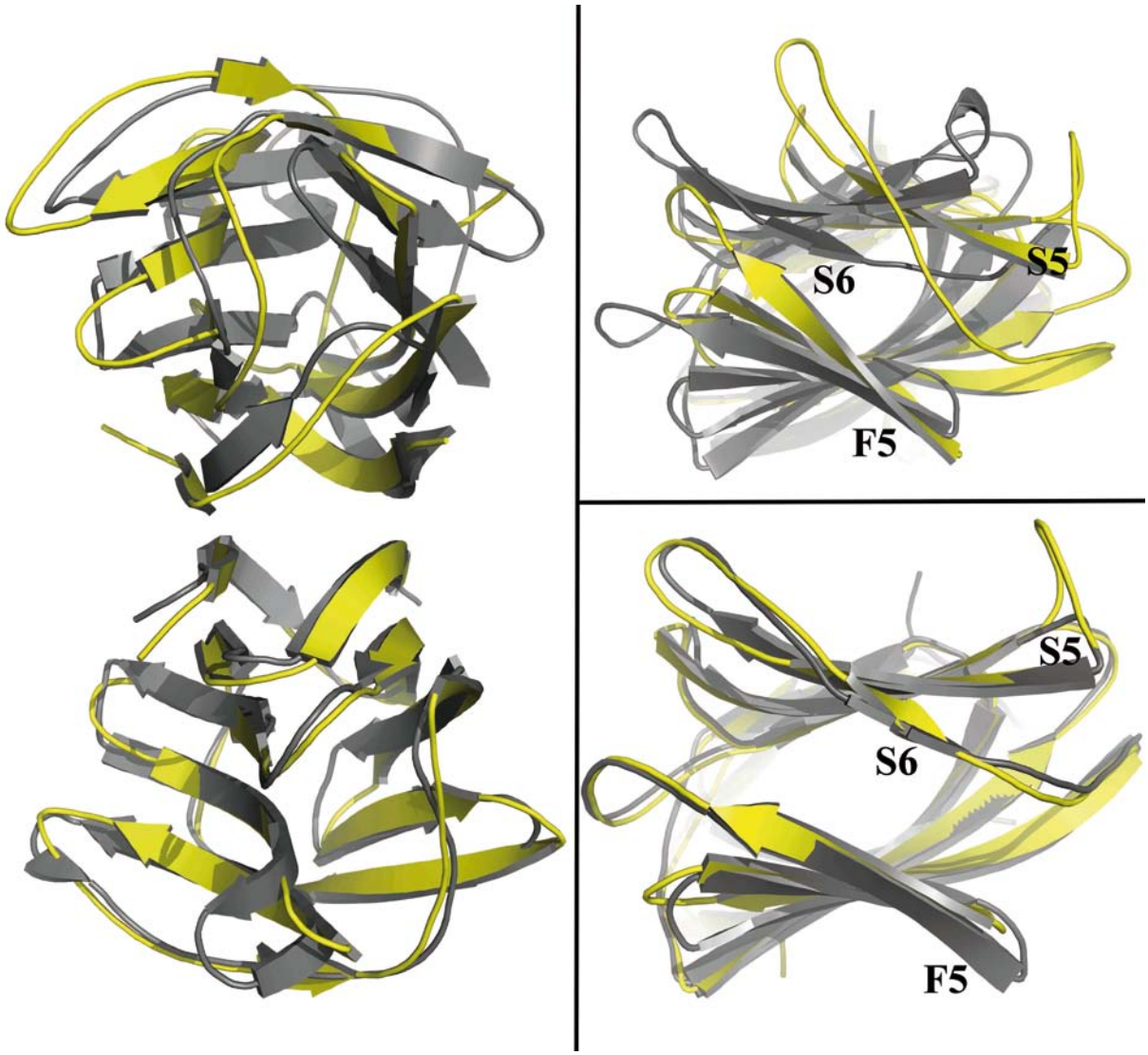


Figure 5

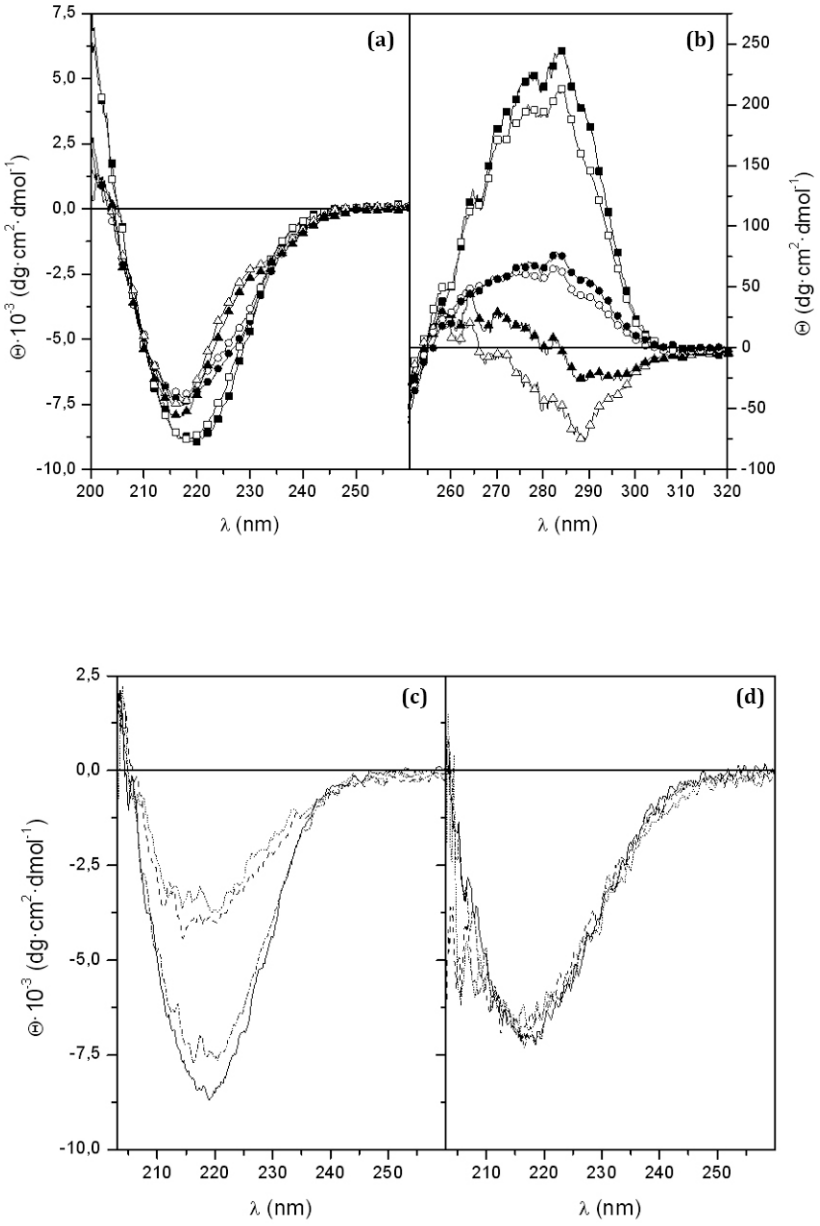


Figure 6

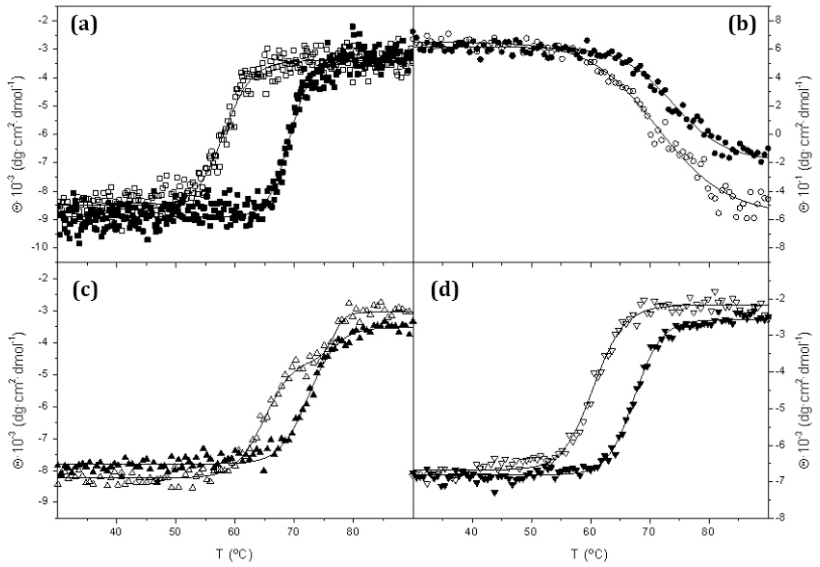
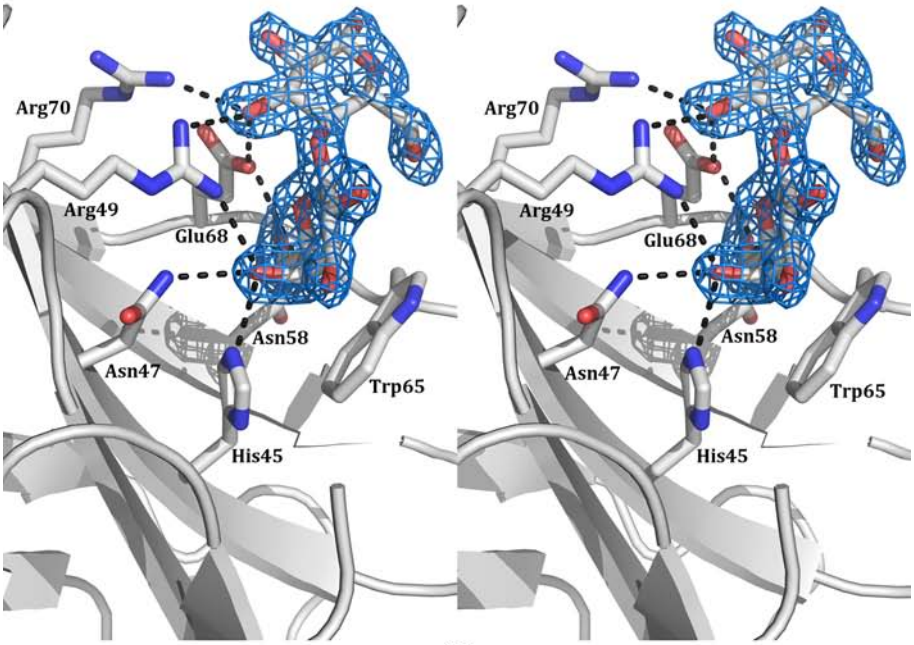
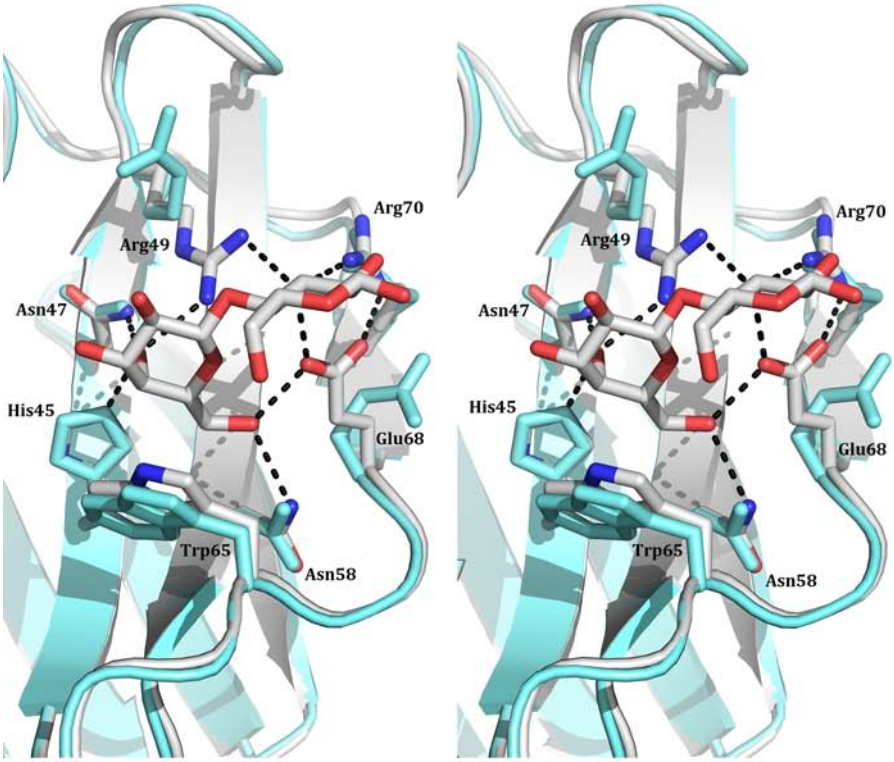


Figure 7


Arg70
Arg49
Glu68
Asn58
Asn47
Trp65
His45
Arg70
Arg49
Glu68
Asn58
Asn47
Trp65
His45

(a)


Arg70
Arg49
Asn47
His45
Glu68
Trp65
Asn58
Arg70
Arg49
Asn47
His45
Glu68
Trp65
Asn58

(b)

Figure 8