Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Biochimica et Biophysica Acta 1860 (2016) 2285–2297

Contents lists available at ScienceDirect

Biochimica et Biophysica Acta

journal homepage: www.elsevier.com/locate/bbagen

Galectin-related protein: An integral member of the network of chicken galectins 1. From strong sequence conservation of the gene confined to vertebrates to biochemical characteristics of the chicken protein and its crystal structure

Gabriel García Caballero [a,1], Andrea Flores-Ibarra [b,1], Malwina Michalak [c,1], Nailya Khasbiullina [d], Nicolai V. Bovin [d], Sabine André [a], Joachim C. Manning [a], Sabine Vértesy [a], Federico M. Ruiz [b], Herbert Kaltner [a], Jürgen Kopitz [c], Antonio Romero [b], Hans-Joachim Gabius [a,*]

[a] Institute of Physiological Chemistry, Faculty of Veterinary Medicine, Ludwig-Maximilians-University Munich, Veterinärstr. 13, 80539 Munich, Germany
[b] Chemical and Physical Biology, Centro de Investigaciones Biológicas, CSIC, Ramiro de Maeztu 9, 28040 Madrid, Spain
[c] Department of Applied Tumor Biology, Institute of Pathology, Medical School of the Ruprecht-Karls-University, Im Neuenheimer Feld 224, 69120 Heidelberg, Germany
[d] Shemyakin and Ovchinnikov Institute of Bioorganic Chemistry, Russian Academy of Sciences, ul. Miklukho-Maklaya 16/10, Moscow, Russia

ARTICLE INFO

Article history:
Received 16 March 2016
Received in revised form 11 May 2016
Accepted 2 June 2016
Available online 3 June 2016

Keywords:
Adhesion
Crystallography
Lectin
Phylogenesis
Proliferation

ABSTRACT

Background: Endogenous lectins are multifunctional effectors in cell physiology. Adding the sixth member of the galectin family in chicken, a model organism for systematic profiling of these adhesion/growth-regulatory proteins, is a step toward comprehensive network monitoring.
Methods: Database mining and computational data processing are applied for gene detection, chromosomal location and sequence alignments. Cloning, recombinant production and fusion-protein technology gain access to the protein, mass spectrometry and gel electrophoresis/filtration provide analytical data. Haemagglutination, glycan microarray and cell assays assess binding capacity, and crystallography of a shortened variant (also analyzed by ultracentrifugation and small angle X-ray scattering) determines its structure.
Results: The gene for the galectin-related protein (GRP) is present exclusively in vertebrates with high-level sequence conservation and similar chromosomal positioning. The chicken protein is monomeric and has lost the canonical galectin property of binding lactose. The crystal structure of the variant without the 36-amino-acid extension at the start provides explanations for this lack of binding.
Conclusions: Chicken GRP is special within this family of six proteins by being unable to bind lactose. The documented high degree of sequence conservation among vertebrate orthologues confers the status of a model for delineating an assumedly shared functionality to this GRP.
General significance: Biochemical characterization of a product of a gene under strong positive selection is a prerequisite for functional characterization. It is also essential for network monitoring by adding a new member to this lectin family.

1. Introduction

The concept of the sugar code (for recent reviews, please see [1]) rests on the existence of routes to turn glycan-encoded information into effects on the cellular level. In this respect, a broad physiological significance has been delineated for receptors of distinct sugar determinants (lectins), what prompted their thorough structural characterization [2–7]. Proceeding from the identification of a protein fold with capacity to accommodate glycans and a common sequence signature, the systematic search for homologous proteins in a species and in phylogenesis is the next step on the way toward network analysis of these effectors. The results of respective database mining then set the stage to characterize expression and the profile of localization as well as functional cooperation of all members of the corresponding family of lectins, which have arisen from an ancestral gene by divergence through duplications/losses and sequence deviations. The choice of lectin class and species for a comprehensive case study of general relevance is favored, if the proteins found in a species represent the full diversity of known types of structural organization at a relatively small number of individual proteins. Guided by these

* Corresponding author.
E-mail addresses: gabius@tiph.vetmed.uni-muenchen.de, gabius@lectins.de (H.-J. Gabius).
[1] These authors contributed equally.

http://dx.doi.org/10.1016/j.bbagen.2016.06.001
0304-4165/

criteria, complete fingerprinting of biochemical properties and tissue positivity at the lowest degree of study complexity will be possible. Turning to the adhesion/growth-regulatory galectins, a family of β-sandwich proteins with carbohydrate specificity to β-galactosides and derivatives thereof [8–10], this general prerequisite is best fulfilled in chicken with a total of only five canonical proteins [11]. However, as pointed out [10], the occurrence of an expressed sequence tag (EST) with similarity to a mammalian galectin-related protein (GRP) found in chicken bursal lymphocytes indicates that a sixth member of this family is present in this organism (C-GRP; AJ453496) predestined for model study of the galectin network. Obviously, comparative biochemical characterization from the gene to the mature product as well as expression profiling and tissue localization of this protein are thus called for.

The detection of GRP has its origin in cataloguing of gene expression of human CD34-positive haematopoietic stem/progenitor cells (HSPCs) that led, among 300 cDNA clones, to an mRNA termed HSPC159 [12]. Systematic alignments of its predicted amino acid sequence disclosed similarity to galectins. It encompasses 51 positions of the set of 64 amino acids most likely shared among these proteins [10]. Presence of the gene was not confined to man. GRP sequences had been found in mammals (man, mouse), chicken, frog and fish (puffer and zebrafish), and, intriguingly, initial comparison revealed evidence for an exceptionally high degree of similarity that implies a "very strong positive selection, as generally seen for genes encoding proteins with multiple aspects involved in critical interactions" [10]. Given i) this feature emerging from inter-species considerations that imply special functionality and ii) the obvious requirement to bring characterization of this chicken protein (i.e. C-GRP) to the same level as has been done for the five canonical chicken galectins (CGs), i.e. the three proto-type (homodimeric) CG-1A, CG-1B and CG-2, the chimera-type CG-3 and the tandem-repeat-type CG-8 [13–17], we here take a two-step approach. First, we present an overview on occurrence of the GRP gene and its organization in phylogenesis, examine C-GRP's biochemical properties and describe the crystallographic structure of a shortened variant. In the accompanying second part, details on C-GRP expression and its localization in the positive tissue as well as detection of reactive sites for C-GRP are given, set in relation to corresponding results with the five canonical CGs.

2. Materials and methods

2.1. Processing sequence information and constructing phylogenetic trees

Gene sequences were downloaded from the Ensembl Genome Browser (www.ensembl.org/index.html; Ensembl release 83, December 2015) and the NCBI Genbank (www.ncbi.nlm.nih.gov/genbank/index.html) with its divisions CoreNucleotide (www.ncbi.nlm.nih.gov/nuccore), EST database (www.ncbi.nlm.nih.gov/nucest) and Genome Survey Sequence database (www.ncbi.nlm.nih.gov/nucgss). Information on copy-number variation was obtained by processing whole-shotgun sequencing data of each species, provided by the University of California Santa Cruz genome browser (www.genome.ucsc.edu) and by NCBI Genome (www.ncbi.nlm.nih.gov/genome) as contigs, unplaced scaffolds, chromosomal genomic scaffolds and assemblies, then analyzed for presence of distinct exon sequences, thereafter routinely for full-length coding sequence, as carried out for canonical galectins of mammals recently [18]. In addition, the Basic Local Alignment Search Tool (BLAST) (http://blast.ncbi.nlm.nih.gov/Blast.cgi) search algorithms were applied to cover the full range of sequences, which satisfy stringent criteria of homology. Information on entries for orthologues of the GRP gene in species of different branches of the phylogenetic tree was displayed applying the NCBI Taxonomy Browser (www.ncbi.nlm.nih.gov/Taxonomy/CommonTree/wwwcmt.cgi) and the visualization software TreeView (http://taxonomy.zoology.gla.ac.uk/rod/treeview.html). Genomic sequences were examined manually in each case for annotation errors, presence of non-sequenced stretches, exon/intron boundaries and completeness of the coding sequence including the presence of start/stop codons, applying the sequence text view tool implemented in NCBI Gene (www.ncbi.nlm.nih.gov/gene; [19]). Sequences were edited using the EditSeq software version 12.1.0 (DNAstar Inc., Madison, WI, USA). The principle to stringently apply homology criteria was rigorously followed in each case. Amino acid sequences of the predicted gene products were deduced unless available in the NCBI package retrieved by BLASTP/Position-Specific Iterative BLAST search algorithms from the NCBI Protein (www.ncbi.nlm.nih.gov/protein) and the UniProt Knowledgebase (UniProtKB, Expasy Proteomics Server; www.expasy.ch).

Multiple alignments of amino acid sequences were performed using the Clustal Omega software (www.ebi.ac.uk/Tools/msa/clustalo/; [20]) and edited in Jalview (v. 2.7; [21]). Aligned sequences both for intra-family (C-GRP vs CGs) and inter-species (C-GRP vs GRP of other species) comparisons were processed manually to spot positions of highly conserved amino acids relevant for binding the canonical sugar ligand lactose (Lac) (sequence signature of galectins).

Analysis of evolutionary relationships and construction of phylogenetic trees were done using the Maximum Likelihood method implemented in the MEGA6 software package [22]. The tree with the highest log likelihood is presented. The test of phylogeny was performed using bootstrap analysis (with 1000 replicates); the percentage of tree(s), in which the associated taxa clustered together, is shown next to branches. Initial tree(s) for the heuristic search were obtained by applying the Neighbor-Joining method to a matrix of pairwise distances estimated using a Jones–Taylor–Thornton model. For selected organisms, the chromosomal environment of the GRP gene was comparatively analyzed using the Ensembl Genome Browser (www.ensembl.org/index.html) and the NCBI Map Viewer (www.ncbi.nlm.nih.gov/mapview/).

2.2. Cloning, expression and purification

Total RNA from kidney of a 14-day-old chicken embryo was isolated using the RNeasy kit (Qiagen, Heidelberg, Germany) following the manufacturer's instructions, and 2.5 μg were used as a template to yield cDNAs for full-length C-GRP and the shortened version of C-GRP without the N-terminal section termed C-GRP-C (encoding amino acid residues 37–171). PCR amplification of both sequences was directed by the sense primer 5′-CGCTAGGGATCCGTGGCCGAGCGGGAC-3′ (*Bam*HI restriction site underlined) for full-length C-GRP or 5′-CTGGGATCCTGCGGGCACATCAAAGGAGGG-3′ (*Bam*HI restriction site underlined) for C-GRP-C, in both cases using the antisense primer 5′-CGTACGGTCGACTCAGCCAAGTTTTGTAAGCTGAAG-3′ (*Sal*I restriction site underlined). The reaction was performed with Phusion High Fidelity Polymerase™ under conditions recommended by the manufacturer (New England BioLabs, Frankfurt, Germany). The amplification products were separated from PCR reagents by gel electrophoresis in 3% agarose, then extracted from the gel and digested with *Bam*HI/*Sal*I endonucleases. Insertions into *Bam*HI/*Sal*I-linearized pGEX-6P-2 vector (GE Healthcare, München, Germany) in frame with the present coding sequence for glutathione S-transferase (GST) directed the production of the respective fusion proteins that contain a cleavage site for the human rhinovirus 3C protease between the C-terminus of the GST and the start of C-GRP/C-GRP-C, which is then necessarily extended by the Gly-Pro dipeptide. Reconstitution of the first three amino acids (Ala, Gly and Thr for C-GRP; Val, Pro and Phe for C-GRP-C), which had to be changed to establish the *Bam*HI site by a modified QuikChange™ site-directed mutagenesis protocol (Agilent Technologies, Waldbronn, Germany), was applied. For recombinant production, pGEX-6-P-2/C-GRP or pGEX-6-P-2/C-GRP-C plasmids were transfected into *E. coli* BL21 (DE3)-pLysS cells (Promega, Mannheim, Germany). Optimal yields in both cases were obtained with incubation at 22 °C using TB medium (Roth, Karlsruhe, Germany), at a final concentration of 100 μM isopropyl β-D-thiogalactopyranoside with incubation overnight. Cells were thereafter lysed by sonification using 7 mL 20 mM phosphate-

buffered saline (PBS), pH 7.3, containing 1 mM dithiothreitol (DTT) per gram of wet cell paste. After centrifugation of the suspension, the supernatant was processed by affinity chromatography on glutathione Sepharose 4B (GE Healthcare), using 50 mM Tris–HCl buffer, pH 8.0, containing 20 mM reduced glutathione for elution of bound GST. Removal of GST after the proteolytic cleavage of the fusion proteins in 50 mM Tris–HCl buffer, pH 7.5, containing 150 mM NaCl, 1 mM EDTA and 1 mM DTT by human rhinovirus 3C protease (fused to GST; working concentration 1:100 (w/w)) was done batchwisely by incubation of the solution with glutathione-presenting Sepharose 4B beads for 16 h at 4 °C.

2.3. Analytical procedures

Gel electrophoretic analyses were carried out in SDS polyacrylamide gels with 4% stacking/15% running gels and in the two-dimensional system, as described previously [16,17]. Proteins were visualized by silver staining; measured pI values were compared to computationally calculated results using the respective program of the EXPASY (Expert Protein Analysis System) proteomics server at the Swiss Institute of Bioinformatics (Basel, Switzerland; http://web.expasy.org/compute_pi/). Matrix-assisted laser desorption/ionization (MALDI) time-of-flight (TOF) mass spectrometry (MS) was applied for C-GRP(-C) characterization in a top-down approach by in-source decay (ISD) and molecular mass determination, using sinapinic acid (SA; Bruker Daltonik, Bremen, Germany) as matrix. Protein samples were dissolved in water to reach a final concentration of 100 pmol/μL, and the solution was diluted with 0.1% trifluoroacetic acid (TFA) to a final concentration of 40 pmol/μL. Then, 0.5 μL of a saturated solution of SA in ethanol was pipetted on individual spots of the MALDI target. After drying, 1 μL of protein solution was added on top of the thin SA layer, immediately followed by 1 μL of a saturated solution of SA in 0.1% TFA with 30% acetonitrile (TA30). Spotted samples were dried at ambient temperature prior to analysis. MALDI mass spectra were obtained on an Ultraflex™ TOFTOF I instrument (Bruker Daltonik) equipped with a nitrogen laser (20 Hz), as described for engineered variants of human galectin-3 [23]. MALDI TOF MS experiments were run in the positive-ion linear mode using the following settings: ion acceleration voltage at 25.0 kV and first extraction plate at 23 kV. Reflectron ISD (reISD) spectra were recorded in the positive-ion reflectron mode using the following settings: ion acceleration voltage at 25.0 kV, reflector voltage at 26.3 kV, first extraction plate at 21.75 kV; linear ISD (linISD) spectra were acquired in the positive-ion linear mode with settings for ion acceleration voltage at 25.0 kV and first extraction plate at 23.2 kV, as carried out for an engineered tandem-repeat-type-like galectin-1 variant [24]. Experimental information from up to 5000 individual laser shots was routinely accumulated. Calibration of spectra was performed externally by linear fit between trypsinogen and protein A using protein calibration standard II for molecular mass determination and by a quadratic fit using the protein calibration standard I in the case of linISD spectra or the peptide calibration standard II in reISD spectra (Bruker Daltonik). FlexControl version 2.4 was used for instrument control and FlexAnalysis version 2.4 for processing the data of the spectra. Annotated spectra were further analyzed by BioTools 3.0 (Bruker Daltonik).

For quaternary-structure determination by gel filtration, C-GRP/C-GRP-C (100 μg in a volume of 50 μL PBS) were chromatographed on a Superose™ HR10/30 column using an ÄKTApurifier 10 system at a flow rate of 0.5 mL/min and 4 °C. The column was calibrated with the following molecular weight markers: blue dextran (Mr > 2000), aldolase (Mr = 158 kDa), albumin (Mr = 67 kDa), ovalbumin (Mr = 44 kDa), chymotrypsinogen (Mr = 25 kDa) and vitamin B_{12} (Mr = 1.35 kDa). Small-angle X-ray scattering (SAXS) data on C-GRP-C were collected on BM29 at the European Synchrotron Radiation Facility (ESRF, Grenoble, France) using the BioSAXS robot and a Pilatus 1M detector (Dectris AG, Baden-Daettwil, Switzerland) with synchrotron radiation at a wavelength λ = 0.1 nm. Protein samples were prepared at a series of concentrations of 2, 4, 6 and 8 mg/mL in PBS containing 5 mM DTT. For each measurement, ten frames were obtained at 1 s exposures of a 100 μL sample flowing continuously through a 1 mm diameter capillary during X-ray exposure, protein-free buffer processed as control. The scattering images were spherically averaged and buffer scattering intensities subtracted using in-house software. The radius of gyration (R_g) was calculated with GNOM [25], which also gives the distance distribution function P(r) and the particle maximum dimension (D_{max}). Bead models were obtained using the ATSAS software package [26]. Each model was produced from 20 runs of DAMMIN [27] that were combined and filtered to produce an averaged model using DAMAVER [28]. Analytical ultracentrifugation was run using C-GRP-C samples at concentrations of 0.2, 1 and 2 mg/mL after clearing them by a centrifugation step for 10 min at 16,000*g*. Sedimentation velocity experiments were run at 4 °C in an Optima XL-I instrument (Beckman Coulter, Krefeld, Germany) equipped with an AN50-Ti rotor at 48 krpm. Differential sedimentation coefficients were calculated by least-squares boundary modeling of the experimental data using the c(s) method implemented in the program *SedFit* version 14.7 [29].

2.4. Haemagglutination, cell assay and array binding

Haemagglutination assays with trypsin-treated, glutaraldehyde-fixed rabbit erythrocytes in 2-fold serial dilution were performed in 96-well microtiter plate wells as described [16,17]. Binding of fluorescent C-GRP(-C) (obtained by reaction of 0.25 mg/mg dye (fluorescein isothiocyanate) protein for 1 h at room temperature in 0.1 M Na_2CO_3 buffer, pH 9.0, containing 0.9% (w/v) NaCl) to cells (Chinese hamster ovary (CHO) Pro^{-5} parental line, the Lec1, Lec2, Lec8 and Lec19 glycosylation mutants and an α2,6-sialyltransferase-expressing variant of the parental line as well as human pancreatic carcinoma cells (Capan-1) expressing the tumor suppressor $p16^{INK4a}$) was analyzed by flow cytometry as described [16,30].

The glycan (50 μM)/polysaccharide (10 μg/mL) array was established by printing as described [31]. The full structure representation of the polysaccharides can be found at http://csdb.glycoscience.ru/bacterial. Biotinylated C-GRP as well as CG-1A, CG-1B and CG-2 as controls, all labeled by conjugation of biotin using the *N*-hydroxysuccinimide ester derivative (Sigma, München, Germany) under activity-preserving conditions as described [16], were comparatively tested at 50 μg/mL in PBS containing 0.1% Tween-20, 1% bovine serum albumin and 0.01% NaN_3 for 1 h at 37 °C in a humidified chamber. The glass surface had been pretreated with PBS containing 0.1% Tween-20 for 15 min. After thorough washing to remove the labeled protein, probing with fluorescent streptavidin (labeled with ALEXA-555 dye) followed for 45 min at 20 °C. After washing with PBS–0.001% Tween-20 and then with deionized water, slides were scanned on an Innoscan 1100 AL scanner (Innopsys, Carbonne, France) using an excitation wavelength of 543 nm at 10 μm resolution. The obtained data were processed using ScanArray Express 4.0 software and the fixed 70 μm-diameter circle method as well as Microsoft Excel. Ten spots represent each oligosaccharide or polysaccharide on the array, and data are reported as median RFU (relative fluorescence units) of replicates. Median deviation was measured as interquartile range. A signal, whose fluorescence intensity exceeded the background value by a factor of five, was considered as significant. The solid-phase assay using asialofetuin and proteoglycans as surface-presented test substances was performed using microtiter plates from various sources in the initial phase, then working with MaxiSorp® plates (Thermo Fisher, Darmstadt, Germany), as described [16,17].

2.5. Crystallization, data collection and refinement

Protein-containing solutions were extensively dialyzed against PBS containing 4 mM β-mercaptoethanol in a 3.5 kDa Slide-A-Lyzer™ dialysis cassette (Thermo Fisher). Any aggregates were removed by gel filtration as follows: the sample was loaded onto a HiPrep 16/60 Sephacryl S-100 high-resolution column (GE Healthcare) equilibrated with

20 mM Tris–HCl, pH 7.5, 150 mM NaCl, 1 mM EDTA and 1 mM DTT, kept at 4 °C. Eluted fractions containing C-GRP-C were checked by one-dimensional gel electrophoresis, pooled and concentrated with Centricon devices (10 kDa cut-off; Merck Chemicals, Darmstadt, Germany) to a final concentration of 16 mg/mL. The protein concentration was determined measuring its absorbance at 280 nm using the calculated extinction coefficient of 10,220 M^{-1} cm^{-1} (http://web.expasy.org/protparam/).

Systematic screening for conditions to crystallize C-GRP(-C) was performed in 96-well sitting drop plates (Swissci MRC, Suffolk, England) at 22 °C using a Cartesian Honeybee robot (Digital; Honeybee Robotics, Brooklyn, NY, USA) and commercially available 96-well kits: JBScreen Classic (Jena Bioscience™), Wizard Classic Screen I–III (Emerald Bio™) and Index Crystal Screen (Hampton Research™). The drops were 0.4 μL in volume, a mixture of 0.2 μL of the protein solution and 0.2 μL of precipitant, and they were equilibrated against 50 μL of the reservoir solution. Diffracting crystals grew after a couple of weeks in the presence of 2.0 M $(NH_4)_2SO_4$, 100 mM MES (pH 6.5) and 5% (w/v) PEG 400. Crystals were cryo-protected with the reservoir solution supplemented with a 1.0 M sodium malonate solution, mounted on nylon loops and flash-cooled in liquid nitrogen.

X-ray diffraction data were collected on a single crystal using a Pilatus 6M detector at the BL13 XALOC beamline at the ALBA synchrotron (Barcelona, Spain). A total of 960 rotation images were collected with an oscillation angle of 0.25°. The resulting data set was processed using XDS [32,33] and scaled using AIMLESS [34,35]. The crystals belonged to a body-centered orthorhombic space group $I2_12_12_1$ with cell dimensions a = 38.6 Å, b = 106.9 Å, c = 114.1 Å. The Matthews coefficient [36] assuming one molecule in the asymmetric unit was 3.91 $Å^3$ Da^{-1} (68.6% solvent content).

The structure was solved by molecular replacement using PHASER [37] and the human GRP structure (PDB code 3B9C) as a search model. The initial model was first refined with *Refmac5* [38] and alternating manual building with *Coot* [39]. The final model was obtained by repetitive cycles of refinement using *PHENIX* [37]. Solvent molecules were added automatically and inspected visually for chemically plausible positions. The placement of PEG, ethylene glycol and sulfate molecules was done using *Coot* [39]. The quality of the model was assessed by the *MolProbity* software [40], indicating that 99.2% of the residues lie in the allowed regions of the Ramachandran plot. Data collection statistics are shown in Table 1. Superimpositions were done using the *lsq* (least-squares) option in *Coot*, and structural figures were drawn with *PyMOL* (http://www.pymol.org/).

Table 1
Data collection and refinement statistics.

Data collection	
Beamline	BL13-XALOC (ALBA)
Wavelength (Å)	0.9794
Space group	$I2_12_12_1$
Unit cell parameters (Å)	a = 38.6, b = 106.9, c = 114.1
Resolution range (Å)	57.04–1.55 (1.63–1.55)
No. of observations	300,868 (40,181)
No. of unique reflections	34,844 (4882)
Multiplicity	8.6 (8.2)
Completeness (%)	99.7 (97.6)
Mean $I/s(I)$	12.6 (2.2)
R_{merge} [a]	0.069 (0.977)
R_{meas} [b]	0.074 (1.043)
$CC_{1/2}$ [c]	99.9 (84.8)
Mosaicity	0.20
Wilson B-factor	21.62
Refinement	
R_{work}/R_{free}	0.168 (0.267)/0.189 (0.30)
Working reflections	33,160 (2285)
Testing reflections	1593 (101)
Protein atoms (non-H)	1064
PEG and ethylene glycol	36
Sulfate ions	10
Water molecules	102
Mean B factors ($Å^2$)	
Protein	27.23
PEG and ethylene glycol	57.45
Sulfate ions	58.01
Water molecules	42.42
rmsd bond lengths (Å)	0.008
rmsd angles (°)	1.36
Ramachandran plot statistics	
Favored (%)	99.2
Outliers (%)	0
PDB accession ID	5IT6

Statistics for the highest-resolution shell are shown in parentheses.
[a] $R_{merge} = S_{hkl}S_i\,|I_i(hkl) - <I(hkl)>|\,/\,S_{hkl}S_iI_i(hkl)$.
[b] $R_{meas} = \Sigma_{hkl}\,(N-1)^{-1/2}\,\Sigma_i|I_i(hkl) - <I(hkl)>|\,/\,\Sigma_{hkl}\,\Sigma_i\,I_i(hkl)$, where $I_i(hkl)$ is the intensity measured for the *i*th reflection and $(I(hkl))$ is the average intensity of all reflections with indices *hkl*.
[c] $CC_{1/2}$ is the correlation coefficient between two random half datasets [84].

3. Results and discussion

3.1. GRP on the level of the gene

Genes for galectins are widely present in organisms from hydrozoans to vertebrates so that a full-scope search in eukaryotes needed to be performed for GRP. Database mining reveals that the presence of the GRP gene is strictly limited to vertebrates. By taking the number of studied species on the level of the genome to 123 and on the level of the amino acid sequence to 119, as listed in Fig. S1, it is evident that the gene is broadly present in vertebrates (Fig. 1). In contrast, occurrence of canonical galectins is known to occur in other organisms such as in fungi, nematodes or insects or, in special cases such as galectins-5 or -6, restricted to a single species among vertebrates (rat or mouse, respectively) [10,11]. In all studied genomes, the copy number of the gene was one, a parameter that is variable among canonical family members and species [18]. The chromosomal environment revealed no vicinity to any of the genes for canonical galectins in the chicken genome (Fig. S2A). In inter-species comparison, it was quite similar (Fig. S2B), an indicator for the gene's origin prior to intra-vertebrate divergence and a lack of positional dynamics. The organization of the gene in exons/introns, shown in Fig. S3, reflected its status as a member of the galectin family, as already indicated previously [41]. As such, no signal sequence was present to direct the protein to the secretory route. Thus, the GRP gene is not a recent acquisition by duplication of an ancestral gene within a distinct species but an integral component of the vertebrate genome.

The compilation of predicted amino acid sequences confirmed and extended the assumption for a strongly positive selection. The region of the so-called carbohydrate recognition domain (CRD) following the N-terminal tail was highly conserved, as was the second part of the 36-amino-acid N-terminal section (Fig. S4). A peculiar feature is seen at the central position in the CRD of canonical galectins, i.e. the Trp moiety, which establishes C–H/π-contacts to a ligand's galactose residue: birds, fish and amphibians have this amino acid in its place, whereas mammals consistently present a substitution by Arg/Lys (Fig. S4). At other positions, GRP sequences appear exceptionally conserved, validating the initial observation that human and mouse GRP "only differ in one amino acid, residue 24" [10]. Among the CGs, the predicted amino acid sequence of C-GRP is rather distant from those of the three prototype CGs (Fig. 2). Based on these comparisons, it is most closely related to the C-terminal CRD of the tandem-repeat-type CG-8, termed CG-8C (Fig. 2). For mammals, the likewisely calculated phylogenetic tree had placed GRP in closest vicinity of the C-terminal CRD of tandem-repeat-type galectin-12 [41]. This family member is not represented in the chicken genome. It is an effector of growth regulation via induction of apoptosis and cell cycle arrest at the G_1 phase in murine 3T3-L1 adipocytes, exhibiting nuclear and cytoplasmic localization and marker status

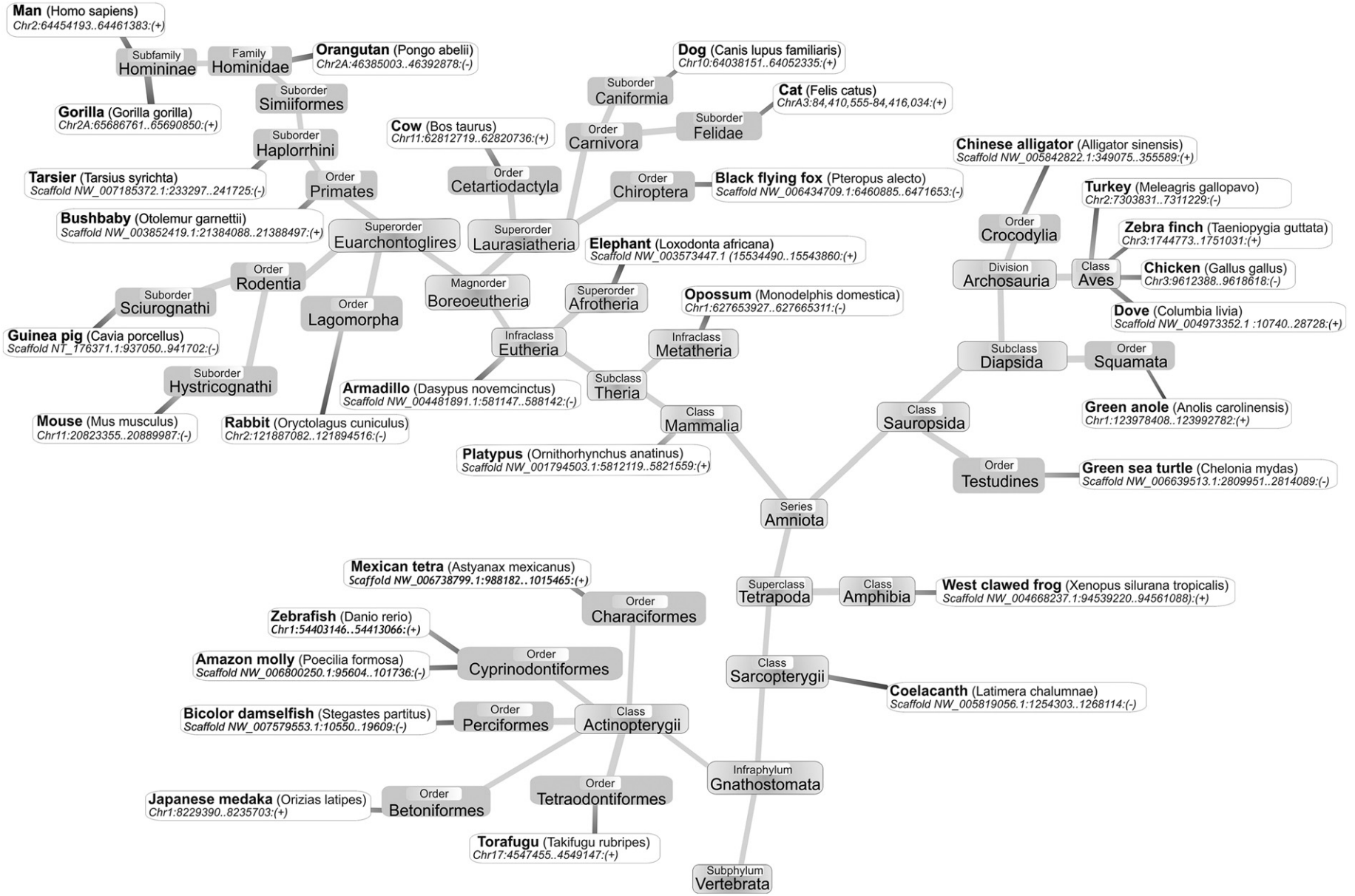

Fig. 1. Compilation of information on GRP gene in man and 31 vertebrate species from different branches of the taxonomic tree available in databases. Database-retrieved information on the genes, present either in forward (+) or in reverse (−) direction, is listed as chromosome or scaffold number together with the number of nucleotides, characterizing the precise position and length of the gene.

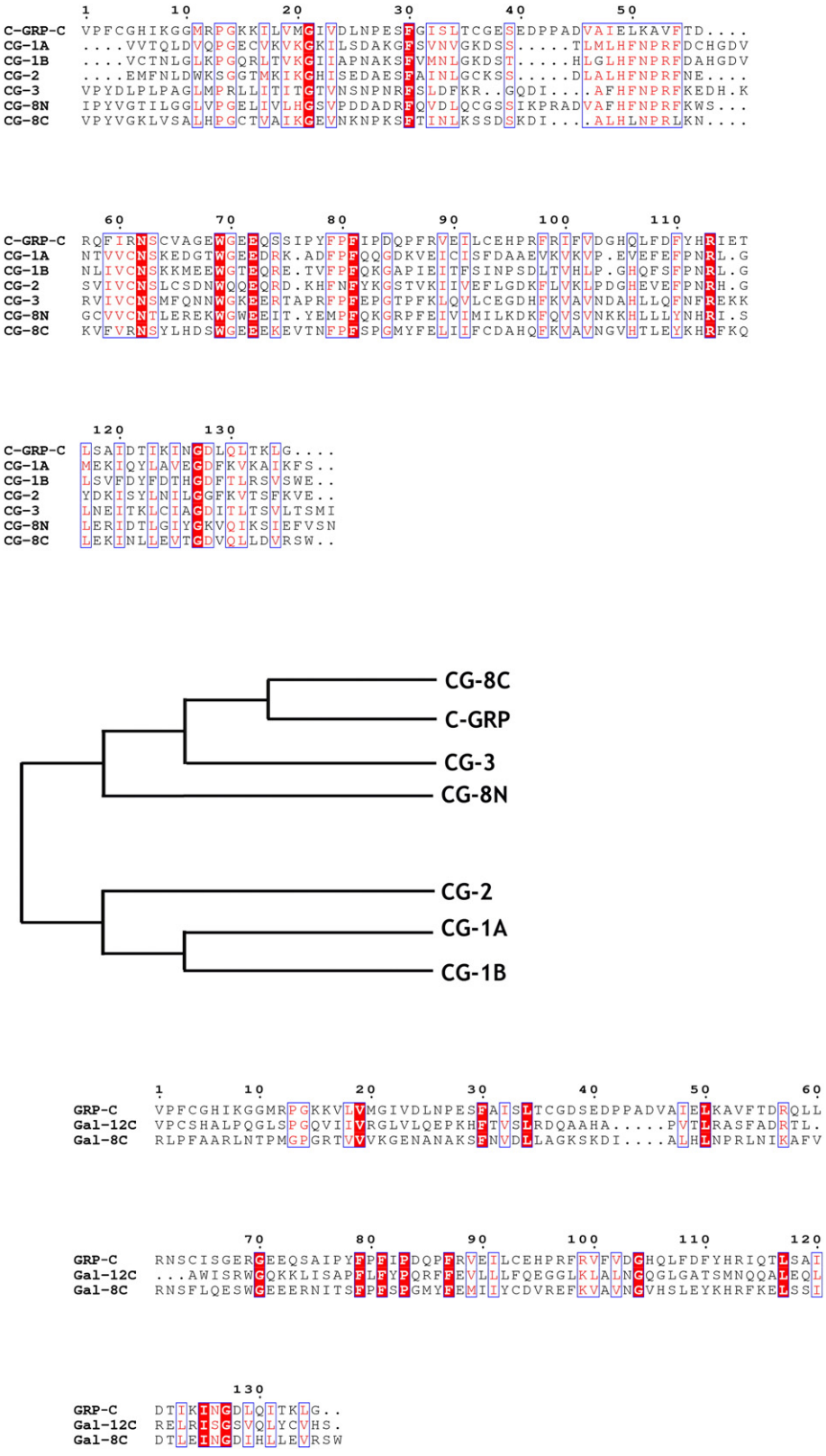

Fig. 2. Comparison of amino acid sequences of the CRDs for the shortened version of C-GRP, i.e. C-GRP-C (amino acid 37 is set to position 1 for alignment), versus the five canonical CGs (top), the respective phylogenetic family-tree diagram (middle) and the alignment of the amino acid sequence of the shortened version of human GRP, i.e. human GRP-C (amino acid 37 is set to position 1 for alignment) with those of the CRDs of human Gal-8 and -12, referred to as Gal-8C/-12C (bottom). Strictly conserved (red background) and homologous residues (>70% conservation; boxed red letters) are highlighted by coloring.

in butyrate-induced differentiation of human colon cancer cells [42–44]. Looking at the sequence signature for binding lactose, marked deviations yet occur (please see Section 3.3. on crystallography for details), raising the question on C-GRP's lectin activity. In order to study protein properties such as this feature, recombinant expression of C-GRP had to be established.

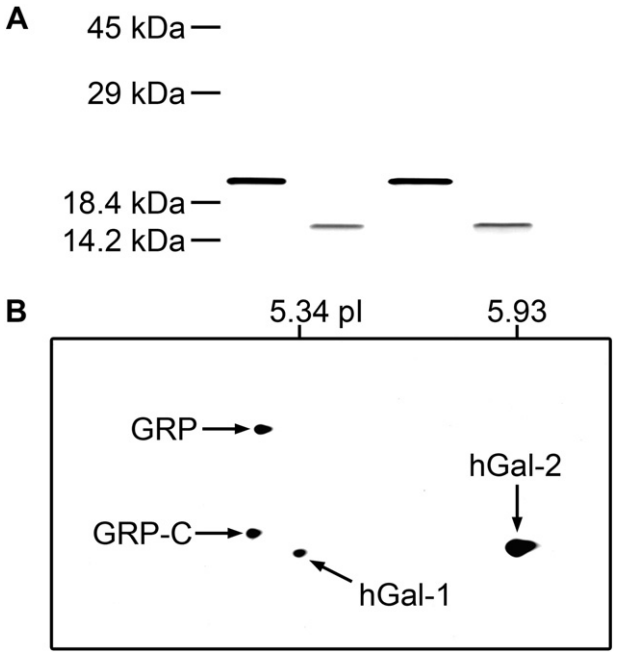

Fig. 3. Documentation of one- and two-dimensional gel electrophoretic analysis of recombinant C-GRP and C-GRP-C. (A) Electrophoretic mobility of purified C-GRP and C-GRP-C in one-dimensional polyacrylamide gel electrophoresis under denaturing conditions in the presence of 2-mercaptoethanol in a 15% running gel. The bands for C-GRP and C-GRP-C are shown in pairwise arrangement (60 and 80 ng per lane, respectively). Positions of relevant marker proteins are given on the left (ovalbumin, 45 kDa; carbonic anhydrase, 29 kDa; β-lactoglobulin, 18.4 kDa; lysozyme, 14.2 kDa). (B) The relevant section of a gel after two-dimensional gel electrophoresis and staining is shown. The given section presents the spots for C-GRP and C-GRP-C together with those of human galectins-1 (pI 5.34) and -2 (pI 5.93).

3.2. C-GRP on the level of the protein

Recombinant production led to an inducible expression of a protein at the expected position in gel electrophoretic analysis of extracts that failed to bind to the affinity resin commonly used for galectin purification (i.e. Sepharose 4B presenting lactose attached to the beads after their activation by divinyl sulfone). As consequence, C-GRP purification required its expression as fusion protein. Among the tested systems, involvement of GST via a rhinoviral C3 protease-sensitive cleavage site proved most favorable. It yielded C-GRP of calculated mobility and isoelectric point, as measured in gel electrophoretic analysis, at good yield (about 30 mg/mL) (Fig. 3). Mass spectrometry took the analysis of the purified protein to the most sensitive level, ascertaining the expected mass of the product by MALDI TOF runs as well as the sequence by peptide fingerprinting and re/linISD processing (Fig. 4, Fig. S5, Table S1A).

To determine its quaternary structure, gel filtration was performed and revealed elution nearly exclusively at the position of the monomer (Fig. 5A). Presence of lactose, commonly applied in this protocol when exposing canonical galectins to the resin to preclude retardation of elution by carbohydrate-dependent interaction of the protein with the beads, did not affect the elution volume (Fig. 5A). Thus, C-GRP, formally a proto-type protein based on its sequence features, is the first member of this group in chicken, which is monomeric in solution.

Considering its monomeric status in solution and its sequence addition to the CRD, C-GRP resembles rat galectin-5 from the group of prototype galectins. Galectin-5, too, is monomeric, as measured by gel filtration and nano-electrospray ionization MS [45,46]. This proto-type lectin unique for rat is constituted by a short segment of the N-terminal portion of galectin-9 and the C-terminal CRD of this tandem-repeat-type protein, functionally involved in exosome generation and routing during rat reticulocyte maturation [45,47,48]. However, galectin-5 is a lectin. It binds lactose and other β-galactosides, this activity leading to haemagglutination of rabbit erythrocytes at 150–300 μg/mL [45,49]. Similarly, the chimera-type galectin-3 aggregated erythrocytes at a minimal concentration of 25 μg/mL, pointing to a role of its non-triple-helical collagen-like repeats in ligand-mediated cell aggregation (for overview on structural aspects of this lectin, please see [23]). Tested as further controls, homodimeric proto-type CG-1A and -2 at concentrations of 10–15 ng/mL or 140–160 ng/mL, respectively, led to a positive result, as tandem-repeat-type CG-8 did at around 400–500 ng/mL, corroborating published data [15,16]. In contrast, C-GRP was consistently inactive in this assay at concentrations up to 500 μg/mL. Obviously, capacity for glycan-dependent cell binding and/or aggregation are impaired in the case of C-GRP. Of note, deviations in positions of the signature sequence do not necessarily mean that a member of this family will lose its lectin activity. In the cases of congerin P, the galectin of peritoneal cells of the conger eel [50], and the third protein of the galectin family in the inky cap mushroom *Coprinopsis cinerea* termed CGL3 [51], glycan binding is still operative despite such alterations. These precedents prompted us to perform further assays using a microarray and cytofluorimetry of cells with defined changes in their surface glycomes.

The microarray presents a total of 642 compounds with about 600 different glycans as mono- to oligosaccharides, bacterial (lipo)polysaccharides, glycosaminoglycans and synthetic (glyco)peptides to cover a wide spectrum of glycan structures (for complete listing, please see Table S2). Human tandem-repeat-type galectins have previously been successfully probed in this setting for selectivity among bacterial polysaccharides [52], and CG-1A/-1B/-2 were run in parallel as positive controls. Under these conditions, few signals with significant intensity were obtained by testing: 3′-sulfated Tn antigen (but no other sulfated saccharide or sialylated Tn), β-linked ribose, a bacterial polysaccharide, the α-anomer of the TF antigen (CD176), this disaccharide with unnatural β2,6-sialylation at the core and *N*-acetyllactosamine (but not its dimer, which is reactive with all canonical CGs, whereas the disaccharide is not a binder for them in the array) (for detailed listing of fluorescence data, please see Table S2). Further probing of the labeled protein in a solid-phase assay with glycoprotein (the potent galectin ligand asialofetuin) and proteoglycans revealed absence of reactivity to the β-galactoside termini of *N*-glycan chains and indicated weak reactivity to heparin depending on the type of microtiter plate (not shown). Evidently, *N*-acetyllactosamine of *N*-glycans and in repeats does not associate with C-GRP, in contrast to binding to canonical CGs, whereas the disaccharide did associate in the array.

Turning to cell surface binding, flow cytofluorimetry was performed using CHO parental cells and four glycosylation mutants, a sensitive set of tools to trace glycan-dependent binding activity [53]. Fluorescent C-GRP was found to associate with the surface of CHO cells in a concentration-dependent manner (Fig. 6A). Taking note of the fact that the anionic dye fluorescein itself can interact with scavenger receptors and also non-specifically [54], rigorous controls were added to document very low-level binding of an inert protein with this label, inhibition by lactose for binding of labeled CGs and loss of activity of C-GRP by heat treatment (not shown). In contrast to canonical CGs, the extent of binding (measured as percentage of positive cells/mean fluorescence intensity) to CHO cells was not affected by all tested glycome alterations, i.e. reduced presence of complex/hybrid-type *N*-glycans (Lec1) (Fig. 6B), introduction of α2,6-sialylation to *N*-glycans, a molecular switch blocking galectin binding (Fig. 6C), reduced level of overall sialylation (Lec2) (Fig. 6D) and reduced level of galactosylation (Lec8, Lec19; the latter also affecting glycoprotein routing and surface presentation [55]) (Fig. 6E, F). In addition to a lack of effect of any tested alteration in the surface glycome, cell binding to parental and Lec2/8 cells was not affected at all by presence of glycocompounds tested as potential inhibitors, i.e. lactose, *N*-acetylglucosamine, D-glucuronic acid, L-fucose, D-mannose, maltose and L-rhamnose at up to 100 mM, heparin at up

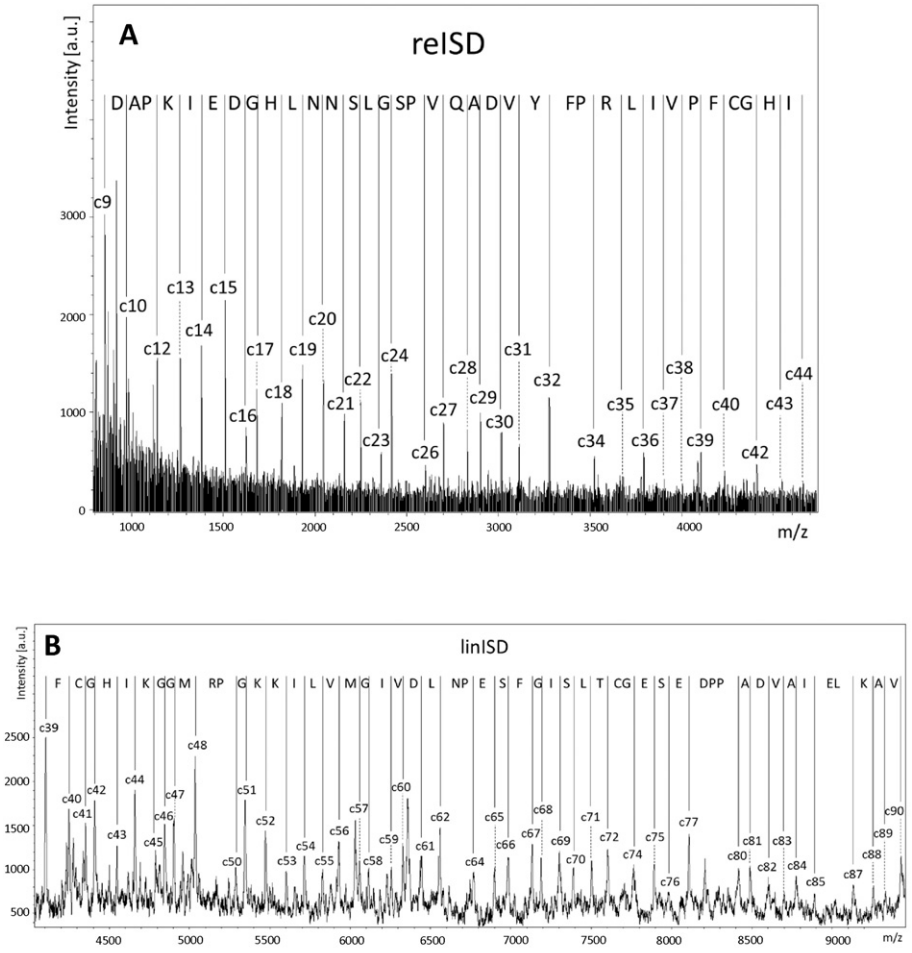

Fig. 4. Sequencing of the N-terminus of C-GRP by reISD and linISD. (A) reISD mass spectrum of C-GRP shows the smallest detectable peptide at the c9/c10 and the ensuing stepwise peptide ladder. (B) linISD spectrum shows the smallest detectable peptide at c39/c40 and the ensuing stepwise peptide ladder up to c89/90. (C) Sequence coverage by ISD. Experimental and calculated mass values are given in Table S1A.

to 250 μg/mL as well as chondroitin sulfate and carageenan at 100 μg/mL (not shown). C-GRP reactivity was also measured to human pancreatic carcinoma (Capan-1) cells, which likewise was not inhibitable by glycocompounds. Canonical galectins such as CG-8 avidly associate with them in a lactose-inhibitable manner [16,56]. This study part thus provides no evidence for a galectin-like reactivity to a common glycan, tested either with glycocompounds or glycosylation mutants as ligands (microarray, cell assay) or inhibitor (cell assay). Also, the cell assays in the presence of glycans exclude the possibility of an allosteric activation of lactose binding by the test substances, revealed previously for mannose in the case of the mentioned congerin P [50]. In consequence, C-GRP appears to have lost lectin activity despite the presence of the central Trp residue, which separates avian from mammalian GRP (Fig. S4). Considering i) the strong sequence conservation that implies a common functionality, ii) this Trp/Arg substitution at a central site and iii) availability of structural information on an engineered variant of human GRP by crystallography [57,58], we next set the aim to report such structural details for C-GRP.

3.3. Crystal structure of a C-GRP variant

Initial attempts to crystallize C-GRP were not successful. Whereas crystals of proto-type galectins and of separate galectin CRDs had been obtained, presence of sequence extensions of a CRD, either in tandem-repeat-type or chimera-type proteins, had proven unfavorable, except for a single case: a variant of human galectin-3 constituted by the CRD and a tail of three non-triple-helical collagen-like repeats and the N-terminal peptide with its two sites for Ser phosphorylation [59]. Evidently, for C-GRP, a well-ordered structure of the N-terminal extension could not be attained to enable crystal formation.

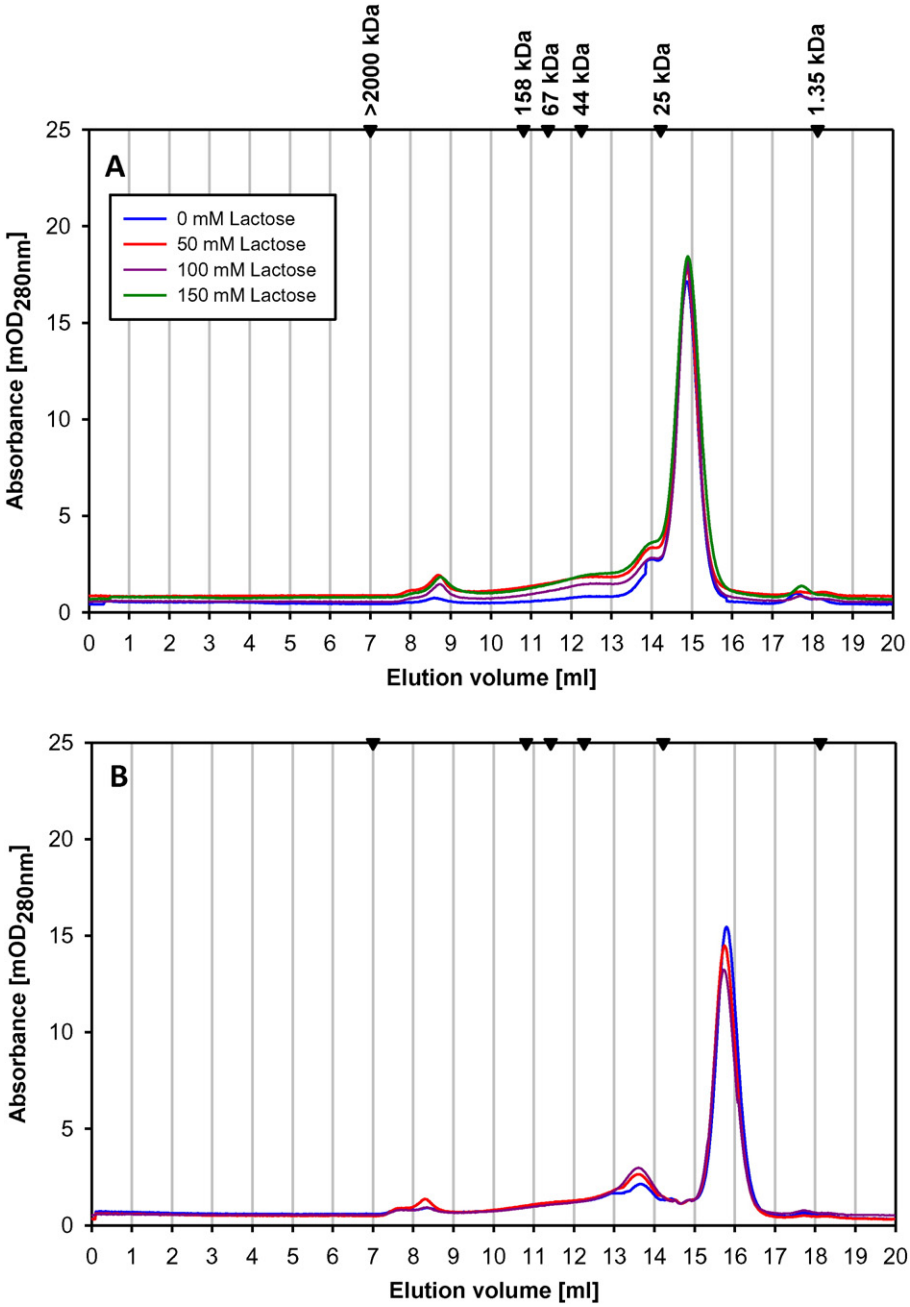

Fig. 5. Gel filtration analysis of C-GRP and C-GRP-C. (A) Elution profiles of the physiological protein (C-GRP with molecular mass of 19,069 Da; please see Fig. S5A for MALDI TOF MS-based information) and (B) its shortened version (C-GRP-C with molecular mass of 15,339 Da; please see Fig. S6A for respective MS-based information) in the absence and in the presence of lactose at increasing concentrations. Black arrowheads indicate elution volume of molecular weight markers used for calibration.

Therefore, as previously performed for His-tagged human GRP [60], this sequence portion was deleted (total of 36 amino acids) to produce a shortened form termed C-GRP-C, in analogy to human GRP-C [57,58]. This variant underwent the same analytical protocols, to prove its purity in gel electrophoresis (Fig. 3), its sequence on the level of amino acids (Fig. S6, Fig. S7, Table S1B) and its quaternary structure (Fig. 5B). Tested at loading concentrations in the range of 0.2–2 mg/mL in sedimentation equilibrium analyses, C-GRP and the shortened C-GRP-C maintained the same monomeric status, as demonstrated by sedimentation velocity experiments. In detail, the calculated sedimentation coefficients were 1.137 ± 0.008 with a frictional ratio of 1.52 for C-GRP (Fig. S8A) and 1.161 ± 0.006 S with a frictional ratio of 1.30 for C-GRP-C (Fig. S8B). The monomer status was independently confirmed by SAXS for C-GRP-C. Here, experimental data enabled to calculate a monomer with a maximum dimension of 48.2 ± 2.3 Å at a molecular mass of 16.2 ± 1.6 kDa, the ab initio model in solution providing an internal standard for the crystal structure (Fig. S9). When probed for cell-binding capacity, always run in parallel with the physiological form to ensure identical conditions, fluorescent C-GRP-C showed very similar reactivity (not shown). These data preclude an influence of the non-CRD portion for this property.

The crystal structure of C-GRP-C was determined at 1.55 Å resolution by molecular replacement (for details on refinement statistics, please see Table 1). The typical galectin fold is composed of two anti-parallel five-stranded (F1–F5) and six-stranded (S1–S6) β-sheets, along with a short 3_{10} helix between the F5/S2 strands (Fig. S10). As in solution, C-GRP-C in crystals is monomeric. Human GRP-C, in gel filtration also monomeric [57,58], formed a dimer [57] or a dimer of dimers [58] in

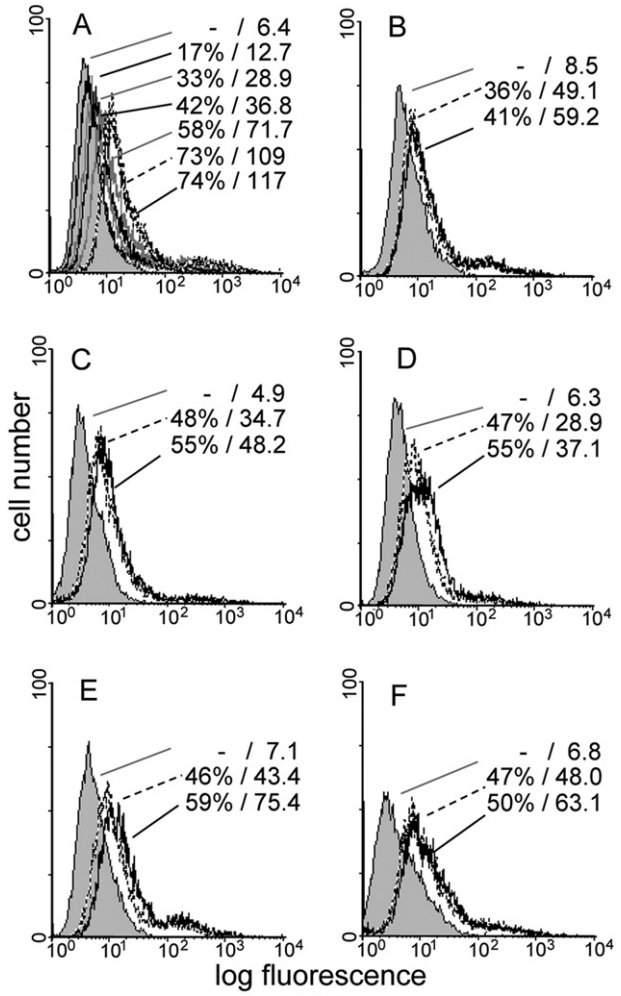

Fig. 6. Semilogarithmic representation of fluorescent cell surface staining by labeled C-GRP. Quantitative data on percentage of positive cells (%) and mean fluorescence intensity are given for each curve in each panel. The control value (background) is drawn as grey-shaded area. (A) Staining of parental CHO Pro^{-5} (lacking expression of β1,4-galactosyltransferase VI) cells measured with increasing concentrations of C-GRP from 2 μg/mL to 5 μg/mL, 10 μg/mL, 20 μg/mL and 40 μg/mL (from top to bottom) and inhibition of staining (40 μg/mL C-GRP) by a lactose concentration of 10 mM (dashed line). (B–F): Staining with 10 μg/mL (dashed line) and 20 μg/mL C-GRP (black line). Staining profiles of CHO glycosylation mutant and transfected lines are shown: Lec1 (decreased expression of complex- and hybrid-types *N*-glycans) (B), parental cells over-expressing α2,6-sialyltransferase I (C), Lec2 (decreased level of sialylation) (D), Lec8 (decreased level of galactosylation) (E) and Lec19 (decreased level of galactosylation) (F).

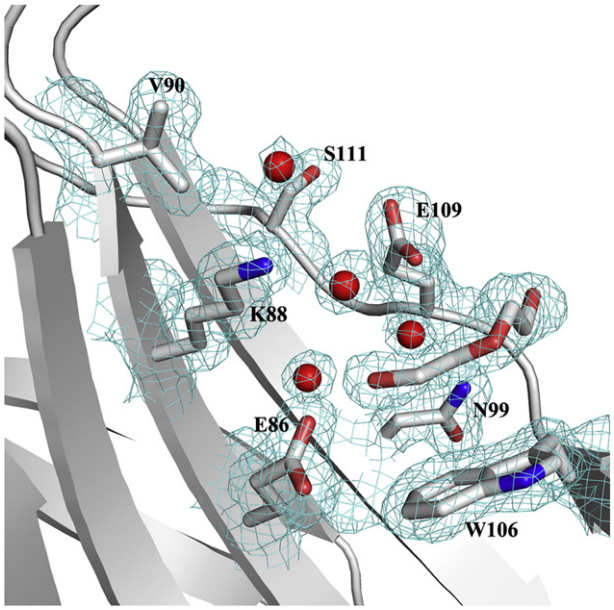

Fig. 7. Close-up view of the putative contact site for a β-galactoside in C-GRP-C. Respective residues in C-GRP-C and an ethylene glycol molecule are show in ball-and-stick mode, water molecules are depicted as red spheres. The observed electron density map 2Fo-Fc is contoured at 1.0 σ.

crystals, depending on the actual experimental conditions. Such a sensitivity of quaternary structure on the environment had been reported in the case of human galectin-1 and its appearance as dimer of dimers in an aprotic solvent [61]. In terms of the root-mean-square difference (rmsd) for positions of the Cα atoms, values between 1.07 Å and 1.54 Å relative to the structures of CG-8N (1.07 Å; [62]), CG-2 (1.33 Å; [63]), CG-1A (1.42 Å; [64]) and CG-1B (1.54 Å; [65]) underscore the close relationship. The superimpositions of the C-GRP-C fold with that of the listed canonical CGs, however, disclosed notable differences in several regions involving loops that are between adjacent β-strands in the concave face of the groove, the site of accommodation of lactose in canonical CGs (Fig. 8). In detail, the S3–S4 loop, which connects anti-parallel β-strands F3–F4, is five amino acids longer in C-GRP-C (and CG-8N) than in proto-type CGs, for whom (CG-1A/B) the S4-S5 loop is extended (Fig. S11).

Looking at the equivalent of the contact site for lactose in canonical CGs, the sequence signature for operative binding (His45, Asn47, Arg49, Asn58, Trp65, Glu68 and Arg70 for example in CG-2 [63]) is turned into Glu49, Lys51, Val53, Asn62, Trp69, Glu72 and Ser74 in C-GRP-C. Only three of seven positions are thus maintained (Fig. 7). As shown in Fig. 8, the set of contacts of the 4′, 6′-hydroxyls of galactose will be conspicuously reduced (in comparison to CG-2), and – as in human GRP-C with rmsd values of 1.06 Å - 1.59 Å relative to galectins-1, -3 and -7 [57,58] – Lys51 protrudes to an extent that

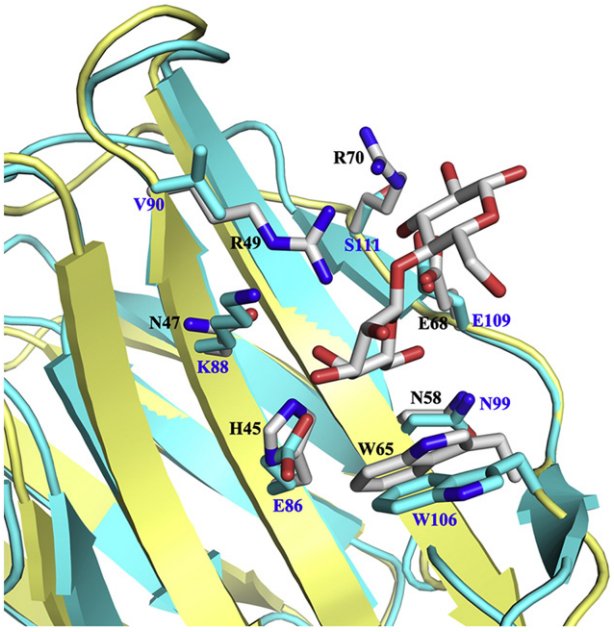

Fig. 8. Superposition of the structure of the contact site for a β-galactoside in CG-2 with the structurally equipositioned of C-GRP-C shown in Fig. 7. Amino acids of CG-2 are colored in yellow and those of C-GRP-C in cyan. Residues making contact to lactose of CG-2 and those of equivalent positions of C-GRP-C (Glu49, Lys51, Val53, Asn62, Trp69, Glu72 and Ser74) are shown in ball-and-stick mode, and their carbon atoms are colored in white and cyan, respectively. Lactose is shown in stick mode.

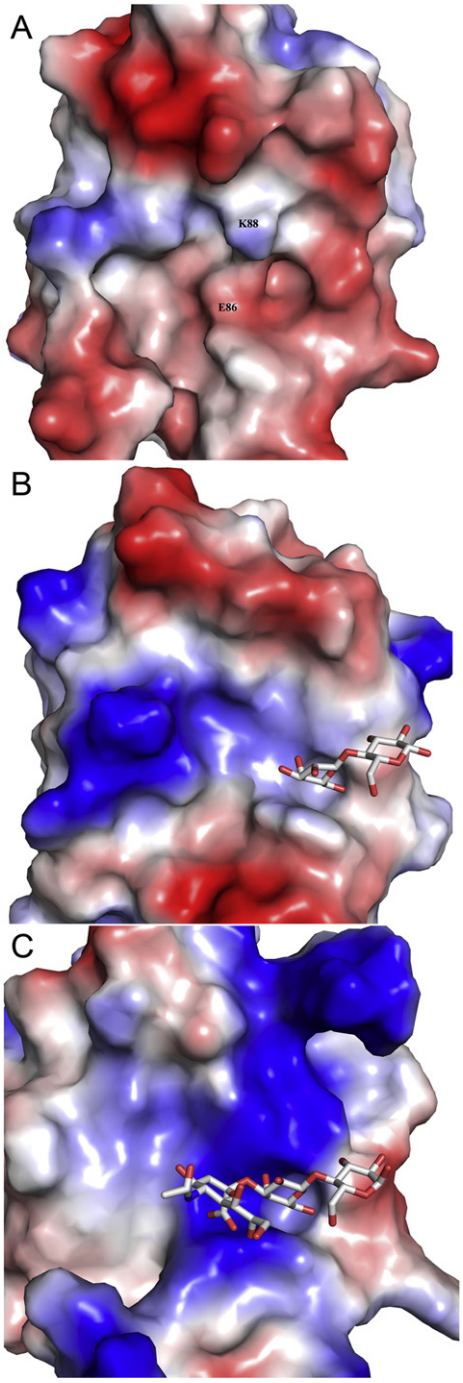

Fig. 9. Electrostatic surface potential maps of C-GRP-C (A), CG-2 (B) and CG-8N (C), contoured from −10 kT/e (intense red) to +10 kT/e (intense blue). Differences in distribution of acidic residues between C-GRP-C and the two canonical CRDs as well as absence of a tunnel-like cavity present in CG-2 and CG-8N, which accommodates the galactose ring, are readily noted. Lactose and 3′-sialyllactose are shown in stick mode.

disfavors a snug fit for lactose so that not even the presence of Trp69 can compensate this distortion, these factors explaining the loss of lectin activity. The prominent Lys51 position, along with a comparative view on the electrostatic surfaces and absence (C-GRP-C)/presence (CG-2/CG-8N) of a tunnel-like cavity hosting galactose, is illustrated in Fig. 9. That binding to lactose is no longer possible at this site does not mean that C-GRP(-C) is devoid of capacity for specific interactions. A close look into the literature attests that new virtues are acquired by a sequence remodeling of galectins.

In fact, illustrating the plasticity of the galectin fold to serve as platform for building complementarity to protein surfaces, this route has likely been taken in the development of coronavirus spike protein N-terminal domains, after they were imported from the host [66,67]. A constellation with hydrophilic/hydrophobic patches, as it is also seen in C-GRP-C, is suggestive to facilitate a contact area for a specific binding partner. The case study of the galectin fold of mouse hepatitis coronavirus, describing details of its interaction with murine carcinoembryonic antigen-related cell adhesion molecule 1a, teaches the lesson that hydrogen bonding and a bifurcated salt bridge "help bring the adjacent hydrophobic patches into place" [66]. Other examples of protein–protein interactions by the galectin fold concern mammalian galectins-1, -3 and -7, with oncogenic Ras proteins or anti-apoptotic Bcl-2/pro-apoptotic Bax, in this case likely involving the central Trp residue for contact [68–72]. Similarly, the micronemal protein 1 (MIC1) of the protozoan parasite *Toxoplasma gondii* harbors a galectin-like fold in its C-terminal portion, which binds the micronemal protein 6 by protein-protein recognition, as likewise seen for MIC2/MIC2AP binding [73,74]. Most intriguingly in view of the close relationship between GRP and galectin-8's C-terminal CRD, this CRD of the human lectin associates with the cargo receptor NDP52 in the process of inducing anti-bacterial autophagy, again via protein-protein recognition [75,76].

Equally important, already a single-site sequence deviation, and this far away from contact positions to the ligand, can alter a galectin's functionality. This principle is illustrated for a natural variant of human galectin-8 arising by single nucleotide polymorphism (F19Y), measured in terms of thermodynamics of ligand binding, its association with autoimmune disease (rheumatoid arthritis, myasthenia gravis) and reduction in its bridging capacity [77–79]. Sequence conservation at this high level, as seen for the GRP genes in vertebrates (Fig. S4), is a sign for the development of a distinct function in the monomeric protein at the expense of the canonical lectin activity. Conceptually, this insight and the given detailed biochemical characterization of the chicken protein pave the way to advance the galectin fingerprinting in this model organism from the canonical CGs to define the where of C-GRP expression, and this in relation to all canonical CGs. This endeavor will at the same time likely teach instructive lessons for the understanding of more complex systems. Moving from localizing a single galectin to increasing the scope of analysis to other family members in disease states already attests the potential of this approach for human galectins [80,81].

4. Conclusions

A detailed view on GRP genes reveals its presence exclusively in vertebrates, and here with high-level sequence conservation. In contrast to most other proto-type galectins, C-GRP is monomeric. In bead, microarray, solid-phase and cell assays used to probe into glycan-binding activity, no activity was detected. C-GRP thus is the first member of the CG family without lectin activity. In consequence, its place in the listing of galectins in the phylogenetic tree is best as proto-type-like GRP, in analogy to proteins with C-type lectin-like domains that have lost capacity to bind sugars but gained reactivity to other types of epitopes [82]. The next step toward tracing GRP's mission that explains the absence of usual sequence drifts among vertebrate species will now be to define the expression pattern of C-GRP in tissues and its relation to the five canonical CGs as well as to trace tissue reactivity. The corresponding results are reported in the second part of our report [83].

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.bbagen.2016.06.001.

Transparency Document

The Transparency document associated with this article can be found, in online version.

Acknowledgements

We cordially thank the staff at XALOC beamline (ALBA Synchrotron) and the BM29 beamline (ESRF) for their support, Drs. Y. Knirel, P. Kosma and H. Kunz for their gifts of glycocompounds (bacterial polysaccharides and lipopolysaccharides as well as synthetic glycopeptides), and Drs. J. Domingo-Ekark, B. Friday, M. Ilsüm and A. Leddoz for inspiring discussions as well as the reviewers for their valuable advice. This work was generously supported by funding by grants BFU2014-55448-P and CSD2009-00088 from the Spanish Ministry of Science and Innovation, by the Regional Government of Madrid (S2010/BMD-2353), by the Russian Science Foundation (14-50-00131) and by the European Union's Seventh Framework Program FP7/2007-2013 under REA grant agreement no. 317297 ("GLYCOPHARM").

References

[1] H.-J. Gabius, The magic of the sugar code, Trends Biochem. Sci. 40 (2015) 341.
[2] J.M. Rini, Lectin structure, Annu. Rev. Biophys. Biomol. Struct. 24 (1995) 551–577.
[3] R. Loris, T. Hamelryck, J. Bouckaert, L. Wyns, Legume lectin structure, Biochim. Biophys. Acta 1383 (1998) 9–36.
[4] H.-J. Gabius, S. André, H. Kaltner, H.-C. Siebert, The sugar code: functional lectinomics, Biochim. Biophys. Acta 1572 (2002) 165–177.
[5] M.E. Taylor, K. Drickamer, Convergent and divergent mechanisms of sugar recognition across kingdoms, Curr. Opin. Struct. Biol. 28 (2014) 14–22.
[6] H.-J. Gabius, H. Kaltner, J. Kopitz, S. André, The glycobiology of the CD system: a dictionary for translating marker designations into glycan/lectin structure and function, Trends Biochem. Sci. 40 (2015) 360–376.
[7] D. Solís, N.V. Bovin, A.P. Davis, J. Jiménez-Barbero, A. Romero, R. Roy, K. Smetana Jr., H.-J. Gabius, A guide into glycosciences: how chemistry, biochemistry and biology cooperate to crack the sugar code, Biochim. Biophys. Acta 1850 (2015) 186–235.
[8] S.H. Barondes, Galectins: a personal review, Trends Glycosci. Glycotechnol. 9 (1997) 1–7.
[9] J. Hirabayashi (Ed.), Recent topics on galectins, Trends Glycosci. Glycotechnol., 9 1997, pp. 1–180.
[10] D.N.W. Cooper, Galectinomics: finding themes in complexity, Biochim. Biophys. Acta 1572 (2002) 209–231.
[11] H. Kaltner, H.-J. Gabius, A toolbox of lectins for translating the sugar code: the galectin network in phylogenesis and tumors, Histol. Histopathol. 27 (2012) 397–416.
[12] Q.H. Zhang, M. Ye, X.Y. Wu, S.X. Ren, M. Zhao, C.J. Zhao, G. Fu, Y. Shen, H.Y. Fan, G. Lu, M. Zhong, X.R. Xu, Z.G. Han, J.W. Zhang, J. Tao, Q.H. Huang, J. Zhou, G.X. Hu, J. Gu, S.J. Chen, Z. Chen, Cloning and functional analysis of cDNAs with open reading frames for 300 previously undefined genes expressed in $CD34^{+}$ hematopoietic stem/progenitor cells, Genome Res. 10 (2000) 1546–1560.
[13] E.C. Beyer, S.E. Zweig, S.H. Barondes, Two lactose-binding lectins from chicken tissues. Purified lectin from intestine is different from those in liver and muscle, J. Biol. Chem. 255 (1980) 4236–4239.
[14] Y. Oda, K.-i. Kasai, Purification and characterization of β-galactoside-binding lectin from chick embryonic skin, Biochim. Biophys. Acta 761 (1983) 237–245.
[15] H. Kaltner, D. Solís, J. Kopitz, M. Lensch, M. Lohr, J.C. Manning, M. Mürnseer, M. Schnölzer, S. André, J.L. Sáiz, H.-J. Gabius, Proto-type chicken galectins revisited: characterization of a third protein with distinctive hydrodynamic behaviour and expression pattern in organs of adult animals, Biochem. J. 409 (2008) 591–599.
[16] H. Kaltner, D. Solís, S. André, M. Lensch, J.C. Manning, M. Mürnseer, J.L. Sáiz, H.-J. Gabius, Unique chicken tandem-repeat-type galectin: implications of alternative splicing and a distinct expression profile compared to those of the three proto-type proteins, Biochemistry 48 (2009) 4403–4416.
[17] H. Kaltner, D. Kübler, L. López-Merino, M. Lohr, J.C. Manning, M. Lensch, J. Seidler, W.D. Lehmann, S. André, D. Solís, H.-J. Gabius, Toward comprehensive analysis of the galectin network in chicken: unique diversity of galectin-3 and comparison of its localization profile in organs of adult animals to the other four members of this lectin family, Anat. Rec. 294 (2011) 427–444.
[18] H. Kaltner, A.-S. Raschta, J.C. Manning, H.-J. Gabius, Copy-number variation of functional galectin genes: studying animal galectin-7 (p53-induced gene 1 in man) and tandem-repeat-type galectins-4 and -9, Glycobiology 23 (2013) 1152–1163.
[19] D. Maglott, J. Ostell, K.D. Pruitt, T. Tatusova, Entrez Gene: gene-centered information at NCBI, Nucleic Acids Res. 39 (2011) D52–D57.
[20] F. Sievers, A. Wilm, D. Dineen, T.J. Gibson, K. Karplus, W. Li, R. Lopez, H. McWilliam, M. Remmert, J. Soding, J.D. Thompson, D.G. Higgins, Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega, Mol. Syst. Biol. 7 (2011) 539.
[21] A.M. Waterhouse, J.B. Procter, D.M. Martin, M. Clamp, G.J. Barton, Jalview Version 2: a multiple sequence alignment editor and analysis workbench, Bioinformatics 25 (2009) 1189–1191.
[22] K. Tamura, G. Stecher, D. Peterson, A. Filipski, S. Kumar, MEGA6: molecular evolutionary genetics analysis version 6.0, Mol. Biol. Evol. 30 (2013) 2725–2729.
[23] J. Kopitz, S. Vértesy, S. André, S. Fiedler, M. Schnölzer, H.-J. Gabius, Human chimera-type galectin-3: defining the critical tail length for high-affinity glycoprotein/cell surface binding and functional competition with galectin-1 in neuroblastoma cell growth regulation, Biochimie 104 (2014) 90–99.
[24] S. Vértesy, M. Michalak, M.C. Miller, M. Schnölzer, S. André, J. Kopitz, K.H. Mayo, H.-J. Gabius, Structural significance of galectin design: impairment of homodimer stability by linker insertion and partial reversion by ligand presence, Protein Eng. Des. Sel. 28 (2015) 199–210.
[25] A.V. Semenyuk, D.I. Svergun, Gnom: a program package for small-angle scattering data-processing, J. Appl. Crystallogr. 24 (1991) 537–540.
[26] P.V. Konarev, M.V. Petoukhov, V.V. Volkov, D.I. Svergun, ATSAS 2.1, a program package for small-angle scattering data analysis, J. Appl. Crystallogr. 39 (2006) 277–286.
[27] D.I. Svergun, Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing, Biophys. J. 76 (1999) 2879–2886.
[28] V.V. Volkov, D.I. Svergun, Uniqueness of ab initio shape determination in small-angle scattering, J. Appl. Crystallogr. 36 (2003) 860–864.
[29] P. Schuck, H. Zhao, C.A. Bräutigam, R. Ghirlanda, Basic Principles of Analytical Ultracentrifugation, CRC Press, Boca Raton, USA, 2015.
[30] M. Amano, H. Eriksson, J.C. Manning, K.M. Detjen, S. André, S.-I. Nishimura, J. Lehtiö, H.-J. Gabius, Tumour suppressor $p16^{INK4a}$: anoikis-favouring decrease in N/O-glycan/cell surface sialylation by down-regulation of enzymes in sialic acid biosynthesis in tandem in a pancreatic carcinoma model, FEBS J. 279 (2012) 4062–4080.
[31] O. Blixt, S. Head, T. Mondala, C. Scanlan, M.E. Huflejt, R. Alvarez, M.C. Bryan, F. Fazio, D. Calarese, J. Stevens, N. Razi, D.J. Stevens, J.J. Skehel, I. van Die, D.R. Burton, I.A. Wilson, R. Cummings, N. Bovin, C.H. Wong, J.C. Paulson, Printed covalent glycan array for ligand profiling of diverse glycan binding proteins, Proc. Natl. Acad. Sci. U. S. A. 101 (2004) 17033–17038.
[32] K. Diederichs, P.A. Karplus, Improved R-factors for diffraction data analysis in macromolecular crystallography, Nat. Struct. Biol. 4 (1997) 269–275.
[33] W. Kabsch, Integration, scaling, space-group assignment and post-refinement, Acta Crystallogr. D66 (2010) 133–144.
[34] P.R. Evans, An introduction to data reduction: space-group determination, scaling and intensity statistics, Acta Crystallogr. D67 (2011) 282–292.
[35] P.R. Evans, G.N. Murshudov, How good are my data and what is the resolution? Acta Crystallogr. D69 (2013) 1204–1214.
[36] B.W. Matthews, Solvent content of protein crystals, J. Mol. Biol. 33 (1968) 491–497.
[37] A.J. Mccoy, R.W. Grosse-Kunstleve, P.D. Adams, M.D. Winn, L.C. Storoni, R.J. Read, Phaser crystallographic software, J. Appl. Crystallogr. 40 (2007) 658–674.
[38] G.N. Murshudov, A.A. Vagin, E.J. Dodson, Refinement of macromolecular structures by the maximum-likelihood method, Acta Crystallogr. D53 (1997) 240–255.
[39] P. Emsley, B. Lohkamp, W.G. Scott, K. Cowtan, Features and development of Coot, Acta Crystallogr. D66 (2010) 486–501.
[40] V.B. Chen, W.B. Arendall, J.J. Headd, D.A. Keedy, R.M. Immormino, G.J. Kapral, L.W. Murray, J.S. Richardson, D.C. Richardson, MolProbity: all-atom structure validation for macromolecular crystallography, Acta Crystallogr. D66 (2010) 12–21.
[41] D. Houzelstein, I.R. Gonçalves, A.J. Fadden, S.S. Sidhu, D.N.W. Cooper, K. Drickamer, H. Leffler, F. Poirier, Phylogenetic analysis of the vertebrate galectin family, Mol. Biol. Evol. 21 (2004) 1177–1187.
[42] K. Hotta, T. Funahashi, Y. Matsukawa, M. Takahashi, H. Nishizawa, K. Kishida, M. Matsuda, H. Kuriyama, S. Kihara, T. Nakamura, Y. Tochino, N.L. Bodkin, B.C. Hansen, Y. Matsuzawa, Galectin-12, an adipose-expressed galectin-like molecule possessing apoptosis-inducing activity, J. Biol. Chem. 276 (2001) 34089–34097.
[43] R.Y. Yang, D.K. Hsu, L. Yu, J. Ni, F.T. Liu, Cell cycle regulation by galectin-12, a new member of the galectin superfamily, J. Biol. Chem. 276 (2001) 20252–20260.
[44] E.-M. Katzenmaier, S. André, J. Kopitz, H.-J. Gabius, Impact of sodium butyrate on the network of adhesion/growth-regulatory galectins in human colon cancer *in vitro*, Anticancer Res. 34 (2014) 5429–5438.
[45] M.A. Gitt, M.F. Wiser, H. Leffler, J. Herrmann, Y. Xia, S.M. Massa, D.N.W. Cooper, A.J. Lusis, S.H. Barondes, Sequence and mapping of galectin-5, a β-galactoside-binding lectin, found in rat erythrocytes, J. Biol. Chem. 270 (1995) 5032–5038.
[46] S. André, H. Kaltner, M. Lensch, R. Russwurm, H.-C. Siebert, C. Fallsehr, E. Tajkhorshid, A.J.R. Heck, M. von Knebel-Döberitz, H.-J. Gabius, J. Kopitz, Determination of structural and functional overlap/divergence of five proto-type galectins by analysis of the growth-regulatory interaction with ganglioside GM1 *in silico* and *in vitro* on human neuroblastoma cells, Int. J. Cancer 114 (2005) 46–57.
[47] J. Wada, Y.S. Kanwar, Identification and characterization of galectin-9, a novel β-galactoside-binding mammalian lectin, J. Biol. Chem. 272 (1997) 6078–6086.
[48] C. Barrès, L. Blanc, P. Bette-Bobillo, S. André, R. Mamoun, H.-J. Gabius, M. Vidal, Galectin-5 is bound onto the surface of rat reticulocyte exosomes and modulates vesicle uptake by macrophages, Blood 115 (2010) 696–705.
[49] A.M. Wu, T. Singh, J.H. Wu, M. Lensch, S. André, H.-J. Gabius, Interaction profile of galectin-5 with free saccharides and mammalian glycoproteins: probing its fine-specificity and the effect of naturally clustered ligand presentation, Glycobiology 16 (2006) 524–537.
[50] M. Watanabe, O. Nakamura, K. Muramoto, T. Ogawa, Allosteric regulation of the carbohydrate-binding ability of a novel conger eel galectin by D-mannoside, J. Biol. Chem. 287 (2012) 31061–31072.
[51] M.A. Wälti, P.J. Walser, S. Thore, A. Grunler, M. Bednar, M. Kunzler, M. Aebi, Structural basis for chitotetraose coordination by CGL3, a novel galectin-related protein from *Coprinopsis cinerea*, J. Mol. Biol. 379 (2008) 146–159.
[52] Y.A. Knirel, H.-J. Gabius, O. Blixt, E.M. Rapoport, N.R. Khasbiullina, N.V. Shilova, N.V. Bovin, Human tandem-repeat-type galectins bind bacterial non-βGal polysaccharides, Glycoconj. J. 31 (2014) 7–12.
[53] S.K. Patnaik, P. Stanley, Lectin-resistant CHO glycosylation mutants, Methods Enzymol. 416 (2006) 159–182.
[54] P. van der Sluijs, H.P. Bootsma, B. Postema, F. Moolenaar, D.K.F. Meijer, Drug targeting to the liver with lactosylated albumins: does the glycoprotein target the drug or is the drug targeting the glycoprotein? Hepatology 6 (1986) 723–728.

[55] H.-J. Gabius, M. van de Wouwer, S. André, A. Villalobo, Down-regulation of the epidermal growth factor receptor by altering N-glycosylation: emerging role of β1,4-galactosyltransferases, Anticancer Res. 32 (2012) 1565–1572.
[56] H. Sanchez-Ruderisch, C. Fischer, K.M. Detjen, M. Welzel, A. Wimmel, J.C. Manning, S. André, H.-J. Gabius, Tumor suppressor p16^{INK4a}: downregulation of galectin-3, an endogenous competitor of the pro-anoikis effector galectin-1, in a pancreatic carcinoma model, FEBS J. 277 (2010) 3552–3563.
[57] D. Zhou, H. Ge, J. Sun, Y. Gao, M. Teng, L. Niu, Crystal structure of the C-terminal conserved domain of human GRP, a galectin-related protein, reveals a function mode different from those of galectins, Proteins 71 (2008) 1582–1588.
[58] M.A. Wälti, S. Thore, M. Aebi, M. Kunzler, Crystal structure of the putative carbohydrate recognition domain of human galectin-related protein, Proteins 72 (2008) 804–808.
[59] A. Flores-Ibarra, F.M. Ruiz, S. Vértesy, S. André, H.-J. Gabius, A. Romero, Preliminary X-ray crystallographic analysis of an engineered variant of human chimera-type galectin-3 with a shortened N-terminal domain, Acta Crystallogr. F71 (2015) 184–188.
[60] D. Zhou, J. Sun, W. Zhao, X. Zhang, Y. Shi, M. Teng, L. Niu, Y. Dong, P. Liu, Expression, purification, crystallization and preliminary X-ray characterization of the GRP carbohydrate-recognition domain from *Homo sapiens*, Acta Crystallogr. F62 (2006) 474–476.
[61] L. He, S. André, H.-C. Siebert, H. Helmholz, B. Niemeyer, H.-J. Gabius, Detection of ligand- and solvent-induced shape alterations of cell-growth-regulatory human lectin galectin-1 in solution by small angle neutron and X-ray scattering, Biophys. J. 85 (2003) 511–524.
[62] F.M. Ruiz, U. Gilles, I. Lindner, S. André, A. Romero, D. Reusch, H.-J. Gabius, Combining crystallography and hydrogen-deuterium exchange to study galectin-ligand complexes, Chem. Eur. J. 21 (2015) 13558–13568.
[63] F.M. Ruiz, I.S. Fernández, L. López-Merino, L. Lagartera, H. Kaltner, M. Menéndez, S. André, D. Solís, H.-J. Gabius, A. Romero, Fine-tuning of prototype chicken galectins: structure of CG-2 and structure–activity correlations, Acta Crystallogr. D69 (2013) 1665–1676.
[64] P.F. Varela, D. Solís, T. Díaz-Mauriño, H. Kaltner, H.-J. Gabius, A. Romero, The 2.15 Å crystal structure of CG-16, the developmentally regulated homodimeric chicken galectin, J. Mol. Biol. 294 (1999) 537–549.
[65] M.F. López-Lucendo, D. Solís, J.L. Sáiz, H. Kaltner, R. Russwurm, S. André, H.-J. Gabius, A. Romero, Homodimeric chicken galectin CG-1B (C-14): crystal structure and detection of unique redox-dependent shape changes involving inter- and intrasubunit disulfide bridges by gel filtration, ultracentrifugation, site-directed mutagenesis, and peptide mass fingerprinting, J. Mol. Biol. 386 (2009) 366–378.
[66] G. Peng, D. Sun, K.R. Rajashankar, Z. Qian, K.V. Holmes, F. Li, Crystal structure of mouse coronavirus receptor-binding domain complexed with its murine receptor, Proc. Natl. Acad. Sci. U. S. A. 108 (2011) 10696–10701.
[67] G. Peng, L. Xu, Y.L. Lin, L. Chen, J.R. Pasquarella, K.V. Holmes, F. Li, Crystal structure of bovine coronavirus spike protein lectin domain, J. Biol. Chem. 287 (2012) 41931–41938.
[68] B. Rotblat, H. Niv, S. André, H. Kaltner, H.-J. Gabius, Y. Kloog, Galectin-1(L11A) predicted from a computed galectin-1 farnesyl-binding pocket selectively inhibits Ras-GTP, Cancer Res. 64 (2004) 3112–3118.
[69] J.L. Wang, R.M. Gray, K.C. Haudek, R.J. Patterson, Nucleocytoplasmic lectins, Biochim. Biophys. Acta 1673 (2004) 75–93.
[70] K.C. Haudek, K.J. Spronk, P.G. Voss, R.J. Patterson, J.L. Wang, E.J. Arnoys, Dynamics of galectin-3 in the nucleus and cytoplasm, Biochim. Biophys. Acta 1800 (2010) 181–189.
[71] C. Villeneuve, L. Baricault, L. Canelle, N. Barboule, C. Racca, B. Monsarrat, T. Magnaldo, F. Larminat, Mitochondrial proteomic approach reveals galectin-7 as a novel bcl-2 binding protein in human cells, Mol. Biol. Cell 22 (2011) 999–1013.
[72] Y. Harazono, D.H. Kho, V. Balan, K. Nakajima, T. Zhang, V. Hogan, A. Raz, Galectin-3 leads to attenuation of apoptosis through Bax heterodimerization in human thyroid carcinoma cells, Oncotarget 5 (2014) 9992–10001.
[73] S. Saouros, B. Edwards-Jones, M. Reiss, K. Sawmynaden, E. Cota, P. Simpson, T.J. Dowse, U. Jakle, S. Ramboarina, T. Shivarattan, S. Matthews, D. Soldati-Favre, A novel galectin-like domain from *Toxoplasma gondii* micronemal protein 1 assists the folding, assembly, and transport of a cell adhesion complex, J. Biol. Chem. 280 (2005) 38583–38591.
[74] M.H. Huynh, B. Liu, M. Henry, L. Liew, S.J. Matthews, V.B. Carruthers, Structural basis of *Toxoplasma gondii* MIC2-associated protein interaction with MIC2, J. Biol. Chem. 290 (2015) 1432–1441.
[75] B.W. Kim, S.B. Hong, J.H. Kim, H. do Kwon, H.K. Song, Structural basis for recognition of autophagic receptor NDP52 by the sugar receptor galectin-8, Nat. Commun. 4 (2013) 1613.
[76] S. Li, M.P. Wandel, F.D. Li, Z.H. Liu, C. He, J.H. Wu, Y.Y. Shi, F. Randow, Sterical hindrance promotes selectivity of the autophagy cargo receptor NDP52 for the danger receptor galectin-8 in antibacterial autophagy, Sci. Signal. 6 (2013).
[77] Z. Pál, P. Antal, S.K. Srivastava, G. Hullám, A.F. Semsei, J. Gál, M. Svébis, G. Soós, C. Szalai, S. André, E. Gordeeva, G. Nagy, H. Kaltner, N.V. Bovin, M.J. Molnár, A. Falus, H.-J. Gabius, E.I. Buzás, Non-synonymous single nucleotide polymorphisms in genes for immunoregulatory galectins: association of galectin-8 (F19Y) occurrence with autoimmune diseases in a Caucasian population, Biochim. Biophys. Acta 1820 (2012) 1512–1518.
[78] F.M. Ruiz, B.A. Scholz, E. Buzamet, J. Kopitz, S. André, M. Menéndez, A. Romero, D. Solís, H.-J. Gabius, Natural single amino acid polymorphism (F19Y) in human galectin-8: detection of structural alterations and increased growth-regulatory activity on tumor cells, FEBS J. 281 (2014) 1446–1464.
[79] S. Zhang, R.-O. Moussodia, S. Vértesy, S. André, M.L. Klein, H.-J. Gabius, V. Percec, Unraveling functional significance of natural variations of a human galectin by glycodendrimersomes with programmable glycan surface, Proc. Natl. Acad. Sci. U. S. A. 112 (2015) 5585–5590.
[80] H. Dawson, S. André, E. Karamitopoulou, I. Zlobec, H.-J. Gabius, The growing galectin network in colon cancer and clinical relevance of cytoplasmic galectin-3 reactivity, Anticancer Res. 33 (2013) 3053–3059.
[81] S. Toegel, D. Bieder, S. André, K. Kayser, S.M. Walzer, G. Hobusch, R. Windhager, H.-J. Gabius, Human osteoarthritic knee cartilage: fingerprinting of adhesion/growth-regulatory galectins *in vitro* and *in situ* indicates differential upregulation in severe degeneration, Histochem. Cell Biol. 142 (2014) 373–388.
[82] J.N. Gready, A.N. Zelensky, Routes in lectin evolution: case study on the C-type lectin-like domains, in: H.-J. Gabius (Ed.), The Sugar Code. Fundamentals of glycosciences, Wiley-VCH, Weinheim, Germany 2009, pp. 329–346.
[83] H. Kaltner, G. García Caballero, F. Sinowatz, S. Schmidt, J.C. Manning, S. André, H.-J. Gabius, Galectin-related protein: an integral member of the network of chicken galectins 2, From expression profiling to its immunocyto- and histochemical localization and application as tool for ligand detection. Biochim. Biophys. Acta. 1860 (2016) 2285–2297.
[84] P.A. Karplus, K. Diederichs, Linking crystallographic model and data quality, Science 336 (2012) 1030–1033.