Chicken (*Gallus gallus*) as Model for Network Analysis of Adhesion-/Growth-regulatory Galectins

Biochemical characterization of C-GRIFIN/C-GRP and first complete histochemical analysis for the galectin family in bursa of Fabricius

von Gabriel García Caballero

Inaugural-Dissertation
zur Erlangung der Doktorwürde
(Dr. rer. biol. vet.)
der Tierärztlichen Fakultät der
Ludwig-Maximilians-Universität
München

Chicken (*Gallus gallus*) as Model for Network Analysis of Adhesion-/Growth-regulatory Galectins

Biochemical characterization of C-GRIFIN/C-GRP and first complete histochemical analysis for the galectin family in bursa of Fabricius

von Gabriel García Caballero
aus Zaragoza, Spanien

München 2018

Aus dem Veterinärwissenschaftlichen Department der
Tierärztlichen Fakultät der Ludwig-Maximilians-Universität München

Lehrstuhl für Physiologische Chemie

Arbeit angefertigt unter der Leitung von: Prof. Dr. Herbert Kaltner

Gedruckt mit der Genehmigung der Tierärztlichen Fakultät der Ludwig-Maximilians-Universität München

Dekan:	Univ.-Prof. Dr. Reinhard K. Straubinger, Ph.D
Berichterstatter:	Prof. Dr. Herbert Kaltner
Korreferent/en:	Univ.-Prof. Dr. Hermann Ammer Univ.-Prof. Dr. Dr. Michael H. Erhard Univ.-Prof. Dr. Dr. habil. Joris Peters Priv.-Doz. Dr. Monika Rinder

Tag der Promotion:10. Februar 2018

Die vorliegende Arbeit wurde gemäß §6 Abs. 2 der Promotionsordnung für die Tierärztliche Fakultät der Ludwig-Maximilians-Universität München in kumulativer Form verfasst.

Folgende wissenschaftliche Arbeiten sind in dieser Dissertationsschrift enthalten:

García Caballero, G., Kaltner, H., Michalak, M., Shilova, N., Yegres, M., André, S., Ludwig, A.K., Manning, J.C., Schmidt, S., Schnölzer, M., Bovin, N.V., Reusch, D., Kopitz, J., and Gabius, H.-J. (2016) Chicken GRIFIN: A homodimeric member of the galectin network with canonical properties and a unique expression profile. *Biochimie* **128-129,** 34-47

García Caballero, G., Flores-Ibarra, A., Michalak, M., Khasbiullina, N., Bovin, N.V., André, S., Manning, J.C., Vértesy, S., Ruiz, F.M., Kaltner, H., Kopitz, J., Romero, A., and Gabius, H.-J. (2016) 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. *Biochim Biophys Acta* **1860**, 2285-2297

Kaltner, H., García Caballero, G., Sinowatz, F., Schmidt, S., Manning, J.C., André, S., and Gabius, H.-J (2016) 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**, 2298-2312

Weitere Veröffentlichungen, die nicht in die Dissertationsschrift eingehen:

Manning, J.C., Romero, A., Habermann, F.A., García Caballero, G., Kaltner, H., and Gabius, H.-J. (2017) Lectins: a primer for histochemists and cell biologists. *Histochem Cell Biol* **147**, 199-222

Manning, J.C., García Caballero, G., Knospe, C., Kaltner, H., and Gabius, H.-J. (2017) Network analysis of adhesion/growth-regulatory galectins and their binding sites in adult chicken retina and choroid. *J Anat* **231**, 23-37

Kaltner, H., Toegel, S., García Caballero, G., Manning, J.C., Ledeen, R.W., and Gabius, H.-J. (2017) Galectins: their network and roles in immunity/tumor growth control. *Histochem Cell Biol* **147**, 239-256

The permissions for reuse of the full articles in this thesis was provided by the publishers.

The position as an ‘early stage researcher’ was generously funded by the Marie Curie Initial Training Network GLYCOPHARM (PITN-GA-2012-317297).

For my parents, with love,
your son

TABLE OF CONTENTS

ABBREVIATIONS

AGE	advanced glycation end-product
Alix	ALG-2-interacting protein X
axin	axis inhibition protein
Arg	arginine
Asn	asparagine
ATP	adenosine triphosphate
Bam32	B lymphocyte adaptor molecule of 32 kDa
bax	bcl-2-associated X protein
bcl-2	B cell lymphoma 2
CD	cluster of differentiation
cDNA	complementary deoxyribonucleic acid
CEA	carcinoembryonic antigen
CG	chicken galectin
CRD	carbohydrate recognition domain
DMBT-1	Deleted in malignant brain tumors 1 protein
DNA	deoxyribonucleic acid
EST	expressed sequence Tag
Gal	galactose
GalNAc	N-acetylgalactosamine
Gemin4	Gem-associated protein 4
Glc	glucose
Glu	glutamate
Gly	glycine
GRIFIN	galectin-related interfiber protein
GRP	galectin-related protein
H	hydrogen
His	histidine
hnRNP	heterogeneous ribonucleoprotein particle
hsp90	heat shock protein 90
Ig	Immunoglobulin
kDa	kilodalton
LacNAc	N-acetyllactosamine
lamp-1	lysosomal-associated membrane protein 1
Lys	lysine
Mac	macrophage antigen
mRNA	messenger ribonucleic acid
mSufu	mouse suppressor of fused
MAG	myelin-associated glycoprotein
Mdm2	mouse double minute 2 homolog
MUC	mucin
Neu5Ac	N-acetylneuraminic acid
Nup	nucleoporin
NG2	neuron-glial antigen 2

OCA-B	Oct co-activator from B cells
PBMC	peripheral blood mononuclear cell
pCIP	p300/CBP-cointegrator protein
PDB	protein data bank
PIAS1	protein inhibitor of activated STAT1
PP13	placental protein 13
Pro	proline
RNA	ribonucleic acid
RPTP	receptor protein-tyrosine phosphatase
RT-PCR	reverse transcription polymerase chain reaction
Ser	serine
TCR	T-cell receptor
TgMIC4	*Toxoplasma gondi* microneme protein 4
Thy-1	thymocyte differentiation antigen 1
Trp	tryptophan
TTF-1	thyroid transcription factor-1
Val	valine

1. INTRODUCTION

"To an observer trying to obtain a bird's eye view of the present state of biochemistry, life may until very recently have seemed to depend on only two classes of compounds: nucleic acids and proteins" [1]. This classical statement by Prof. Dr. N. Sharon serves to introduce the necessity of considering complex carbohydrates as intrinsic part of the dogma of molecular biology in relation to the flow of genomic information (Fig. 1).

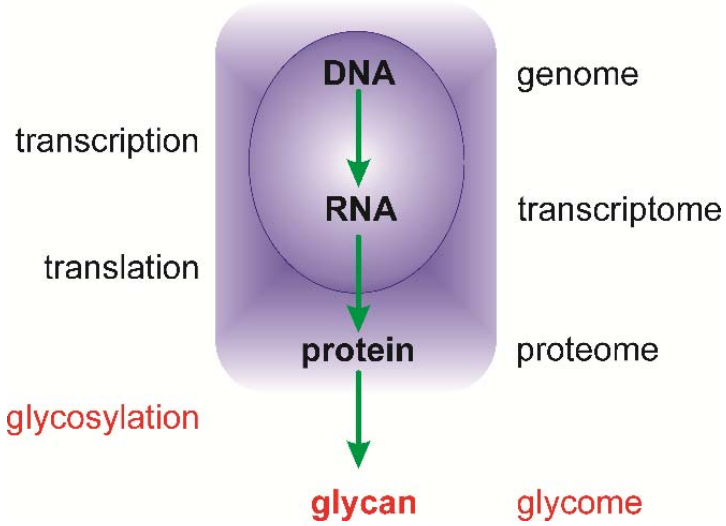


Fig. 1: Schematic illustration of the change in our view on the flow of genetic information. Modified from [2], with permission.

Intuitively, Saul Roseman, a pioneer of glycobiology, reasoned that "these complex structures must serve essential roles in cell surface phenomena, but we are only beginning to understand what some of these functions are" [3]. As illustrated in [4] and [5], sugars (the third alphabet of life) are truly "ideal for generating compact units with explicit informational properties" [6]. A key route to translate the sugar-encoded information into cellular responses is via sugar receptors (lectins). Examples for emerging physiological importance of lectin-glycan recognition are shown in Fig. 2.

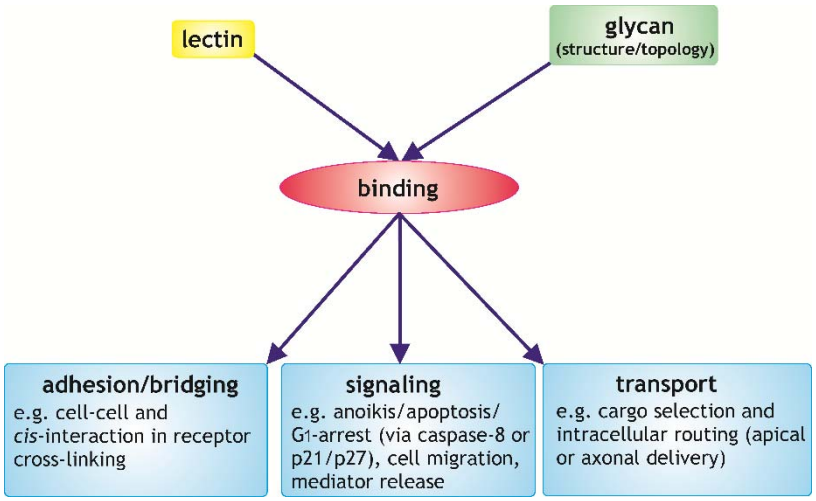


Fig. 2: Schematic illustration of examples of functional consequences of lectin-glycan recognition. From [2], with permission.

Obviously, research efforts on glycans and lectins are complementary within the quest to crack the sugar code. In order to address this challenge, by working on the protein side, a focus on a particular lectin group will be mandatory. The respective work should reach the status of thorough structural analysis and functional assessment. At best, the group size should be of manageable complexity, and the lectins should have physiological importance.

These prerequisites are fulfilled by the family of galectins, potent effectors of various cell activities such as adhesion, anoikis/apoptosis, cytokine induction and release or cell migration [2,7]. Of particular importance, the *c*arbohydrate *r*ecognition *d*omain (CRD), whose characteristics are shared by all members of this family, is presented in three phylogenetically conserved forms of topological display (Fig. 3).

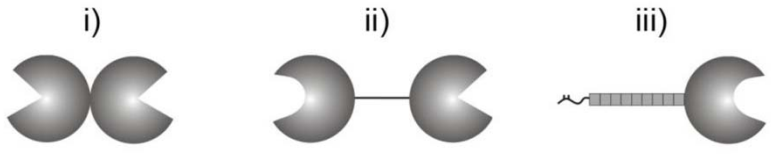


Fig. 3: The three types of natural design of galectins.

In principle, the galectin CRD can form i) non-covalently associated homodimers (proto-type) or ii) linker-peptide-connected heterodimers (tandem-repeat-type) and iii) occur

associated to an N-terminal stalk consisting of a peptide with two sites for serine phosphorylation and non-triple-helical (collagen-like) repeats [7]. These three types of CRD presentation are common to all organisms expressing galectins, with changes in numbers of proteins in each subgroup (Fig. 4).

The comparatively small chicken galectinome with five proteins enables comprehensive galectin biochemical and immunohistochemical studies with focus on monitoring the network of galectins in biorelevant processes as done in [8] and [9] (for a review on roles in immunity/tumor growth control please see [10]). The discovery of two new non-canonical members of the galectin family in this organism prompted us to pursue their characterization and, therewith, fill these gaps in the galectin network puzzle of *Gallus gallus*.

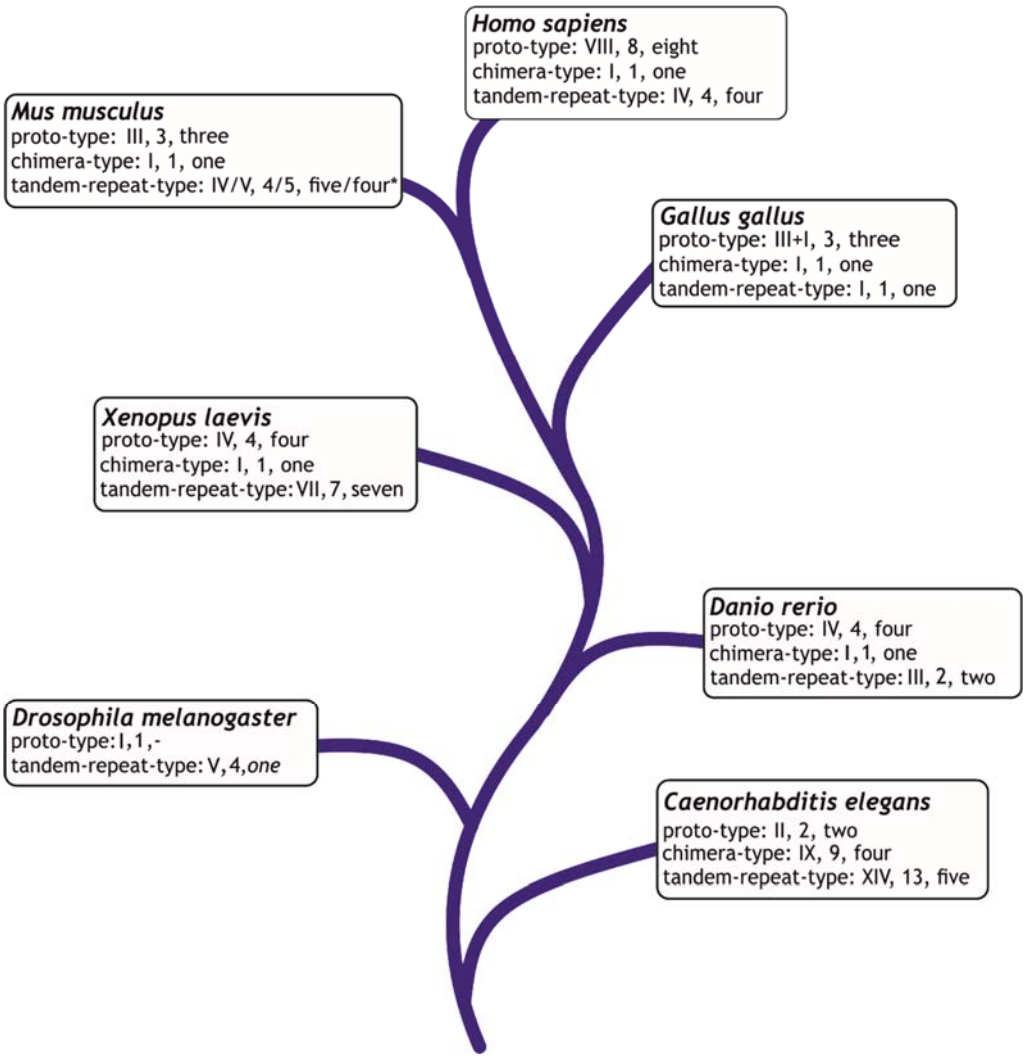


Fig. 4: Schematic representation of occurrence of galectins on the level of the gene (Roman number), of the mRNA (Arabic number) and of the protein (as word) in selected (model) organisms. From [2], modified and updated, with permission.

2. LITERATURE REVIEW

2.1. Galectins: structural requirements

The term lectin (from Latin *legere*, to choose) is rooted in the works with plant proteins that presented specificity for a certain blood-group [11]. It emerged in order to fulfill the necessity of finding a new word for substances which are proteins and bind a determined antigen but are not produced in response to this antigen [12]. Nowadays, the term lectin is the generic name for (glyco)proteins that recognize and bind to carbohydrates, different from sugar-specific antibodies and enzymes as well as from sensor/transport proteins for free mono- and oligosaccharides [11].

Structural analysis of these proteins by crystallography allowed characterization of their active site. The structure of the folds that are able to accommodate a ligand proved to be very rich and varied, with more than a dozen folds being seen in vertebrates [13,14]. Of note, a certain type of fold can present its contact site in different places, resulting in a wide range of proteins structurally related but functionally diverse, being able to bind different carbohydrates or even to recognize different presentations of the same molecule. The complex architecture of the CRD and its network of interactions with the ligand ensure specificity of lectins and enable them to act both as precise readers and translators of the sugar code [15].

Three different types of CRDs are exemplarily depicted in Fig. 5. The C-type fold (Fig. 5a) teams up with a Ca^{2+}-ion in order to guarantee the correct ligand specificity, adding a further degree of complexity to its interaction with the protein [16,17]. The β-sandwich motif shown in Fig. 5b is the common structural feature of galectins, a family of proteins with carbohydrate specificity to β-galactosides and their derivatives [18,19]. Here, several loops connect two antiparallel β-sheets of six and five β-strands, building a β-sandwich structure [20].

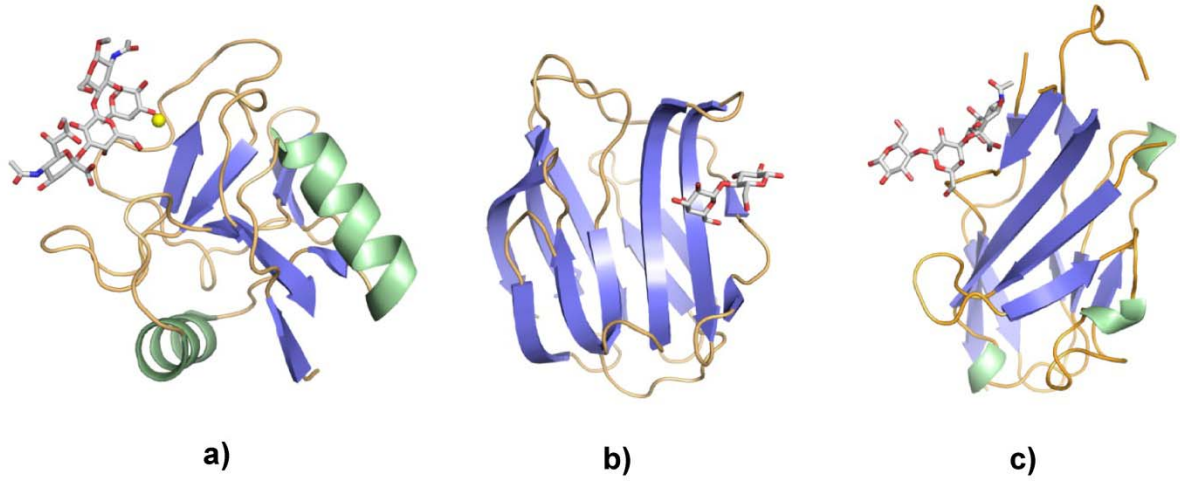


Fig. 5: (a) C-type CRD of human E-selectin with its Ca^{2+} (yellow sphere) in complex with the tetrasaccharide sialyl Lewis x (CD15s) (PDB code 1G1T), **(b)** β-sandwich of human galectin-1 in complex with lactose (PDB code 1GZW), **(c)** the I(gG)-type CRD of murine siglec sialoadhesin (siglec-1) in complex with the trisaccharide 3'-sialyllactose (PDB code 1QFO). From [15], modified, with permission.

This group of lectins has a conserved sequence signature in which a Trp residue plays a crucial role in asserting the correct binding of the β-galactoside ligand (please see Fig. 6 for typical binding partners) by allowing C-H/π interactions [16,18,19,21]. The amino acid residues that constitute this signature are shown in the structural environment of the CRD of three homodimeric chicken galectins (CGs) (Fig. 7a), as well as marked in the alignment of the CRD's sequences of the whole set of five canonical CGs in Fig. 7b. Hereby, we have just described the three defining characteristics of the galectin family: specificity to β-galactosides, β-sandwich fold structure, and amino acid sequence containing the typical conserved signature.

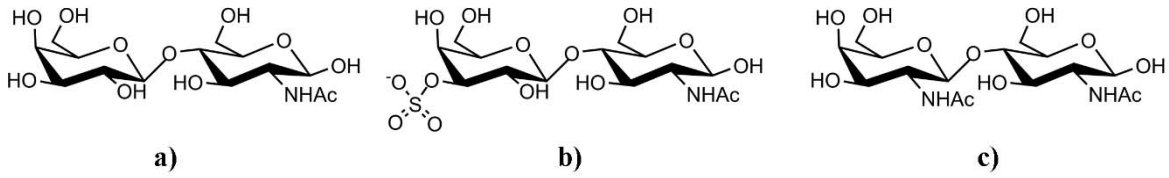


Fig. 6: Illustration of three β-galactoside ligands. a) LacNAc (Galβ1,4GlcNAc), b) 3'-O-sulfated LacNAc (3'-O-sulfo-Galβ1,4GlcNAc)[1] and c) LacdiNAc (GalNAcβ1,4GlcNAc)[2].

[1] Allen, H.J. et al., *Binding of synthetic sulfated ligands by human splenic galectin 1, a beta-galactoside-binding lectin.* Glycoconj J, 1998. **15**(7): p. 691-5

[2] van den Berg, T.K., et al., *LacdiNAc-glycans constitute a parasite pattern for galectin-3-mediated immune recognition.* J Immunol, 2004. **173**(3): p. 1902-7

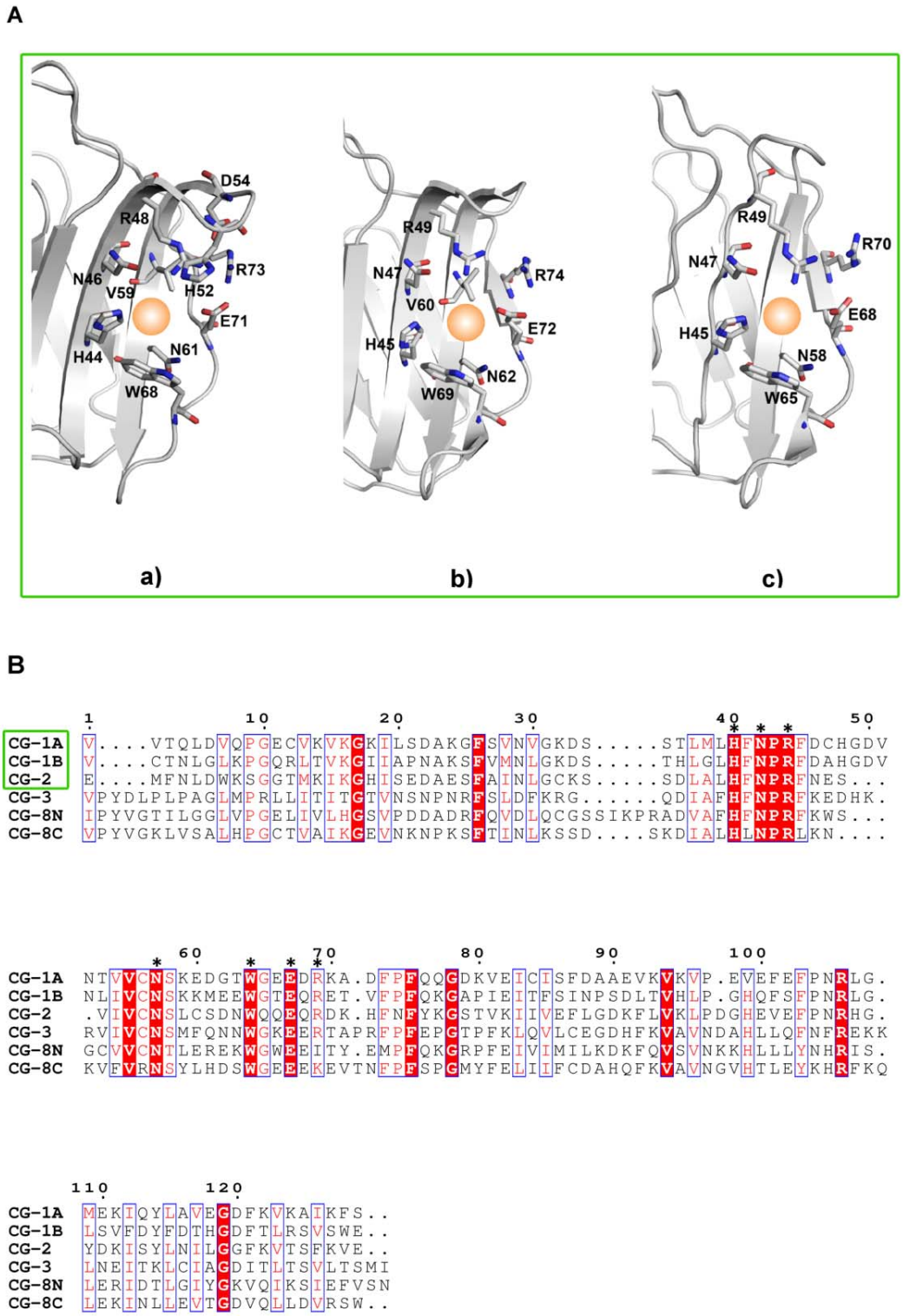


Fig. 7: Comparison of the carbohydrate-binding sites **(A)** of the proto-type chicken galectins CG-1A (a, PDB code 1QMJ), CG-1B (b, PDB code 3DUI) and CG-2 (c, PDB code 2JMZ). Alignment of the section of the amino acid sequences of the CRDs of the five canonical chicken galectins (amino acid 6 of CG-1A is set to position 1 for alignment) **(B)**. Strictly conserved (red background) and homologous residues (>70% conservation; boxed red letters) are highlighted by coloring. The amino acids in contact with the canonical ligand lactose are marked by an asterisk. From [15] extended and modified, with permission.

When taking a closer look into the CRD structure, it is noted that “certain amino acids show a strong propensity to be in the binding site. This applies in particular to the aromatic rings, which can pack against the hydrophobic face of the sugar, and arginine, aspartate and glutamate residues, which can all form bidentate interactions, with adjacent hydroxyls on the sugar” [22]. This assessment, referring to lectins in general, holds true for galectins as depicted exemplarily in Fig. 8 for the case of chicken galectin 2 (CG-2). The lactose molecule is packed inside a cavity in the CRD of the protein, the amino acids present there create a network of H-bonds and CH/π interactions (please note the aromatic chain of amino acid Trp65 facing the hexose ring of the galactose residue) that assures specificity and constitutes the base of the molecular interaction mechanism between the protein and its ligand [16].

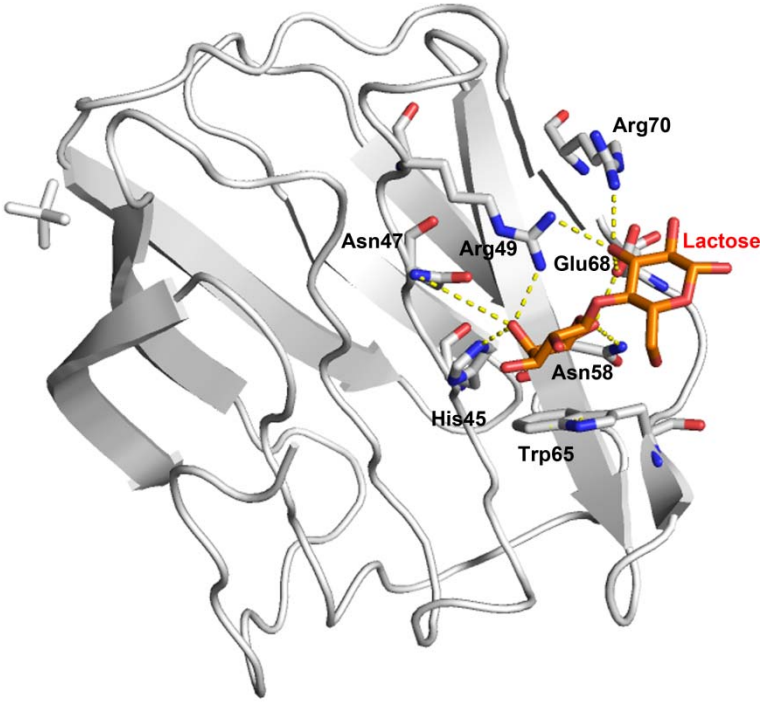


Fig. 8: Close up view of the CRD region of CG-2 accommodating the canonical ligand lactose. The H-bond network interactions between the amino acid residues and the hydroxyl groups of the lactose molecule are shown as yellow dashed points. The 4´-OH group of galactose interacts with the side chains of the key amino acids His45, Asn47 and Arg49, whereas Asn58 and Glu68 are bonded to the 6´-OH group. The hydroxyl group at the 3´position of the glucose ring forms hydrogen bonds with the Arg49, Glu68 and Arg70 residues [20]. PDB code 2YMZ.

In order to successfully complete their varied tasks, galectins must not only possess carefully designed CRDs, but also present them in a suitable way. Three different topological modes of CRD presentation have been described: the proto-type (mono- or homodimeric), the tandem-repeat-type (up to four different CRDs covalently connected by a linker) and the chimera-type, (with the CRD associated with a tail consisting on collagen-like repeats and a

peptide with two sites for Ser phosphorylation) [13,15,19,23,24]. The functional significance of these three ways of presenting the CRD is a fundamental issue, underlined by the observed fact that (ga)lectins with identical ligand specificity but different molecular architecture can initiate different processes or even neutralize the signal triggered by an homologous protein [16], as seen for the case of galectin-1 and galectin-3 (described later). Alternatively, cooperation between galectins is possible as well, as reported for galectins-1 and -8 while promoting plasma cell formation [25] and for galectins-1 and -3 in osteoarthritis pathogenesis [26]. Such occurrences highlight the necessity of studying not only the nature of galectins as individual entities but as part of an interconnected network within the different organisms [9]. In this sense, mapping of galectin presence in vertebrates, first by protein isolation and, more recently, by genomic database mining, has allowed to trace sequences of individual proteins belonging to the three galectin structural groups through the branches of the metazoa phylogenetic tree, disclosing diverse degrees of diversification for each family and number of detected genes.

2.2. Occurrence of galectins

The first member of the galectin family was discovered in 1975, by traditional haemagglutination assays and later purification from the electric organ of *Electrophorus electricus* (electric eel) [27]. Since then, widespread advances in mass sequencing of genomes and screening have surpassed and substituted the traditional galectin identification methods and subsequently expanded our scope over the presence of this protein class among eukaryotic organisms. As such, it is nowadays acknowledged that the galectin family is "evolutionarily ancient with representatives in vertebrates, invertebrates, and even protists" [19].

Although not a bona fide galectin, the fifth apple domain of the TgMIC4 microneme complex of *Toxoplasma gondii* (a parasitic protist) possess lectin activity [28], which can be considered a precursor of the galectin family in the higher eukaryotic kingdoms. An example of a fungus with galectin presence is the mushroom *Coprinopsis cinerea*. Two isolectins with the capacity of binding to lactose have been described [19,29], one of them capable of binding *N*-glycans of invertebrates as previous step to exerting its toxicity over this organisms [30].

However, it is among metazoans where lectins with affinity to β-galactosides are more ubiquitously found. Starting from the lowest clade of this kingdom, two galectin genes have been described for the sponge *Geodia cydonium* [31-33], although only one of the gene products proved to be functional [34]. This protein, termed LECT1, is able to oligomerize to di- and tetrameric complexes.

In nematodes, galectin genes seem to be highly represented. For instance, the genome of *Caenorhabditis elegans* contains in the order of 20,000 genes, of which 26 are candidate for encoding galectin specific information [19]. Twelve mRNAs have been recorded and binding to the canonical ligand lactose has been ascertained in seven cases (Lec-1, -2, -3, -4, -6, -10 and -12 (DC2.3)) [35-37]. *Haemonchus contortus*, one of the most pathogenic gastrointestinal nematodes of small ruminants, presents seven cDNAs associated to the tandem-repeat type galectins [38,39], and an isolated galectin (Hco-gal-m) has been seen to contribute to modulate the immune response of its host organism by binding peripheral blood mononuclear cells [40].

Within the phylum Mollusca, galectins have been identified in various species. The case of CvGal, a protein expressed by the Eastern Oyster (*Crassostrea virginica*), is special because of its CRD architecture, with four of them arranged in a tandem-repeat mode, a unique case among the galectin family [41]. The genome of *Drosophila melanogaster* (fruit fly) encodes five tandem-repeat-type genes and another one, which seems to be very distantly related to the proto-type group [19], and serves to illustrate presence of galectins in arthropods. The mRNAs for five of these sequences have been isolated but only one galectin, belonging to the tandem-repeat group, has been described on the protein level [42,43].

Turning to vertebrates, several galectin-like genes were detected in diverse fish species such as the already mentioned first isolated galectin from electric eel (*Ectrophorus electricus*) [27]. Further examples are found in the conger eel (*Conger myriaster*) [44], catfish (*Ictalurus punctatus*), flounder (*Paralichthys olivaceus*), trout (*Oncorhynchus mykiss*), medaka (*Oryzias latipes*), zebrafish (*Danio rerio*) and two pufferfish (*Takifugu rubripes* and *T. nigroviridis*) [19]. In the case of *D. rerio*, its genome includes eight candidate genes, having already been characterized three proto-type, one chimera-type and two tandem-repeat type galectins, all of them presenting notable similarities with mammalian galectins [45,46]. In amphibians like *Xenopus laevis* (frog) examples of the three groups of galectins are also found [47] as well as in birds like *Gallus gallus* [8].

As to mammals, galectins are commonly found through the branches of the phylogenetic tree. Taking mouse (*Mus musculus*) as an illustrative case, three proto-type proteins have been characterized, together with one chimera-type and four tandem-repeat-type proteins, here with the special occurrence of intra-species variation for galectin-6 [48-50]. *Rattus norvegicus* (rat) has also an exceptional difference, an apparently unique proto-type galectin-5 [51]. Names of animal galectins are given in accordance to their similarities to human galectins (for instance, the mouse galectin-4 is related to the homonymous human galectin, i.e. human galectin-4), presented for the three structural groups: eight proto-type, one chimera-type and four tandem-

repeat-type, all of them characterized both on the level of their genes and protein expression [19].

Having briefly summarized presence of galectins in the animal kingdom, establishment of a model organism to achieve comprehensive coverage of the galectin network is necessary. This organism should fulfill the following requirements: *i)* comparatively low degree of complexity of galectin presence and *ii)* close evolutionary relationship to mammals [52]. In this sense, the easily accessible *Gallus gallus* (chicken) bird meets these criteria by presenting a set of only five canonical CGs with members of the three (proto-, chimera- and tandem-repeat-) structural groups [8,52] as shown in Fig. 9: three proto-type CGs (CG-1A, CG-1B and CG-2) present as homodimers in solution [20,53-60], one chimera-type (CG-3) with four alternative splicing variants but same CRD [61-64] and one tandem-repeat-type (CG-8) with two possible linker lengths [8,65].

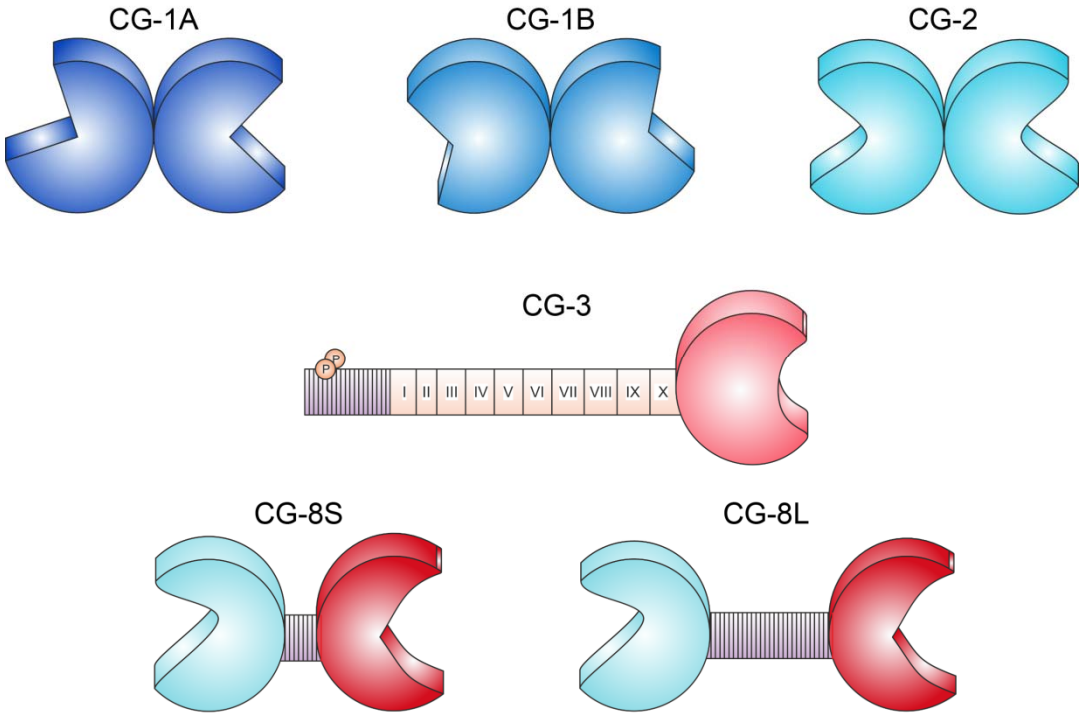


Fig. 9: Schematic illustration of modular organization of the five CGs: homodimeric (proto-type) CG-1A, CG-1B, CG-2; chimera-type CG-3 (with two serine phosphorylation sites in the N-terminal peptide and a collagenase-sensitive section of ten Gly/Pro-rich sequence repeats which consist of either five (one repeat), seven (five repeats) or eight (four repeats) amino acids) and tandem-repeat-type CG-8S/L (linker lengths of nine/28 amino acids). From [10], modified with permission.

2.3. Expression and localization of galectins

Already in the year 1985 it was hypothesized that "families of electrolectin genes could be under the control of different promoters" [66]. This would evidently lead to a diverse

expression profile of galectins [15]. Comparative mapping of genomic sequences disclosed a higher presence of deviations in the promoter regions [8,59,65], a fact that also points into this direction. However, early studies depicted expression of similar amounts of three different galectins during rat tissue maturation [67]. Furthermore, detection of specific mRNA by RT-PCR has taught us that, in general, a cell type does not express a single, but rather a characteristic pattern of co-expressed galectins [68], which prompted questions on identical or overlapping purposes of these proteins.

The possibility of a redundant role is quickly dismissed by the production of specific antibodies and their application in immunohistochemical fingerprinting [69]. Immunohistochemistry is a powerful technique for mapping this expression network, because it allows distinction between different cells types, subcellular localization and histological structures within a tissue. Of note, strict controls and successive purification steps are necessary in order to exclude any cross-reactivity between galectins, as they are homologous proteins with sequence and structural similarities [70-73]. Examples of selectivity of galectin presence are shown in Fig. 10 and Fig. 11. They can be found extracellularly, e.g. galectin-3 in epithelium of murine esophagus (Fig. 10a), possibly related to the well-studied role of galectins as cellular bridges and adhesion molecules [74]. Some galectins mainly are expressed in certain tissues, like galectin-4 in the gastrointestinal tract (Fig. 10c) [75]. Presence in cytoplasm and even in nuclei of very different types of cells is also a standard case encountered while performing immunohistochemical screening of galectin expression (Fig. 10a-f), which prompts the conclusions that a galectin is “variably expressed by different cells and tissues” and “its cellular localization is not restricted to the cell surface” [76]; being each signal presentation characteristic for a given protein.

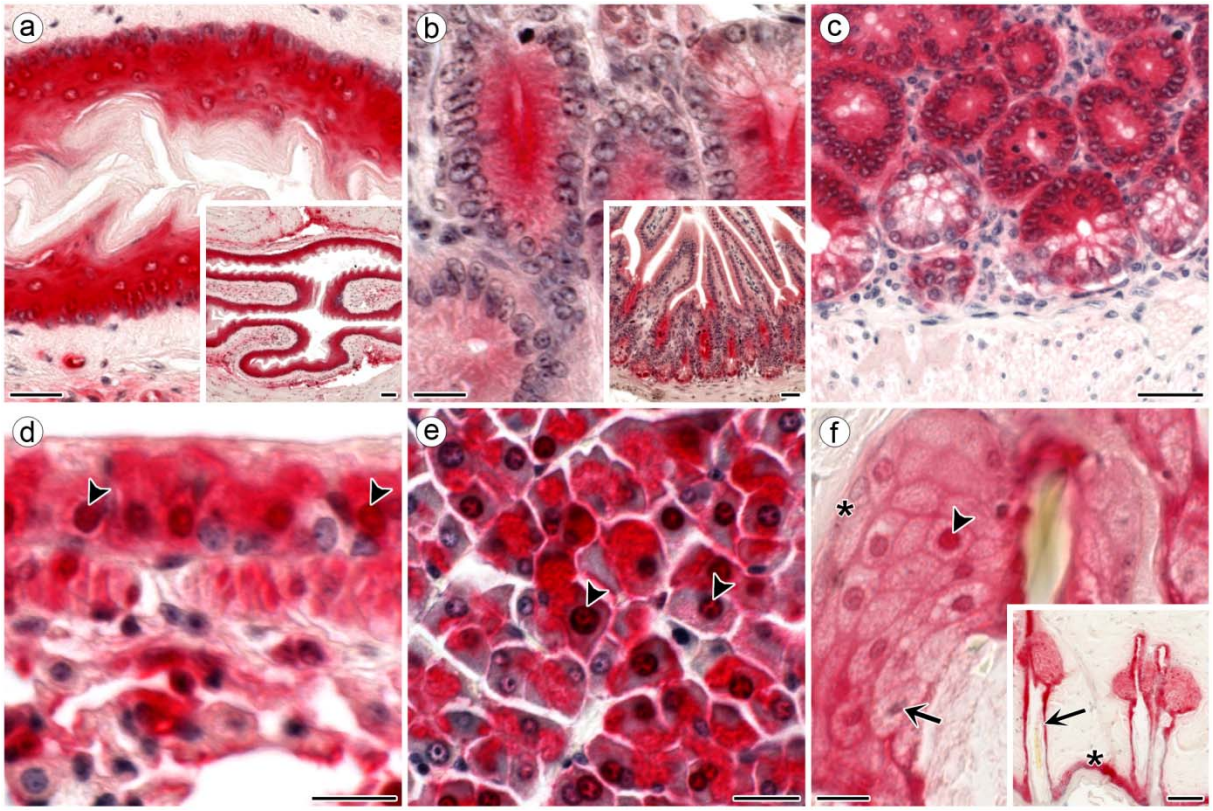


Fig. 10: Immunohistochemical staining profiles revealing subcellular and cell-type selectivity of galectin presence in sections of murine and chicken tissues. **a)** Extracellular presence of galectin-3 in keratinized, stratified epithelium of mouse esophagus (*inset*: section through the esophagus showing the folds of the mucosa lined by the galectin-3-positive epithelium). **b)** Strong apical positivity for galectin-7 in the cytoplasm of epithelial cells in glandulae intestinales of mouse ileum (*inset*: section through the ileum´s wall showing the position of galectin-7-positive glandulae intestinales relative to villi intestinales and muscle layers). **c)** Cytoplasmic staining of epithelial cells in glandulae intestinales in mouse jejunum by anti-galectin-4 IgG. **d)** Nuclear presence of galectin-3 in epithelial cells of a respiratory bronchiole in a section through mouse lung. **e)** Positivity with the antibody against galectin-4 in cell nuclei (*arrowheads*) and cytoplasm of pyramidal acinar cells in a section through the exocrine part of a mouse pancreas. **f)** Longitudinal section through a sebaceous gland in mouse skin. Positivity with anti-galectin-9 IgG was intense in nuclei (*arrowhead*) and weak in cytoplasm of maturing polygonal cells. Basal (*asterisk*) and degenerating cells (with pyknotic nuclei, *arrow*) were negative. The *inset*, showing sebaceous glands as appendages of hair follicles in the dermis of mouse skin, documents that the epidermis (*asterisk*) and epidermis-derived cells of the hair-follicle (*arrow*) were strongly positive. From [10], modified with permission.

As mentioned before, interplay of functionality makes the study of galectin expression as a network an obvious neccessity. Efforts in this direction have been made for different organisms like mouse [77,78] and also in the medical study of human tumors, reviewed in [79]. The production of specific antibodies against the whole set of chicken galectins has made their analysis in a series of anatomic systems possible, highlighting the distinct and specific nature of their co-expression profile (Fig. 11) [9,15,80].

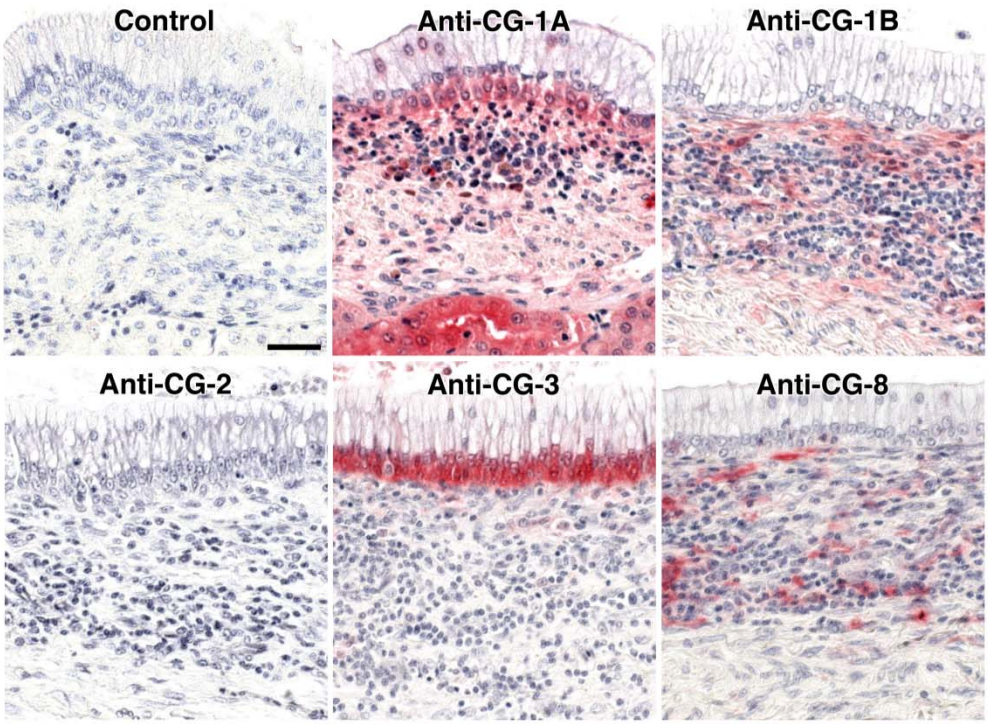


Fig. 11: Expression pattern of the five canonical chicken galectins in sections of fixed adult chicken ureter.CG-3 positivity was confined to the infranuclear portion of epithelial cells, CG-1B/CG-8 to connective tissue, while CG-1A was detected in all three layers. No expression of CG-2 was observed (bar = 25 µm). From [15], modified with permission.

2.4. Functional pairing with counterreceptors

Strategically positioned at the cell surface, glycans are readily accessible to endogenous lectins. For galectins, LacNAc repeats (with/without sulfation) or the pentasaccharide of GM1 (Fig.12) are common suitable binding partners. The interaction with these lectin receptors triggers a cell-type-specific panel of responses (please see Introduction, Fig. 2) relevant for diverse aspects of cellular physiology. It is already known that this class of lectins is capable of regulating cell adhesion growth [81,82] and apoptosis/anoikis [83,84].

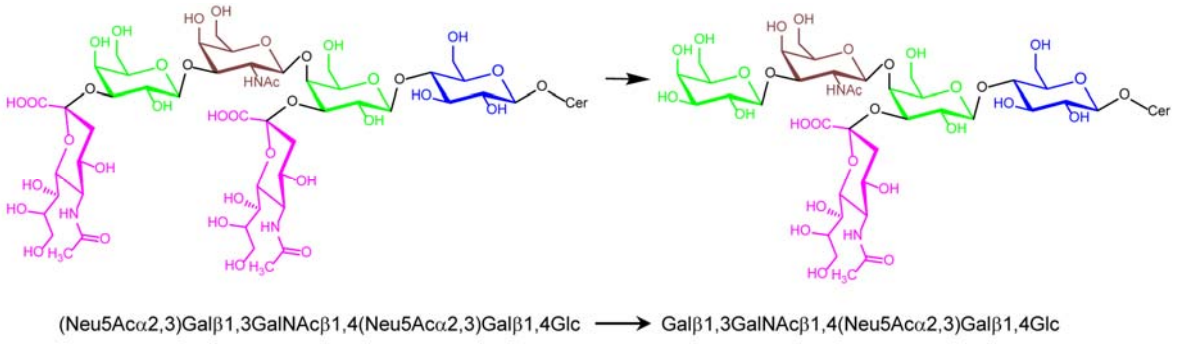


Fig. 12: Dynamic glycan remodeling by the (cell surface) ganglioside neuraminidase (Neu3). The hydrolytic removal of the branch-end sialic acid converts GD1a to GM1. From [98], with permission.

In carcinoma cells, distinct glycan epitopes (glycobiomarkers, please see Table 1 for exemplarly listing) occur, signaling the nature of the aberration on the cellular level and playing active roles in disease progression, by serving as docking points for galectins to induce anoikis/apoptosis or promote tumor invasion. For example, pro-apoptotic galectin-1 can exert its effect on prostate cancer cells only when core 2 β-1,6-N-acetylglucosaminyltransferase (C2GnT) is upregulated [85], resulting in the formation of core 2-branched O-glycans (Galβ1,3(GlcNacβ1,6)GalNAcα-Ser/Thr) which switch on galectin reactivity. A key factor for switching off galectin reactivity is α2,6-sialylation of N-glycans. Its regulation can work at the level of expression of the responsible enzyme (ST6Gal-I, β-galactoside-α2,6-sialyltransferase) and/or at the level of the limiting enzymatic step in sialic acid biosynthesis (i.e. epimerization of UDP-GlcNAc to UDP-ManNAc by the bifunctional UDP-GlcNAc 2-epimerase/ManNAc kinase, GNE). Similarly, Galectin-1 has been identified as a growth regulator as anoikis inductor in susceptible carcinoma cells by recognizing glycan signals and exerting a G_1-arrest, effect that can also be inhibited by N-sialylation [86]. It has also been described to be a central invasive factor in U87 glioblastoma cells [87,88]. This observation was supported by knock-down experiments [89]. Galectin-1 also regulates neuroblastoma cells growth by binding the GM1 ganglioside. The appearance of this molecule on the cell surface is the consequence of the increase of activity for a cell membrane ganglioside sialidase that shifts higher gangliosides toward ganglioside GM1 (Fig. 12). Galectin-3, not active as growth regulator in neuroblastoma cells, has been observed to exert functional antagonism to galectin-1 by competing for the binding to the GM1 ligand present in these cells [72,90,91]. A further example of this opposing activity is found by looking into regulation of tumor supressor $p16^{INK4a}$. Galectin-1 expression was shown to control the reconstitution of $p16^{INK4a}$ with consequent restoration of susceptibility to anoikis in Capan-1 pancreatic cancer cells [84,92] (Fig. 13). Of note, down-regulation of galectin-3, which functions as an inhibitor of galectin-1 in anoikis induction, has been detected

in p16^{INK4a} reconstituted Capan-1 cells [93] (Fig. 13). Interestingly, the C-H/π-interactions in the contact site of galectin-3 allow binding not only to β-galactosides, but also to the anti-apoptotic Bcl-2 protein and pro-apoptotic Bax [94-96].

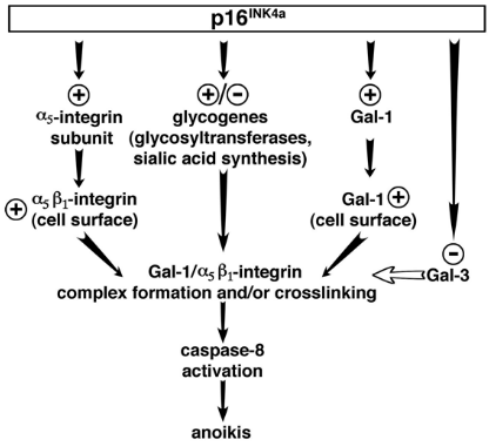


Fig. 13: Scheme depicting orchestrated regulation of cell-surface associated fibronectin receptor, galectin-1 (Gal-1) and galectin-3 (Gal-3) by tumor suppressor p16^{INK4a} in Capan-1 pancreatic carcinoma cells. Gal-1 induces via fibronectin receptor anoikis, antogonizing Gal-3 is down-regulated (with permission [93]).

As mentioned before in section 2.1, this functional diversity for galectins that share a primary affinity for the same ligand, i.e. β-galactosides, is derived from the existence of three topological presentations of CRD domain. In order to elucidate how these differences in molecular arrangment (as well as defined changes in the density of cell surface glycans) affect the functionality of galectins, new test systems have been developed using vesicle-like glycodendrimersomes (vesicles formed by self-assembly of amphiphilic Janus glycodendrimers that contain two identical carbohydrates in their hydrophilic part) and enigineered/natural variants of galectins [97,98]. In this way, enhancement of capacity of galectin-1 for *trans*-interactions (as in haemagglutination) was asserted by covalent connection of its two CRDs [99,100], while presence of a natural sequence deviation reduced it in the case of galectin-8 [101,102].

Equally important to understand the implications of galectin/glycan interactions is the mapping of glycan epitope presence in cells and tissues [98,103-108]. This can be done using specific monoclonal antibodies, plant lectins and, possibly mimicking physiological or pathophysiological interactions, endogenous lectins [109-112]. Furthermore, combination of immuno- and labeled galectin staining, easily analyzable by fluorescence microscopy, allows to detect or exclude co-localization of galectin presence together with its binding sites [9, 113]. Selection of tissue fixation protocols must be done carefully when working on detection of

lectin ligands, as masking or depletion of counterreceptor molecules may occur, prompting to misleading conclusions as noted in [98,114-116].

Table 1: Overview of documented binding partners for mammalian galectins-1 and -3 reactive with the glycan part of cellular glycoconjugates or with proteins.

Type of binding partner	Galectin-1	Galectin-3
Glycoconjugate	Ovarian carcinoma antigen CA125, CD2, CD3, CD4, CD6, CD7, CD43, CD45, CD69, CD95(Fas), CD146, CD166 (ALCAM),carcinoembryonic antigen (CEA), fibronectin (tissue), gastrointestinal mucin, hsp90-like glycoprotein, β1-integrin (CD29), α1/α4/α5/α7β1-, αMβ3- and α4β7-integrins, cell adhesion molecule L1, keratan sulfate, laminin , lamp-1, Mac-2-binding protein, nephrin, neuropilin-1, receptor protein-tyrosine phosphatase (RPTPβ),thrombospondin, Thy-1, tissue plasminogen activator, von Willebrand factor, chondroitin sulfate proteoglycan, distinct neutral glycolipids, ganglioside GM1	CD6, CD7, CD11b of CD11b/CD18 (Mac-1 antigen, CR3), CD131(01000 aminopeptidase N), CD32, CD43, CD44, CD45, CD66a,b, CD71, CD95, CD98, CD147, CD166 (ALCAM), CEA, colon cancer mucin, corneal mucin (MUC16), pancreas cancer mucin-4 and MUC1-D (N-glycan at Asn36), cubilin, C4.4A (member of Ly6 family), desmoglein-2, epidermal growth factor receptor, glycoform of IgE, haptoglobin β-subunit (after desialylation), hensin (DMBT-1), insulin-like growth factor-1 receptor, β1-integrin (CD29), α4/α5/β1- and αvβ3-integrins, keratan sulfate, LI-cadherin, laminin, lamp-1/-2, Mac-2-binding protein, Mac-3, MAG, MP20 (tetraspanin), Na+/K+-ATPase, NG2 proteoglycan, NKp30, TCR complex, tenascin, tissue plasminogen activator, transforming growth factor-β receptor, vascular cell adhesion molecule-1, vascular endothelial growth factor receptor 2, von Willebrand factor, ganglioside GM1
Protein	B lymphocyte adaptor molecule of 32 kDa (Bam32), CaV1.2 L-type calcium channel (α1-subunit), Gemin4, oncogenic H-Ras, OCA-B, pre-B cell receptor (human, not murine system)	AGE products, Alix/AIP-1, ATP synthase b-subunit, axin, bax, bcl-2, bcatenin, Cys/His-rich protein, Gemin4, glycogen synthase kinase-3β, hnRNP Q, mSufu, Mer receptor tyrosine kinase, non-receptor tyrosine kinases c-Abl and Arg, nucleoporin Nup98, nucling, oncogenic K-Ras, OCA-B, pCIP, PIAS1, synexin (annexin VII), TTF-1

From [15] and [117], with permission.

2.5. Non-canonical galectins

The term non-canonical galectin is used to describe proteins that share prominent sequence and structural similarities to the galectin family but have not been seen to possess affinity for β-galactosides [118]. A long time known example is human galectin-10, more commonly referred to as Charcot-Leyden crystal protein, which exhibits a very weak galactoside-binding activity together with a strong disposition to form intracellular crystals, and is present in high levels in eosinophils and basophils [19,119,120]. Though showing low affinity to lactose, this protein is reported to bind mannose [121], while the placental protein PP13 (also known as galectin-13), which shares high sequence homology with galectin-10, presents strong affinity for *N*-acetylgalactosamine and fucose [122]. Advances in genome mapping led to discovery of new members of galectin-related proteins in human and other species. In our model organism chicken, in addition to its relatively low number of canonical galectin genes, two expressed sequence tags (EST)s with similarity to the mammalian non-canonical *g*alectin-*r*elated *i*nter*f*iber prote*in* (GRIFIN) and *g*alectin-*r*elated *p*rotein (GRP) were identified.

GRIFIN was first discovered in rat as a very abundant protein in lens that forms dimers in solution, presents a high amino acid homology to the galectin sequence but has no lactose-binding activity [19, 123]. It was also shown to interact with the small heat shock protein α-crystallin, also present in the ocular lens [124]. Lack of lactose binding can be understood when looking at the otherwise conserved signature sequence of the CRD: two of the key amino acids typically involved in contacting the ligand are changed (Asn48Lys and Arg71Val) [19,123]. This holds true for the mammalian sequences that have been recorded but not for the zebrafish (*Danio rerio*) GRIFIN, which lacks these deviations and is reported to bind lactose like a genuine galectin family member [125]. However, introduction of single, double or triple mutations into the cDNA of rat GRIFIN did not restore its lactose-binding activity [123].

A first sign of GRP presence was deduced from partial ESTs alignment with the genomic sequence [118] and posteriorly confirmed by detection of an mRNA isolated from human haematopoietic stem/progenitor cells (HSPCs), termed as HSPC159 [126]. Despite sharing an important degree of consensus with galectin sequences, binding to the canonical ligand was noted improbable due to deviations in five of the seven key contact residues of the CRD region [19], which was verified in the case of the human GRP-C terminal domain after co-crystallization with lactose and *N*-acetyllactosamine [127]. A compelling characteristic of this protein is the fact that its “sequence appears to be evolutionarily ancient and highly conserved” through different branches of the animal kingdom [19], a sign for presumed strong positive selection and consequent involvement in critical interactions.

With the genomic sequences for both of these two non-canonical galectins having been detected in chicken, characterization of chicken GRIFIN (C-GRIFIN) and GRP (C-GRP), as done for the five canonical galectins [8,11,55,56,59,65], comes out as an evident necessity. The three studies presented here attempt to take initial steps into this direction, aiming for the first complete network analysis of the galectin family in a vertebrate organism.

3. OBJECTIVES

As described, chicken offers the three topological displays combined with a minimum number of galectin family members and cofacilitate to unravel the role(s) of the individual galectins in the network. Our first objective is to run an extensive survey for further nucleotide sequences that may be galectin-like/-related on the genomic level. In this context, the question will be addressed whether the chicken homologues of the non-canonical galectins GRIFIN and GRP, C-GRIFIN/C-GRP are present in the chicken genome.

Furthermore, both proteins will be analyzed (*in silico* and by PCR-technology) on the genomic level to elucidate the gene structures and chromosomal locations. A phylogenetic analysis to calculate the evolutionary relationships between species using alignments of complete sets of GRIFIN/GRP-sequences (after searching in the available genomes) will be performed. Calculations and tree constructions will be done according to the maximum-likelihood method with extensive bootstrapping.

In order to achieve the aim of a full characterization on the protein level, a bundle of tasks is awaiting. These include cloning of the cDNA sequences, insertion into prokaryotic expression vectors (i.a. pET-24a, pGEMEX-1, pGEX-6P-2) and recombinant expression, preferably in the *E.coli* strain BL21. Presumably, both proteins are not lactose-binding, thus the crucial step for the purification of canonical galectins, affinity chromatography on lactose-Sepharose 4B, is not applicable. Alternatively, we have to establish a GST (glutathione-S-transferase) fusion protein approach, with on column cleavage of the GST tag by a specific protease.

In the next step, an accurate analysis of the capacities of both proteins to bind carbohydrates by means of haemagglutination, cell assays and glycan arrays will be performed.

Analysis of the crystal structures (in cooperation with F. Ruiz, A. Romero, Madrid) will give insights on the atomic level, the pattern of hydrogen bonds and may explain the influence of deviations in the contact site on carbohydrate binding.

For the final objectives of our study, we turn from structure to biology. Our experiments include expression profiling for C-GRIFIN/C-GRP in extracts by RT-PCR, and Western Blots as well as in sections of paraffin-embedded/cryogenized tissues with immunohistochemistry. Currently, there are no data available on occurrence and/or bioactivity of C-GRP in chicken.

The full set of canonical/non-canonical galectins now attained, the chicken's bursa of Fabricius will be selected due to its various cell types and structures for a first complete histochemical analysis applying cross-purified antibodies and using the fluorescent-labeled proteins for detection of accessible binding sites.

4. RESULTS

4.1 Publication 1

Chicken GRIFIN: A homodimeric member of the galectin network with canonical properties and a unique expression profile

Gabriel García Caballero[a], Herbert Kaltner[a], Malwina Michalak[b], Nadezhda Shilova[c], Michelle Yegres[d], Sabine André[a], Anna-Kristin Ludwig[a], Joachim C. Manning[a], Sebastian Schmidt[a], Martina Schnölzer[e], Nicolai V. Bovin[c], Dietmar Reusch[d], Jürgen Kopitz[b], Hans-Joachim Gabius[a,*]

[a]*Institute of Physiological Chemistry, Faculty of Veterinary Medicine, Ludwig-Maximilians-University Munich, Veterinärstr. 13, 80539 Munich, Germany*
[b]*Department of Applied Tumor Biology, Institute of Pathology, Medical School of the Ruprecht-Karls-University, Im Neuenheimer Feld 224, 69120 Heidelberg, Germany*
[c]*Shemyakin and Ovchinnikov Institute of Bioorganic Chemistry, Russian Academy of Sciences, ul. Miklukho-Maklaya, Moscow, Russia*
[d]*Pharma Biotech Development Penzberg, Roche Diagnostics GmbH, 82377 Penzberg, Germany*
[e]*Genomics and Proteomics Core Facility, German Cancer Research Center, 69120 Heidelberg, Germany*

[*]Corresponding author. Tel: +49-(0)89-2180-2290; fax: +49-(0)89-2180-992290; E-mail address: gabius@tiph.vetmed.uni-muenchen.de or gabius@lectins.de (H.-J. Gabius).

Abbreviations:

CG, chicken galectin; CHO, Chinese hamster ovary; CRD, carbohydrate recognition domain; GST, glutathione-S-transferase; immunoglobulin G, IgG; linISD, linear in-source decay; MALDI, matrix-assisted laser desorption/ionization; PBS, phosphate-buffered saline; pc, post coitum; reISD, reflectron in-source decay; TBS, Tris-buffered saline; TBS-T, Tris-buffered saline containing 0.5% Tween 20; TOF, time-of-flight

Key words: fiber cells · in-source decay · lectin · lens · phylogenesis

ABSTRACT

Occurrence of the adhesion/growth-regulatory galectins as family sets the challenge to achieve a complete network analysis. Along this route taken for a well-suited model organism (chicken), we fill the remaining gap to characterize its seventh member known from rat as *g*alectin-*r*elated *i*nter-*f*iber prote*in* (GRIFIN) in the lens. Its single-copy gene is common to vertebrates, with one or more deviations from the so-called signature sequence for ligand (lactose) contact. The chicken protein is a homodimeric agglutinin with capacity to bind β-galactosides, especially the histo-blood group B tetrasaccharide, shown by solid-phase/cell assays and a glycan microarray. Mass spectrometric identification of two lactose-binding peptides after tryptic on-bead fragmentation suggests an interaction at the canonical region despite a sequence change from Arg to Val at the site, which impairs reactivity of human galectin-1. RT-PCR and Western blot analyses of specimen from adult chicken organs reveal restriction of expression to the lens, here immunohistochemically throughout its main body. This report sets the stage for detailed structure-activity studies to define factors relevant for affinity beyond the signature sequence and to perform the first complete network analysis of the galectin family in developing and adult organs of a vertebrate.

1. Introduction

Rooted in the concept of the sugar code, work on receptors for glycans (lectins) has convincingly demonstrated the ubiquitous relevance of carbohydrate-protein recognition in a wide variety of physiological processes [1-7]. Consequently, more than a dozen families of animal lectins have been thoroughly mapped, most recently by computational database mining of genomic information (for a case study on C-type lectins, please see [8]). Applied to galectins (β-galactoside-binding proteins with β-sandwich fold; for detailed information, please see [2, 9-14]), occurrence in many metazoan phyla and conservation of a set of seven key amino acids (the sequence signature) are revealed [15, 16]. They are known from crystallographical studies to contact the canonical ligand lactose by hydrogen bonding and C-H/π-interactions (with Trp) [5, 15, 17]. In general, reflecting the assumed significance of this property for the galectins' activity as bioeffector, no sequence deviation at these sites is observed for orthologues in inter-species comparison. A notable exception is the *g*alectin-*r*elated *i*nter-*f*iber prote*in* (GRIFIN), what explains the interest to study this family member.

First detected as developmentally regulated lens-specific protein in rats, its immunohistochemical localization at the interface between adjacent fiber cells led to the given acronym [18]. In extracts of rat nuclear fiber cells, GRIFIN was found in the insoluble fraction by proteomic analysis [19]. Additionally, Barton et al. [20] indicated that "GRIFIN appears to be a cytoplasmic protein (data not shown)", when reporting murine GRIFIN's engagement in complex formation with α-crystallin (at K_D: 13.6 ± 5.3 μM). Looking at its signature sequence, two drastic changes (Asn48Lys, Arg71Val) questioned its ability to bind lactose. In fact, site-specific mutations of these amino acids in human galectin-1, i.e. Asn to Asp and Arg to His, had impaired the affinity to lactose [21]. Seemingly fitting, rat GRIFIN with its natural sequence changes indeed lacked binding capacity [18]. In stark contrast, its orthologue in zebrafish (*Danio rerio*) presented overall conservation in all seven signature-sequence positions and was reported to be a "bona fide galectin family member" [22]. Obviously, despite maintaining its expression in lens fiber cells [18, 22], the GRIFIN gene appears to have undergone a fundamental sequence divergence at crucial sites after phylogenetic separation into the lineages that lead either to fish or to mammals. Tracking occurrence of GRIFIN genes and examining their sequences in organisms at different branches of the evolutionary tree was therefore our first step to investigate this unusual phenomenon.

At the same time puzzling and challenging, conversion to the signature sequence for active galectins by introducing single, double or triple mutations into the cDNA for rat GRIFIN did

not establish capacity of the resulting proteins to bind to an affinity resin presenting lactose [18]. This clear indication for an unsuspected key involvement of other amino acids directed our interest to a natural GRIFIN protein with only one deviation from the sequence signature, i.e. chicken (C-)GRIFIN, which shares the Arg to Val substitution with rat GRIFIN. Whether it is active as lectin was an entirely open question. Embedded in our efforts toward the aim to achieve the first comprehensive network analysis of the galectin family, for which chicken is a suited model organism due to its relatively small number of proteins representing all types of galectins [13, 15, 23-30], we here report on the C-GRIFIN gene and its promoter, its expression and immunohistochemical localization of the protein. Moreover, we answer the question on its ability to bind β-galactosides in solid-phase/cell-based assays, detect oligosaccharides of strong avidity in an array and identify lactose-binding peptides by mass spectrometry after tryptic digestion of C-GRIFIN bound to affinity resin. Hereby, a so far missing piece is added to the network puzzle of chicken galectins (CGs) and an attractive platform is made ready for ensuing structure-activity studies to amend the signature-sequence concept.

2. Materials and methods

2.1. Database mining and processing of sequence information

Systematic searches for sequences, which satisfy stringent criteria of functionality and homology, in databases of genome and expressed sequence tag (EST) information, multiple alignments of amino acid sequences and the delineation of the chromosomal neighborhood of the detected gene for an orthologue of rat GRIFIN were performed, as given in detail previously [27, 31]. Using the Maximum Likelihood method implemented in the MEGA6 software package [32], the sequences of various organisms were arranged based on highest likelihood into a phylogenetic tree, processing involving a bootstrap analysis (with 1000 replicates) and the Neighbor-Joining method applied to a matrix of pairwise distances estimated using a Jones-Taylor-Thornton model. Multiple alignments of amino acid sequences were performed using the Clustal Omega software and edited in Jalview or ESPript. The computational scrutiny of the manually edited proximal promoter sequences (-2500 bp upstream of the translation start, applying the sequence text view tool of NCBI Gene [33]) to compile occurrence of putative sites for binding of transcription factors was carried out with the MatInspector software (http://www.genomatix.de/index.html) based on the TRANSFAC® database and the program-inherent 634 matrices grouped into 279 families (Matrix Library 9.3, released March 2015), setting similarity thresholds as described in previous analyses on the corresponding sequence section of genes of chicken and human galectins [30, 34]. In detail, individual monitoring and comparisons were performed for respective sequences of C-GRIFIN (Gene ID: 427660) as well as of 10 eye-expressed chicken genes, i.e. aquaporin-0 (Gene ID: 374124), αA-crystallin (Gene ID: 418546), βA1-crystallin (Gene ID: 396499), βB1-crystallin (Gene ID: 374000), δ2-crystallin (Gene ID: 417545), connexin-43 (Gene ID: 395278), connexin-46 (Gene ID: 428084), connexin-50 (Gene ID: 395846), filensin (Gene ID: 396056) and phakinin (Gene ID: 396141).

2.2. Cloning, recombinant expression and purification

Total RNA from eyes of a 14-day-old chicken embryo was isolated using an RNeasy kit (Qiagen, Hilden, Germany) according to the manufacturer's information, and SuperScript reverse transcriptase (ThermoFisher, Schwerte, Germany) generated cDNA used for PCR amplification by the sense primer 5´-CATATGGCACTGCGGTTTGAGG-3´ (*NdeI* restriction

site underlined) and the antisense primer 5′- AAGCTTCTAGTAAAGACCACGTTTGG-3′ (*HindIII* restriction site underlined). The reaction was performed with *Taq* DNA polymerase (Qiagen) at conditions recommended by the manufacturer. The amplification products were separated from PCR reagents by gel electrophoresis in 3% agarose, then extracted from the gel and finally ligated into the *EcoRV*-linearized pGEM®-T Easy Vector (Promega, Mannheim, Germany) presenting 3′T overhangs. This vector was used for transformation of the commercially available *E. coli* strain TOP10 (Promega) and ensuing amplification. Resulting plasmids were then purified, its cleavage with restriction enzymes (*NdeI/HindIII*) yielded digestion products, which were separated by agarose gel electrophoresis. Subsequently, the cDNA sequences with the appropriate restriction sites at the 5′-end (*NdeI*) and the 3′-end (*HindIII*) were ligated into the expression vector pGEMEX-1 (Promega) pretreated with *NdeI*/*HindIII*. Alternatively, a cDNA construct for a glutathione S-transferase (GST)–C-GRIFIN fusion protein was established using the pGEX-6p2 vector (GE Healthcare, Munich, Germany). PCR amplification of the C-GRIFIN cDNA sequence for insertion into this vector was directed by the sense primer 5′- CGCTAGGGATCCTTTGAGGCTCTGTACCCA-3′ (*BamHI* restriction site underlined) and the antisense primer 5′-CGTACGGTCGACCTAGTAAAGACCACGTTTG-3′ (*SalI* restriction site underlined). The reaction was performed with PerfectTaq DNA polymerase under conditions recommended by the manufacturer (5Prime GmbH, Hamburg, Germany). The combination of this sequence with the linearized commercial vector generated a cleavage site for a protease (please see below) to separate GST from C-GRIFIN, which is then inevitably N-terminally extended by the Gly-Pro dipeptide. Sequence reconstitution to the first three C-GRIFIN amino acids, altered to establish the *BamH1* site, was accomplished by a modified kit-based QuikChange® site-directed mutagenesis protocol (Agilent Technologies, Waldbronn, Germany).

Both expression vectors were used for protein production in *E. coli* BL21 (DE3)-pLysS cells (Promega), first in systematic test series to optimize temperature, duration of induction/incubation and concentration of the inducer isopropyl-β-D-thiogalactopyranoside. Pellets of cultures were frozen and cells lyzed by sonification on ice in 20 mM phosphate-buffered saline (PBS), pH 7.3, containing 1 mM dithiothreitol (7 ml per gram of wet cell paste). The supernatant after centrifugation of the suspension was subjected to affinity chromatography either on glutathione-presenting Sepharose 4B (GE Healthcare) or home-made lactosylated Sepharose 4B [26, 35]. The fusion protein was cleaved by human rhinovirus 3C protease (fused to GST; used at a ratio of 1:100 (w/w)), and enzyme and released GST were removed by a second step of affinity chromatography. C-GRIFIN-containing solutions after both procedures

were run over PD-10 columns to remove reagents, then frozen, lyophilized and the protein was stored at -20 °C until use.

2.3. Analytical procedures (gel electrophoresis and filtration, mass spectrometry)

One- and two-dimensional gel electrophoretic analyses were performed, as described in detail for other CGs [27, 28]. Mass spectrometric characterization comprised mass determination by matrix-assisted laser desorption/ionization (MALDI) time-of-flight (TOF) measurements of the protein and its tryptic peptides as well as in-source decay (ISD) processing, which led to reflectron ISD (reISD) and linear ISD (linISD) spectra on an Ultraflex® TOFTOF 1 instrument with respective matrices (Bruker Daltonik, Bremen, Germany), as described for human galectins previously [36, 37]. Tryptic peptides were also fractionated and analyzed after protein exposure to two types of stress conditions (addition of H_2O_2 up to 0.05% (v/v) to 50 mM PBS at pH 7 and exposure at 25 °C for 24 h or raising the temperature to 40 °C for up to seven days in 20 mM His-HCl buffer, pH 6, containing 100 mM NaCl) using tandem MS/MS on a LTQ Orbitrap Velos electrospray mass spectrometer, as described [38-40]. Gel filtration (100 µg of protein in 50 µl buffer) was performed on a calibrated Superose HR10/30 column using an ÄKTApurifier 10 system (GE Healthcare) at 4 °C and a flow rate of 0.5 ml/min.

2.4. Expression profiling by RT-PCR and Western blotting

Total RNA from tissue specimen of 4-week-old chickens and the respective cDNA were prepared and tested in PCR amplification assays with the sense primer 5′-AAGCTTATGGCACTGCGGTTTGAGG-3′ and the antisense primer 5′-TCTAGACTAGTAAAGGCCTCGTTTGG-3′, as routinely done in CG expression analysis [26-28]. The calculated length of amplified cDNA for C-GRIFIN is 420 bp. In detail, the analyses were carried out in a volume of 20 µl containing 2.5 µg cDNA, 0.5 U Taq polymerase (Qiagen), 50 µM dNTPs, 1×PCR buffer (commercial mixture of Tris-HCl, KCl, $(NH_4)_2SO_4$, 1.5 mM $MgCl_2$, pH 8.7), 1×Q-solution and 0.1 µM of sense and antisense primers at 20 °C. Amplification of the galectin-specific cDNAs was started with an initial denaturation step at 94 °C for 4 min, then 29 consecutive cycles of the following series of steps were performed: denaturation at 94 °C for 45 s, annealing at 60 °C for 45 s and extension at 72 °C for 1 min. The final extension step was carried out at 72 °C for 10 min. PCR products and reagents were

separated in 2% agarose gel electrophoresis at 80 V for 40 min. The loading control with chicken β-actin-specific cDNA was established with the sense primer 5′-ATGGCTCCGGTATGTGC-3′ and the antisense primer 5′-ACCGTGTTCAATGGGGT-3′. The calculated length of amplified cDNA was 185 bp.

Extracts of lens, retina and bursa of Fabricius were obtained and processed by SDS gel electrophoretic analysis and Western blotting, as described previously [27]. Polyclonal antibodies against C-GRIFIN were raised in rabbits, as described [41], and thoroughly checked for presence of any cross-reactivity against other CGs by ELISAs and Western blotting using the complete panel of purified CGs, which was then removed by affinity chromatography on resin with respective CGs. For detection in extracts, proteins (100 µg of tissue lysates or 12 ng of purified recombinant C-GRIFIN as positive control) were separated on a discontinuous SDS polyacrylamide gel (4% stacking gel, 15% running gel) and then transferred by tank blotting onto a nitrocellulose membrane (0.2 µm pore size; Schleicher & Schuell, Dassel, Germany). Residual sites on the matrix for binding protein were saturated by incubation with a blocking solution of Tris-buffered saline (TBS, pH 7.5) containing 0.1% Tween 20 (TBS-T) and 5% powdered skim milk. Membranes were routinely incubated overnight at 4 °C with immunoglobulin G (IgG) fractions against C-GRIFIN (0.5 µg/ml) in blocking solution. Following a washing step with TBS-T, membranes were incubated with goat anti-rabbit IgG-horseradish peroxidase conjugate (0.5 µg/ml; Sigma-Aldrich, Munich, Germany) in blocking solution for 1 h at room temperature and developed using enhanced chemiluminescence detection, driven by a substrate mixture with 1.25 mM sodium salt of luminol in 2 ml 0.1 M Tris-HCl (pH 8.6), 0.2 ml of a 6.7 mM *p*-coumaric acid solution in dimethyl sulfoxide and 0.6 µl H_2O_2 (30% v/v) at room temperature. Exposure time of the processed blots to CL-XPosure™ X-ray film (Pierce, Bonn, Germany) was set from 2 to 10 min to yield optimal signal intensity at minimal background.

2.5. Immunohistochemistry

Tissue specimen from 4-week-old chickens were immediately fixed by immersion in Bouin's solution for 24 h at 4 °C, dehydrated using solutions of increasing content of ethanol (70%, 80%, 96% and 100%), then isopropanol and finally xylene, embedded in paraffin wax at 61 °C, cut into serial sections of 5 µm in thickness, and these sections were mounted on Superfrost slides (Menzel, Braunschweig, Germany), a protocol applied in previous studies to ensure comparability [28, 30, 42], following systematic titrations of the antibody concentration

to optimize signal-to-background ratio. Alkaline-phosphatase-labeled second-step antibody (goat anti-rabbit, 0.66 µg/ml; Sigma-Aldrich) was applied for 1 h at room temperature, and reagents of the Vector® Red alkaline phosphatase Substrate Kit SK-5100 (Enzo Life Sciences, Lörrach, Germany) were incubated for 20-30 min in the dark. Counterstaining with Mayer's hemalaun and dehydration were followed by mounting sections in Eukitt® (Kindler, Freiburg, Germany). Microphotographs during light microscopy were recorded using an AxioImager.M1 microscope (Carl Zeiss MicroImaging, Göttingen, Germany) equipped with an AxioCam MRc digital camera and the software AxioVision version 4.6.

2.6. Binding assays

Haemagglutination of trypsin-treated glutaraldehyde-fixed rabbit erythrocytes using 2-fold serial solutions of C-GRIFIN and proto-type CGs as positive control were performed in 96-well (V-shaped) microtiter plates, as described [43]. Following biotinylation under activity-preserving conditions with the N-hydroxysuccinimide ester derivative (Sigma-Aldrich), labeled C-GRIFIN was tested in a solid-phase system (microtiter plate wells presenting surface-adsorbed (neo)glycoproteins after incubation of solutions (0.5 µg/50 µl of PBS) at 4 °C overnight, blocking residual sites for protein adsorption and thorough washing), using a streptavidin-peroxidase conjugate (0.5 µg/ml; Sigma-Aldrich) and the chromogenic substrates *o*-phenylenediamine (1 mg/ml)/H_2O_2 (1 µl/ml) as reporters, as described [36, 44]. The glycan array was produced by printing using solutions with glycan (50 µM) or polysaccharide (10 µg/ml) and then applied using biotinylated C-GRIFIN and the proto-type CGs, as described [45]. Lectin association to glycans on the chip (total of 672 slots with about 600 different glycocompounds: for full representation of structures, please see http://csdb.glycoscience.ru/bacterial) was converted to a fluorescent signal using ALEXA Fluor® 555-labeled streptavidin, the signals' intensity was monitored in an Innoscan 1100 AL instrument (Innopsys, Carbonne, France). The resulting data were processed using ScanArray Express 4.0 software, the fixed 70 µm-diameter circle method and Microsoft Excel.

Binding to cell surfaces was quantitated using fluorescent C-GRIFIN, prepared by reacting fluorescein isothiocyanate (isomer I; Sigma-Aldrich) at 0.25 mg/mg protein in 0.1 M Na_2CO_3 solution (pH 9.0) containing 0.9% NaCl and 50 mM lactose for 1 h at room temperature, and a panel of lines, i.e. Chinese hamster ovary (CHO) parental line and its Lec1, 2, 8, 19 glycosylation mutants, kindly provided by P. Stanley (Albert Einstein College of Medicine, Bronx, NY), as well as human pancreatic cancer (Capan-1) line reconstituted for expression of

the tumor suppressor $p16^{INK4a}$, kindly provided by K. M. Detjen, Charité-Universitätsmedizin Berlin, Berlin, Germany, as described [27, 46]. Mass spectrometric identification of lactose-binding peptides obtained from C-GRIFIN bound to the affinity resin (0.2 ml packed volume) after trypsin treatment (lectin:enzyme ratio of 20:1) followed the protocol developed for human galectins-1 and -3 and lactose-presenting Sepharose 4B beads, with five washing steps of 0.5 ml PBS after the in situ cleavage to remove peptides without affinity to lactose and elution with 0.2 ml acetonitrile:trifluoroacetic acid (0.2%) (2:1; v/v) [47].

3. Results and discussion

3.1. The C-GRIFIN gene

Systematic database mining disclosed common presence of the GRIFIN gene in vertebrates (please see Fig. 1 for examples; for a complete listing of the 144 species, please see Supplementary material, Table S1, for complete listing of the deduced amino acid sequences in the form of a molecular phylogenetic analysis, Fig. S1; this information thus extends the already reported evidence for GRIFIN gene presence in dog, guinea pig, mouse and rat, also chicken and zebrafish [16, 18, 19, 22, 48, 49]). In all cases, the copy number was found to be one. This finding excludes an inter-species variation due to duplication events. They had been found for other members of this family such as galectins-7 or -9 in mammals [31]. Mapping the chromosomal neighborhood of the C-GRIFIN gene disclosed no spatial association to any other gene for a CG (Supplementary material, Fig. S2A). In inter-species comparison, the neighborhood of the GRIFIN gene was remarkably similar in organisms from different branches of the vertebrate section of the phylogenetic tree (Supplementary material, Fig. S2B). Apparently, the gene and its chromosomal placement are of ancient origin, without having been subject to copy-number or positional dynamics. Looking closely at the gene's organization, it has the common pattern of exon/intron distribution, as documented in Supplementary material, Fig. S3. Thus, the GRIFIN gene is a conserved part of the vertebrate galectin network. Having collected this information and retrieved sequences from the databases, it was possible to perform alignments and to examine in detail especially the status
of the signature sequence on the level of amino acids, setting C-GRIFIN in relation to the other CGs and to its orthologues in other vertebrates

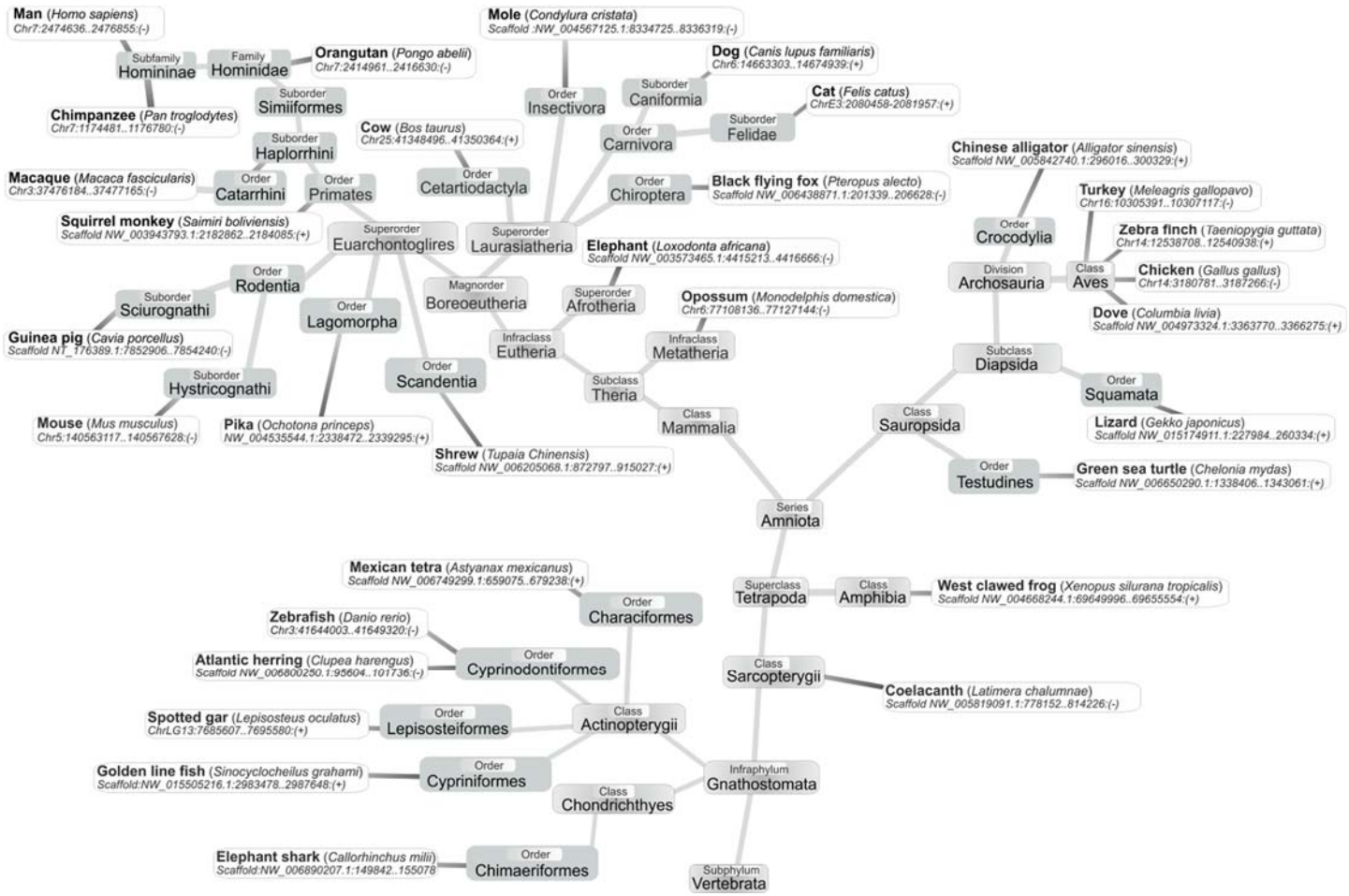


Fig. 1. Information on the GRIFIN gene in man and 32 selected vertebrate species of different branches of the taxonomic tree as available in databases. Respective information on the genes, present either in forward (+) or in reverse (-) direction, is presented as chromosome or scaffold number together with the number of nucleotides, characterizing the precise position and length of the gen.

Respective calculations delineated a close evolutionary relationship to the chimera-type CG-3, separated by a branching point from galectin-related protein (GRP) and the C-terminal carbohydrate recognition domain (CRD) of tandem-repeat-type CG-8 (Fig. 2A; for complete listing of the CG sequences, please see Supplementary material, Fig. S4). Interestingly, the positions of the three proto-type (homodimeric) CG-1A, -1B and -2 were rather distant. This obvious separation argues against immediate origin of C-GRIFIN from this group by a duplication event, as is the case for the paralogue pair CG-1A and -B [50]. This result confirmed and extended a previously reported status of relationship [16], underscoring a genealogically closest vicinity of GRIFIN to CG-3. The degree of conservation in the signature sequence was expectably variable, rather substantial among the CGs and more diverse among GRIFIN in vertebrates (Fig. 2B, Fig. S4, Fig. S5). Intriguingly, a substitution at the equivalent of the Val73 position (from Arg in the homodimeric (proto-type) CGs, which share the occurrence of a gap at the preceding place) appears tolerable in CG-3 and both CRDs of CG-8, which yet have an Arg/Lys residue adjacent to this position and bind well to β-galactosides [27, 28, 51]. Because homodimeric (proto-type) CGs maintain Arg at this position and rat GRIFIN is a homodimer with Val at this site without affinity to lactose [18], it is uncertain whether C-GRIFIN is active as lectin, despite the conservation at the other positions. In addition to answering this question, it was also of interest to clarify the quaternary structure of C-GRIFIN. To address these issues and also obtain material for raising polyclonal antibodies, the cDNA was cloned to produce the protein.

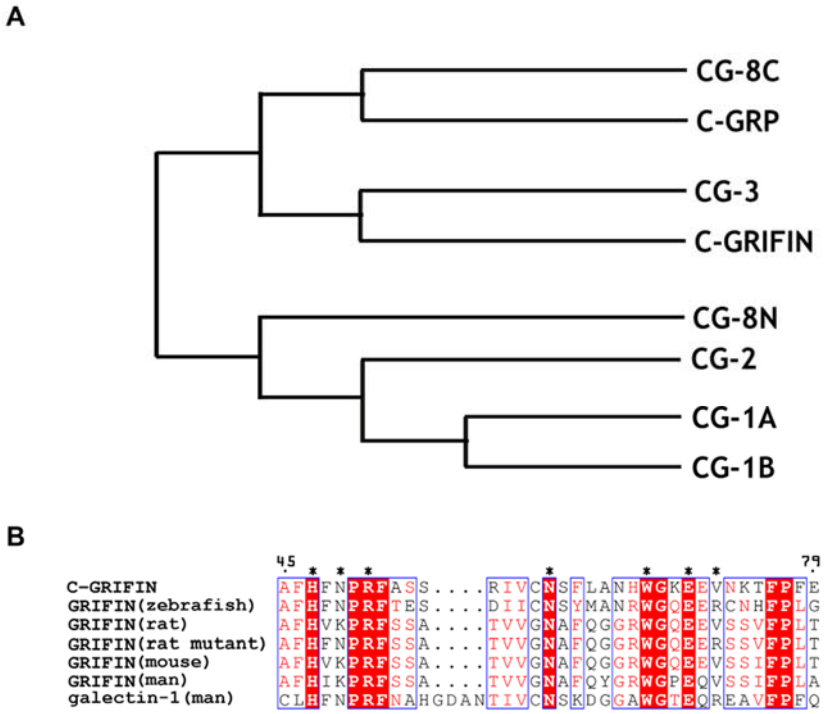


Fig. 2. Phylogenetic family-tree diagram (A), and alignment of the section of the amino acid sequence with the signature region (amino acids 45-79) of C-GRIFIN with those of GRIFIN of zebrafish, rat, the rat triple mutant[18], mouse and man as well as human galectin-1 (B). Strictly conserved (red background) and homologous residues (>70% conservation; boxed red letters) are highlighted by coloring. Amino acids in contact to the canonical ligand lactose are indicated with asterisks.

3.2. Characterization of C-GRIFIN

In the first stage, a design of the expression vector was selected to facilitate the generation of a fusion protein. Hereby, obtaining product became independent from a carbohydrate-binding capacity and yield in glycan-based affinity chromatography. When tested chromatographically for carbohydrate-binding activity with lactose-containing affinity resin, both the fusion protein and the purified C-GRIFIN bound to the beads and could be eluted by the cognate sugar. C-GRIFIN thus is a β-galactoside-binding protein. As consequence, C-GRIFIN-specific cDNA without any artificial extension in combination with this type of affinity chromatography, as routinely used for CGs, could be used. Production directed by the most efficient vector generated a yield of about 60 mg C-GRIFIN/l at 30 °C during induction and incubation overnight.

Gel electrophoretic analyses documented purity with a single spot in isoelectric focusing at a pI value of about 6.2 (calculated pI at 6.15) (Fig. 3). Molecular mass determination of the protein and its tryptic peptides as well as the information obtained by reISD/linISD sequencing, in combination resulting in complete coverage, revealed that C-GRIFIN was neither truncated nor modified posttranslationally (Fig. 4, Fig. S6, Fig. S7, Tables S2A,B). Recalling GRIFIN presence in rat lens and expecting a rather long half-life of the protein at this site, it was of interest to examine the protein's susceptibility to chemically induced modifications, i.e. oxidation and deamidation. Upon exposure to hydrogen peroxide (up to 0.05% in buffer for 24 h), modification in peptide 6-23 occurred at Met14 alone (up to a level of 15.7%) and at Met14/Trp18 together up to 8.9%, in peptide 26-52 up to Met32 of 45.9% and in peptide 59-70 up to Trp68 of 18.6%. Notably, the Trp68 residue at the canonical contact site for lactose obviously is substantially sensitive to oxidation. For electrolectin, the first animal galectin purified from the electric eel, the equivalent Trp was oxidized by O_2 to oxindole and lectin activity lost after 30 min of exposure, the reaction precluded by the protective effect of lactose [52]. Deamidation levels increased linearly over time, reaching 9.2%/13.5% at 25 °C/40 °C after seven days for Asn residue(s) in peptide 58-70 and 3.7%/5.8% in peptide 121-136.

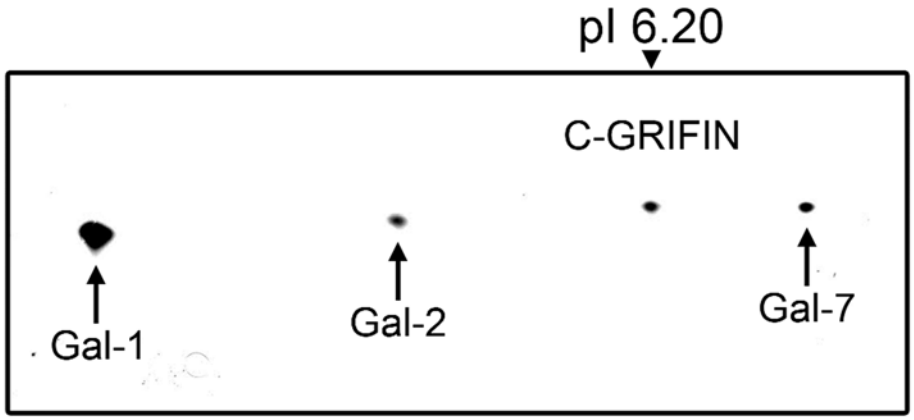
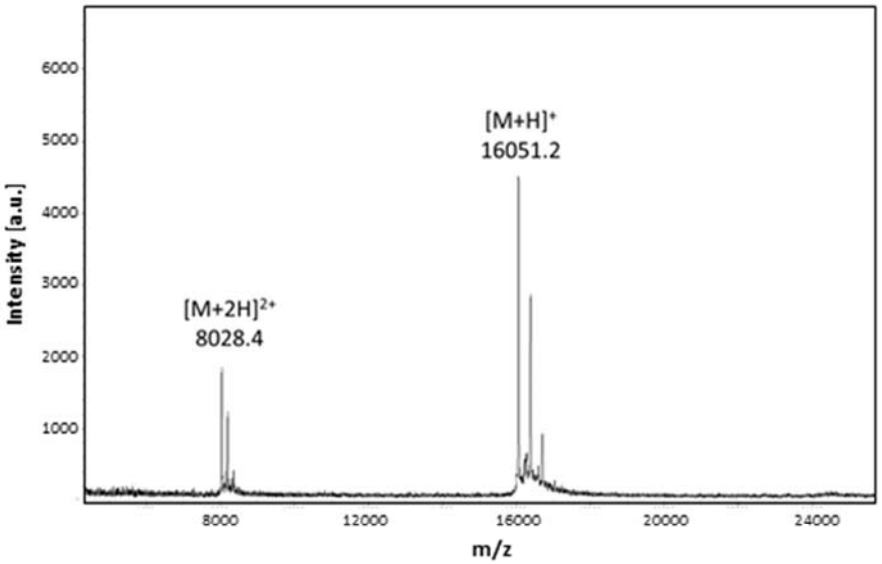


Fig. 3. Analysis of recombinant C-GRIFIN by two-dimensional gel electrophoresis. The relevant part of a gel after two-dimensional gel electrophoresis separating C-GRIFIN and the human galectins-1 (calculated pI: 5.34; experimentally determined pI: 5.28), -2 (calculated pI: 5.93; experimentally determined pI: 5.78) and -7 (calculated pI: 7.00; experimentally determined pI: 6.46) is shown after staining.

A



B

Sequence	[MH]+ (mono) calc	[MH]+ (mono) exp
RGLY	508.3	508.2
FASSR	567.6	567.2
ILQYK	664.4	664.4
TFPFEAK	839.4	839.4
EEVNKTFPFEAK	1438.7	1438.8
IVCNSFLANHWGK	1488.7	1488.8
LQILNDIEISSVEITK	1815.0	1815.1
LQILNDIEISSVEITKR	1971.1	1971.2
FEALYPEGMCPGWSVVVK	2012.0	2012.0
IVCNSFLANHWGKEEVNK	2088.1	2088.4
ALRFEALYPEGMCPGWSVVVK	2352.2	2352.3
EPFQVEIYSDQDYFHIFIDENK	2776.3	2776.4
TSSNTSMFEINFLSHPGDQIAFHFNPR	3094.4	3094.6
GKTSSNTSMFEINFLSHPGDQIAFHFNPR	3279.6	3279.8
TFPFEAKEPFQVEIYSDQDYFHIFIDENK	3596.7	3596.9

C

ALRFEALYPEGMCPGWSVVVKGKTSSNTSMFEINFLSHPGDQIAFHFNPRFASSRIVCNSFLA
NHWGKEEVNKTFPFEAKEPFQVEIYSDQDYFHIFIDENKILQYKHRQKQLSSITKLQILNDIEISS
VEITKRGLY

Fig. 4. MALDI-TOF-MS of C-GRIFIN. (A) Molecular mass determination. The two peaks represent its singly and doubly charged molecular ions. (B, C) Peptide mass fingerprinting. The list of detected peptides with their calculated (calc) and experimentally measured (exp) mass values (B) and the sequence coverage (C, shaded) are shown.

In gel filtration, the protein eluted as a single peak at the calculated position of the homodimer (Supplementary material, Fig. S8). Addition of lactose led to a slight shift in the elution volume, indicating an increase in the protein's shape, along with the appearance of a minor second peak in the range of elution volume expected for a dimer of dimers (Fig. S8). Of note, loading the CRD of a homodimeric galectin with this disaccharide can indeed entail a change in the gyration radius. This had first been detected for human galectin-1 by small angle neutron scattering that made a reduction of the gyration radius from 19.1±1 Å to 18.2±1 Å measurable [53]. The hydrodynamic behavior of homodimeric CGs, too, is prone to change during lactose binding, CG-1B responding with an increase in the diffusion constant of 5.6%, CG-2 with a reduction by 3.8% [29]. In terms of activity and quaternary structure, C-GRIFIN thus is a homodimeric (proto-type) galectin, yet widely separated from the set of the three well-characterized CG-1A, -1B and -2 in the alignment-based scheme showing phylogenetic relationships (Fig. 2A). As the acronym signals, it has a distinct (very restricted) expression profile in rats [18] (recently extended to white adipose tissue in obese rats and to retina after intravitreal sex steroid injection, as determined by proteomic mapping [54, 55]) and in zebrafish, where RT-PCR analysis gave signals in specimen of adult brain, intestine and oocyte besides the expected positivity with RNA from eyes [22]. To monitor C-GRIFIN expression in chicken, we performed RT-PCR and Western blotting as well as immunohistochemistry.

3.3. Expression of C-GRIFIN

Profiling of expression on the levels of mRNA and protein gave signals when testing adult eyes (Fig. 5, Fig. 6). RNA preparations from other organs (heart, jejunum, liver, lung, pancreas, proventriculus, spleen and thymus) were invariably negative (not shown). Material from organs such as bursa of Fabricius or kidney, which also contains other CGs such as CG-3, was negative in Western blotting as were extracts of adult retina. Here, CG-1A had been reported to be expressed in all layers [56, 57]. The absence of any signal in kidney or retina extracts, compared to the strong reaction in the specimen from lens, documented the specificities of expression and of the antibody preparation. As also ascertained with purified CGs, it is apparently not cross-reactive to any other CG when testing organ extracts. Because galectin fingerprinting in the human eye had detected presence of all seven tested family members with regional specificity and characteristic profiles [58], such a cross-reactivity must be rigorously excluded. Recalling the alignment-based positioning of C-GRIFIN and CG-3 on the same branch of the phylogenetic scheme (Fig. 2A), it is noteworthy that galectin-3 immunoreactivity in

mammalian lens had been found in fibers of the outer cortex and also anterior epithelial cells with a heterogenous profile of intensity [59, 60], indicating the possibility for its co-expression with GRIFIN.

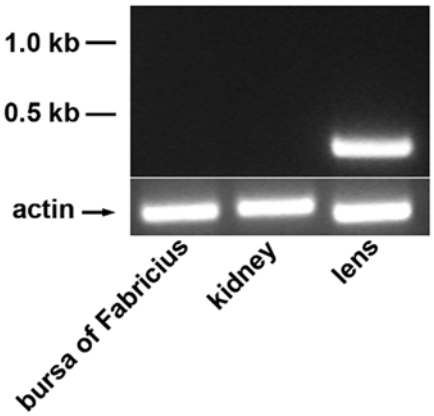


Fig. 5. Expression profiling of C-GRIFIN by RT-PCR. The presence of C-GRIFIN-specific mRNA was detected in samples from tissues of 4-week-old chickens as band at 420 bp. Probing of mRNA for actin served as positive control. Positions of markers are indicated (left).

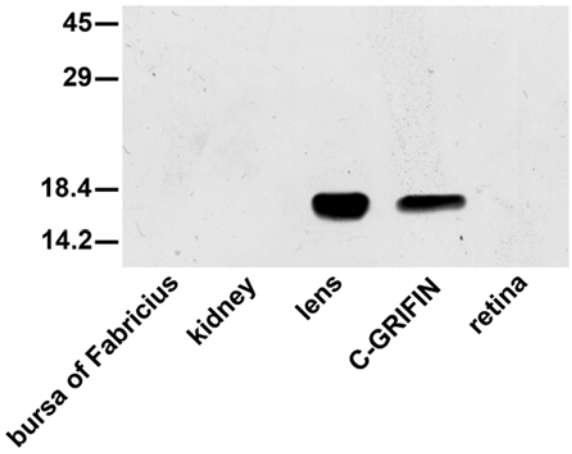


Fig. 6. Expression profiling of C-GRIFIN by Western blotting. Detection of C-GRIFIN in extracts from tissue samples (100 µg total protein) of 4-week-old chickens using a polyclonal IgG preparation against C-GRIFIN (anti-C-GRIFIN: 0.5 µg/ml) free of any cross-reactivity to the other five CGs and to C-GRP. Positions of molecular weight markers are indicated (left), a control (termed C-GRIFIN) with purified recombinant protein (12 ng) is included.

Having ensured absence of cross-reactivity, the anti-C-GRIFIN antibody preparation could be used for immunohistochemistry. Respective results, together with a scheme of the avian lens, are presented in Fig. 7. Strong immunoreactivity for C-GRIFIN was detected in the semicircular fibers of the main lens body and to a weaker extent in the transition zone between the lens body

and the annular pads (Fig. 7). No staining was observed in the hexagonally shaped and radially arranged nucleated fibers of the annular pads, which form an outer ring around the equator of the lens body, the cuboidal cells of the epithelial layer of the lens at the anterior part and the capsule, which covers the whole lens (Fig. 7). When comparing this profile to data from adult rat lens [18], where “immunoreactivity between rows of fiber cells” in a “peripheral region underlying the lens epithelium” had been described, a rather uniform and strong staining throughout the main lens body, both between and within the fibers and with no distinction between peripheral or central regions, characterized C-GRIFIN localization. Thus, C-GRIFIN’s localization justifies the acronym as lens-specific protein, with a rather uniform expression in the lens body.

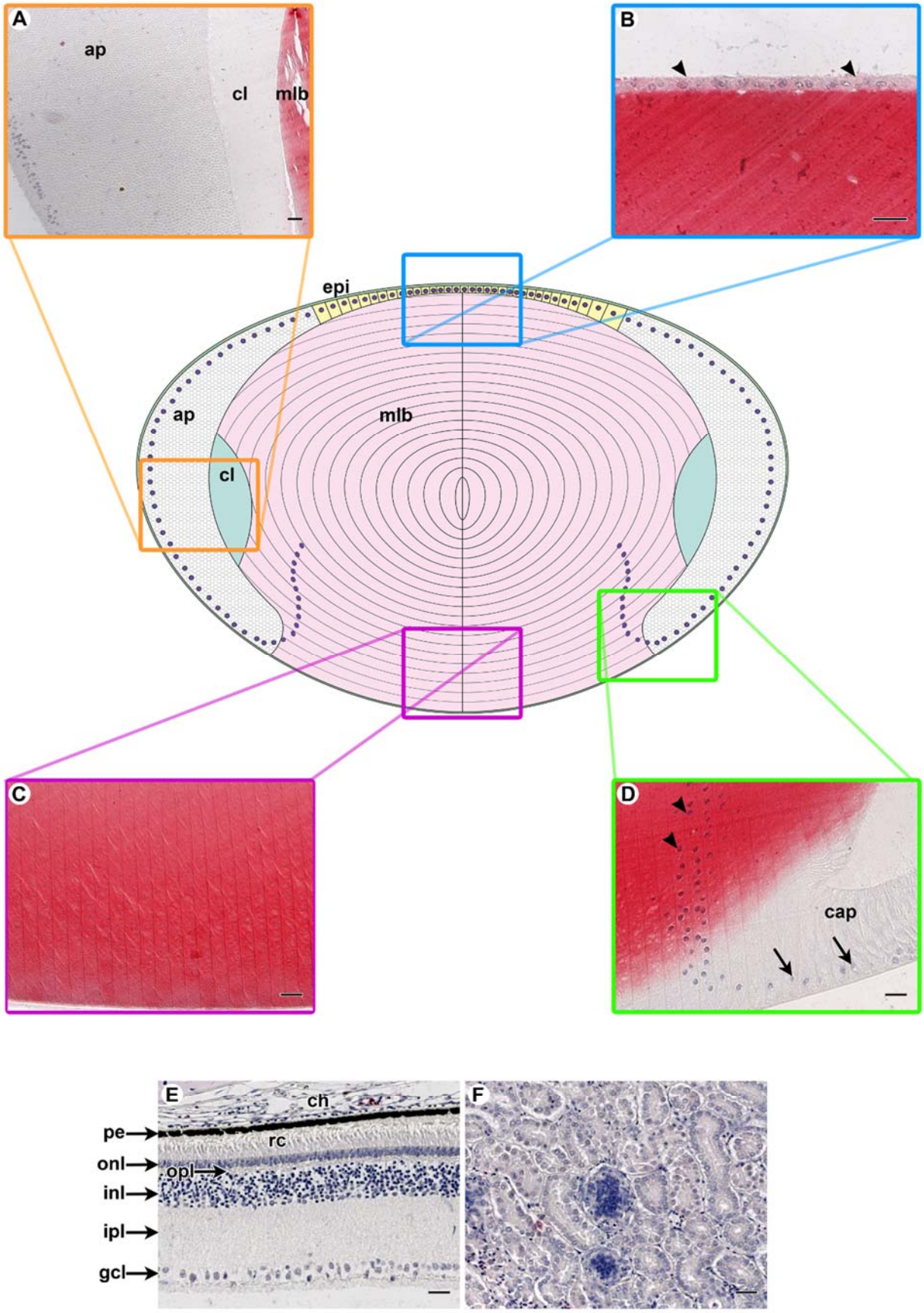


Fig. 7. Immunohistochemical localization of C-GRIFIN in fixed cross-sections through lens together with a scheme of the avian lens (top part). As controls, processed sections of retina and kidney from 4-week-old chickens

are presented (bottom part). The microphotographs of distinct regions of the chicken lens (A-D) are assigned to the scheme (center of top part), which defines their positions by color-coded boxes. (A) No reactivity was observed in the hexagonally shaped nucleated fibers of the annular pads (ap) that form an outer ring around the equator of the lens body and in the lumen of the cavum lenticuli (cl), a wedge-shaped structure between the equatorial-posterior parts of the annular pads and the main lens body (mlb). (B) Strong signal for presence of C-GRIFIN in superficial fibers of the main lens body at the anterior region of the lens. The cells of the cuboidal epithelium (epi, *arrowheads*) were negative. (C) C-GRIFIN presence in typically flat band-like deeper central fibers of the main lens body at the posterior surface of the lens. (D) Transitional zone between annular pad and main lens body in the equatorial-posterior region of the lens. Staining intensity is gradually enhanced with increasing degree of differentiation of C-GRIFIN-negative curved annular pad fibers (cap) into superficial fibers of the main lens body. Of note, spherical nuclei (*arrows*) of annular pad fibers were negative, whereas the ellipsoid nuclei (*arrowheads*) of the superficial fibers were strongly positive. (E) No signal for C-GRIFIN was detected in the layers of the retina: choroid (ch), pigment epithelium (pe), layer of rods and cones (rc), outer nuclear layer (onl), outer plexiform layer (opl), inner nuclear layer (inl), inner plexiform layer (ipl), and ganglion cell layer (gcl). (F) The lack of reactivity for the C-GRIFIN-specific antibody in sections of kidney serves as a further negative control, thus excluding any antigen-independent staining. The scale bars are 25 μm.

Having herewith identified this CG as lens product, we wondered whether the proximal promoter region may harbor sequence motifs of the respective section of genes for proteins exclusively or mostly produced in the lens. To address this issue i) the promoter region of the C-GRIFIN gene was computationally scrutinized for occurrence of putative binding sites for transcription factors (compiled in Table S3), ii) the same procedure was applied to a panel of 10 genes for chicken lens proteins, i.e. aquaporin-0 [61], αA-crystallin [62], βA1-crystallin [63], βB1-crystallin [64], δ2-crystallin [65], connexin-43 [66], connexin-46 [67], connexin-50 [68], filensin [69] and phakinin [70] and iii) potentially relevant transcriptions factors could be singled out on the basis of frequency of occurrence of sequence motifs that satisfy the stringent criteria for entering the list of putative sites for binding a transcription factor, when results of processing the C-GRIFIN sequence were set into relation to those for the lens-specific genes. The obtained list, presented in Table 1, presents candidates, for whom putatively matching sequence motifs are common to all (MARE) or most selected cases (e.g. FOXP1_ES in 10 of 11 promoters). This information nourishes the assumption of shared regulatory mechanisms and provides a guideline for respective study to delineate the molecular basis of transcriptional regulation.

Table 1

Listing of transcription factors with putative reactivity to respective sequence motifs shared by promoters of genes for C-GRIFIN and 10 eye-expressed proteins[a,b]

sequence motifs present in				
11/11 promoters	**10/11 promoters (incl. GRIFIN)**	**9/11 promoters (incl. GRIFIN)**	**8/11 promoters (incl. GRIFIN)**	**7/11 promoters (incl. GRIFIN)**
MARE	FOXP1_ES	GSH1, LHX3, MAFF, PAX6	GSH2, ISL1, OLIG2, SIX1, ZIC3	ASCL2, ISL2, MAFA, NGN/NEUROD, WHN, ZIC2

[a] Full names of transcription factors: ASCL2: Achaete-scute complex homolog-like 2; FOXP1_ES: alternative splicing variant of FOXP1 activated in embryonic stem cells (ESCs); GSH1: homeobox transcription factor Gsh-1; GSH2: homeodomain transcription factor Gsh-2; ISL1: pancreatic and intestinal lim-homeodomain factor; ISL2: ISL LIM homeobox 2; LHX3: homeodomain-binding site in LIM/homeodomain factor LHX3; MAFA: lens-specific musculoaponeurotic fibrosarcoma oncogene homolog (Maf/MafA) sites; MAFF: transcription factor MafF; MARE: Maf recognition element, binding sites for homodimers of large Maf proteins; NGN/NEUROD: Neurogenin and NeuroD binding sites; OLIG2: oligodendrocyte lineage transcription factor 2; PAX6: PAX6 paired domain-binding site; SIX1: sine oculis homeobox homolog 1; WHN: winged helix protein; ZIC2: Zic family member 2; ZIC3: Zic family member 3.

[b] Listing of common names (shown in bold letters), abbreviations and synonyms of the individual chicken proteins: **aquaporin-0** (AQP0) was formerly known as MIP (major intrinsic protein)/MP26; **αA-crystallin**; **βA1-crystallin**; **βB1-crystallin**; **δ2-crystallin**/ASL2 (argininosuccinate lyase 2); **connexin-43**/GJA1 (gap junction α1 protein); **connexin-46**/GJA3 (gap junction α3 protein) was formerly known as connexin-56; **connexin-50**/GJA8 (gap junction α8 protein) was formerly known as connexin-45.6; **filensin**/BFSP1 (beaded filament structural protein 1)/CP95 (cytoskeletal protein 95); **phakinin**/BFSP2 (beaded filament structural protein 2)/CP49 (cytoskeletal protein 49).

With respect to C-GRIFIN's localization, complex formation of murine GRIFIN with α-crystallin, possible inside fiber cells and associated to the membrane, gave reason to propose "a mechanism for facilitating order and efficient packing of protein", hereby "contributing to the refractive index necessary for the lens to focus light on the retina", and an influence on "cell elongation and suture formation" [20]. This description evokes an analogy to a functional aspect of distinct plant lectins: giving structural order to the way storage proteins like vicilin are

arranged by carbohydrate-dependent and -independent recognition had been suggested as an intracellular mission of leguminous lectins, which like galectins have the β-sandwich fold [71]. Even more intriguing, dual specificity of these di- and tetravalent lectins to storage proteins and the protein body membrane suggests them to be a (reversible) glue [71-73]. Molecular interactions by the galectin, for example between cells, vesicles or glycoproteins leading to bridging and aggregate formation documented for CG-1A or human galectins-1 and -3 [74-76], will underlie these activities, explaining the interest to characterize binding features of C-GRIFIN.

3.4. Lectin activity of C-GRIFIN

As a test to reveal carbohydrate-dependent cross-linking capacity, assays for haemagglutination (formation of aggregates of erythrocytes) were performed. C-GRIFIN was active at a minimal concentration of 12 µg/ml. This value was considerably higher than those determined for CG-1A (at about 10 ng/ml) or CG-2 (at about 160 ng/ml) [26]. Cell bridging was blocked by 1.0-1.5 mM lactose but not by other tested sugars used as osmolarity control. Specificity of carbohydrate binding was further measured by solid-phase assays using (neo)glycoproteins. The glycoprotein asialofetuin with three complex-type N-glycans presenting galectin-reactive N-acetyllactosamine [74] and lactosylated albumin served as matrix for concentration-dependent and saturable binding, whereas maltose, mannose or L-fucose as sugar part were completely inert (not shown). In order to learn about the fine-specificity of C-GRIFIN, we tested the labeled lectin in an array. It included a total of about 600 different glycocompounds (mono- to oligosaccharides, glycopeptides, polysaccharides, lipopolysaccharides and glycosaminoglycans) to cover a broad spectrum of natural structures and controls (for complete listing of the panel and for obtained results, please see Table S4). Under conditions, in which the homodimeric CG-1A, -1B and -2 generated significant results, C-GRIFIN, too, gave distinct signals, corroborating its activity and specificity. Its binding activity was best with histo-blood group B tetrasaccharides (type I/II) and the histo-blood group H (type I) epitope extended by N-acetyllactosamine. In comparison to the controls, C-GRIFIN shared reactivity with CG-2 (B tetrasaccharide) and CG-1B (H type I) (please see Table S4 for details).

On the level of cells, binding, too, was dependent on the lectin concentration and inhibitable only by a certain sugar (i.e. lactose), as exemplarily shown for a CHO glycosylation mutant line (Lec2), whose N- and O-glycan sialylation is impaired (Fig. 8). As consequence,

these glycans have a high density of β-galactoside termini. N-Acetylglucosamine, L-fucose, maltose, mannose and L-rhamnose were not inhibitory, underscoring the canonical specificity to β-galactosides. Examined at the same C-GRIFIN concentration (at 10 µg/ml) under identical conditions, CHO wild-type and also other glycosylation mutant (Lec) cells were comparatively analyzed to probe into the correlation of distinct characteristics of the surface glycome and binding properties of C-GRIFIN. This special system of cell lines offers the opportunity to, for example, probe into potential preferences for N- and O-glycosylation (Lec1 mutant) and α2,3-sialylation (Lec2 mutant), when compared to the parental line [77]. The marked extent of decrease upon loss of complex-type N-glycosylation (Lec1) signals only minor reactivity to the surface O-glycans (sialylated mucin-type core 1 and O-fucosylated structures) (Fig. 9A,B), the relative increase from wild-type to Lec2 cells (reduced sialylation) a negative impact by α2,3-sialylation in wild-type cells (Fig. 9A,C) and the significant loss of signal upon reduction of galactosylation in Lec8 cells the importance of this sugar as docking site (Fig. 9A,D). In comparison to other CGs, this pattern of cell binding resembles that of CG-3 [28], whereas CG-8 has marked reactivity with α2,3-sialylated glycans on wild-type cells via its N-terminal CRD [27, 78]. The binding profile of C-GRIFIN was altered, too, upon decrease in expression of β1,4-galactosyltransferases I-VI (Lec19) (Fig. 9A,E). The ensuing consequence to incorporate less galactose moieties into the N-glycans, seen by mass spectrometry and of impact also for glycoprotein processing [79, 80], causes a mild reduction in binding relative to the parental cells. Flanking this series, a human line, strongly reactive with galectins-1 and -3 [81], also bound C-GRIFIN very well (Fig. 9F).

These results further substantiate C-GRIFIN's lectin property, here its reactivity with β-galactosides at N-glycan termini of cell surface glycans. In order to provide a first insight into the structural basis of this activity, a recently developed method based on binding of lectin to the resin, tryptic on-bead protein fragmentation and mass spectrometric identification of eluted peptides was employed [47]. Experimentally, C-GRIFIN on lactose-presenting beads was treated with trypsin, all peptides without affinity to lactose washed out, the remaining peptides eluted and identified. As shown in Fig. 10, two peptides were detected in the eluted fraction after thoroughly removing material no longer in contact to the beads. In line with the tryptic fingerprinting (Fig. 4B), they covered the sequence sections from amino acids 25-51 and 57-69 (Fig. 10). These regions are homologous to the two peptides obtained under identical conditions from human galectins-1 and -3, where they are in contact to bound lactose [47].

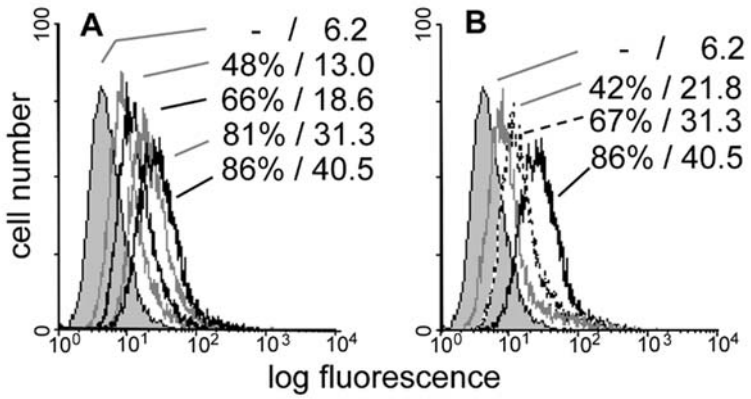


Fig. 8. Semilogarithmic representation of staining of the CHO Lec2 mutant cells with fluorescently labeled C-GRIFIN. The control value without C-GRIFIN addition is given as shaded area. Numbers characterizing staining (percentage of positive cells/mean fluorescence intensity) are given in each panel. (A) Concentration-dependent staining using C-GRIFIN at 5 µg/ml (48% / 13.0), 10 µg/ml (66% / 18.6), 20 µg/ml (81% / 31.3) and 40 µg/ml (86% / 40.5). (B) Inhibition of cell surface binding of C-GRIFIN (40 µg/ml; peak with black line at 86% / 40.5) by 10 mM (dashed line; 67% / 31.3) and 100 mM (grey line; 42% / 21.8) lactose.

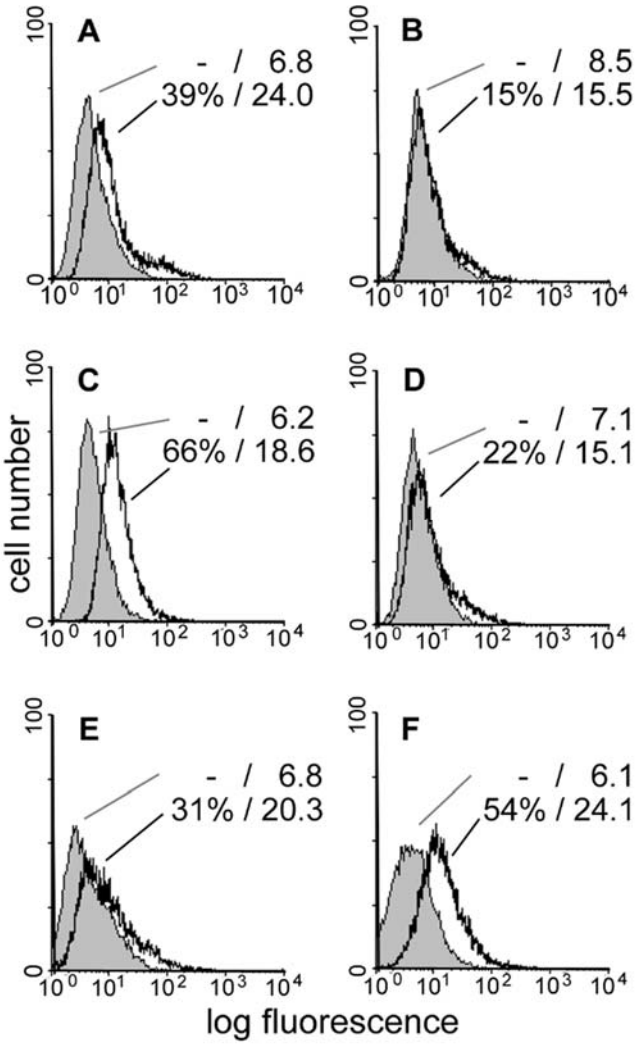


Fig. 9. Cell surface staining by C-GRIFIN (shaded area: control; black line: 10 µg/ml C-GRIFIN). The following lines were tested: (A) CHO wild-type cells, (B) CHO Lec1 mutant, (C) CHO Lec2 mutant, (D) CHO Lec8 mutant, (E) CHO Lec19 mutant and (F) human pancreatic carcinoma cells (Capan-1) expressing tumor suppressor $p16^{INK4a}$.

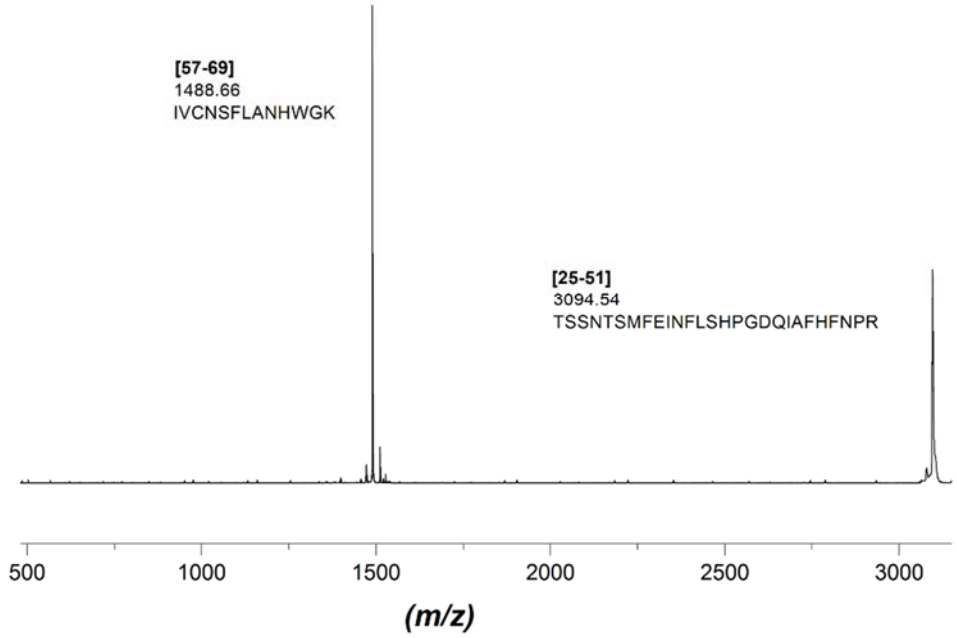


Fig. 10. Mass-spectrometric identification of two lactose-binding peptides of C-GRIFIN after tryptic fragmentation of the protein bound to the ligand on the affinity resin, thorough washing to remove all material without affinity to lactose and elution. The tryptic fingerprinting with full sequence information is given in Fig. 4.

These experiments set the stage to proceed from sequence alignments to a structural comparison in three dimensions. Due to the availability of X-ray structures for CG-1A [82], CG-1B [83], CG-2 [84] and CG-8N [85] and the lessons learned from two mutants of human galectin-1 that a seemingly minor change (i.e. C2S and R111H) can alter loop presentations and cause shifts in the topological presentation of crucial residues [83], the next step to understand C-GRIFIN's lectin activity despite the deviation from the sequence signature will be its crystallographic analysis in the presence of lactose. Equally important, the addition of C-GRIFIN to the list of CGs, with a non-cross-reactive antibody preparation being available, facilitates a complete network analysis of CG expression, including developmental stages. After all, a gradual increase of GRIFIN presence had been reported for the rat protein over the period of day 22 pc to the adult stage [18]. Here, as noted above, special attention will be given to co-monitoring of CG-3. Besides the detection of presence of mammalian galectin-3 in lens, information on a binding partner for this lectin is available. In extracts of ovine lenses, the tetraspanin MP20 was identified to associate with galectin-3 [59, 86]. MP20 is an abundant

intrinsic membrane protein present in junctions between fiber cells [87]. The bovine protein lacks common forms of glycosylation, sugar conjugation only detected as C-mannosylation at Trp, here Trp 43/61 [88] so that in this case pairing with galectin-3 may engage protein-protein association. Sequence sections in the CRD, such as the NWGR motif (BH1 domain) for binding members of the bcl-2 family of proteins [89-91], or in the N-terminal tail, such as a part of the Gly/Pro-rich repeats for involvement in interaction with components of the splicing machinery [92], underlie this functional aspect. This recognition process with a counterreceptor as part of junctions, and galectin-3 is also a part of cell-cell junctions in breast cancer cells [93], may have implications for the architecture of the lens, as the mammalian tandem-repeat-type galectin-4, originally described as 'adherens junction protein' [94], is suggested to have for assembly of adherens junctions [95].

4. Conclusion

The gene for GRIFIN is common in vertebrates without any copy-number variation. As illustrated by compiling the members of the galectin family for selected model organisms in a simplified phylogenetic tree (Fig. 11), numbers for proto- and tandem-repeat-type proteins can vary. But having a gene for GRIFIN and a gene for the chimera-type galectin-3, these two being closely related in sequence alignment, is a general feature, implying a particular functional meaning. Looking at the variability in lectin activity between mammalian and avian/fish GRIFIN, its functional significance can be further explored by cyto- and histochemically mapping binding sites of the labeled protein (and a C-GRIFIN mutant not reactive with β-galactosides) and characterizing binding partners. Of note, the zebrafish gene had also been seen expressed in adult brain, intestine and oocyte [22], and lectin activity may be relevant at these sites. In structural terms, the origin of the molecular compensation for the Arg-to-Val substitution to avoid the loss of reactivity to lactose in C-GRIFIN is presently not known. The same holds true for the relations to the other CGs, in terms of structural details beyond the amino acid sequence and in terms of tissue localization of the proteins and of binding sites detectable by labeled galectin. Because this report completes the individual characterization of all galectins in an organism, it gives chicken and its seven proteins of this family the status of a well-suited model to proceed to the first comprehensive analysis of the galectin network.

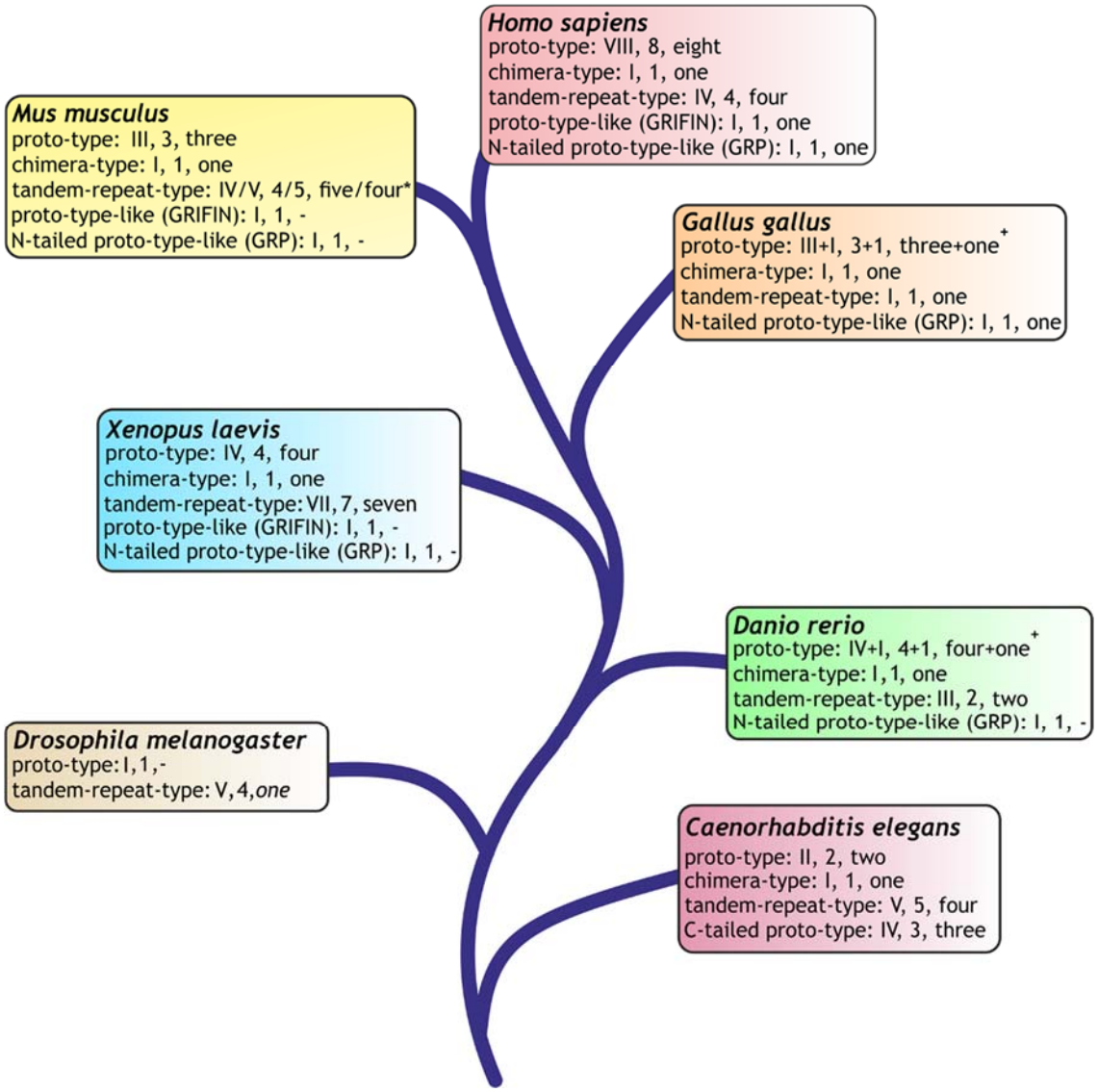


Fig. 11. Overview on the structure of the family of galectins in seven model organisms presented according to the three types of structural design and degree of homology in alignments. Evidence for presence of the gene (Roman number), the mRNA (Arabic number) and the protein (numerical information) determines the entries on numbers in each group. N-Terminal extensions of the common CRD have been described in vertebrates for GRP (36-85 amino acids), C-terminal extensions for galectins of *C. elegans* (40 amino acids for Lec-7, 41 for Lec-8, 53 for Lec-10 and 90 for Lec-11). For galectins of this organism, binding to the canonical ligand lactose has been ascertained in seven cases (Lec-1, -2, -3, -4, -6, -10, and -12 (DC2.3)), the C-tailed proto-type galectins Lec-8 and the proto-type Lec-9 showed relatively weak binding to asialofetuin (21% / 9% relative to the activity of the tandem-repeat-type Lec-1 used as standard), respective activity of Lec-5, Lec-7 and Lec-11 not yet been tested [96,97]. GRIFIN ($^{+}$) is special due to its species-dependent lectin activity, as described in this report, mouse is special due to the intra-species variation in occurrence of the tandem-repeat-type galectin-6 (*) that is found exclusively in this species [98].

Conflict of interests

Funding source was not involved in the collection, analysis and interpretation of data, in the writing of the report, and in the decision to submit the article for publication.

Acknowledgements

The generous financial support by the EC Seventh Framework Program under grant agreement no. 317297 (ITN GLYCOPHARM), the LMU excellence program and the Russian Science Foundation (14-50-00131), the contributions of glycocompounds to the array by Drs. Y. Knirel, P. Kosma and H. Kunz, inspiring discussions with Drs. B. Friday and A. Leddoz as well as valuable comments by the reviewers are gratefully acknowledged.

References

[1] W. Frazier, L. Glaser, Surface components and cell recognition, Annu. Rev. Biochem. 48 (1979) 491-523.
[2] S.H. Barondes, Lectins: their multiple endogenous cellular functions, Annu. Rev. Biochem. 50 (1981) 207-231.
[3] R. Schwartz-Albiez, Inflammation and glycosciences, in: H.-J. Gabius (Ed.), The Sugar Code. Fundamentals of glycosciences, Wiley-VCH, Weinheim, Germany, 2009, pp. 447-467.
[4] D.C. Kilpatrick, Animal lectins: a historical introduction and overview, Biochim. Biophys. Acta 1572 (2002) 187-197.
[5] H.-J. Gabius, S. André, J. Jiménez-Barbero, A. Romero, D. Solís, From lectin structure to functional glycomics: principles of the sugar code, Trends Biochem. Sci. 36 (2011) 298-313.
[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] R.L. Schnaar, Glycans and glycan-binding proteins in immune regulation: a concise introduction to glycobiology for the allergist, J. Allergy Clin. Immunol. 135 (2015) 609-615.
[8] 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.
[9] F.L. Harrison, C.J. Chesterton, Factors mediating cell-cell recognition and adhesion. Galaptins, a recently discovered class of bridging molecules, FEBS Lett. 122 (1980) 157-165.
[10] M. Caron, D. Bladier, R. Joubert, Soluble galactoside-binding vertebrate lectins: a protein family with common properties, Int. J. Biochem. 22 (1990) 1379-1385.
[11] K.-i. Kasai, J. Hirabayashi, Galectins: a family of animal lectins that decipher glycocodes, J. Biochem. 119 (1996) 1-8.

[12] S.H. Barondes, Galectins: a personal review, Trends Glycosci. Glycotechnol. 9 (1997) 1-7.
[13] 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.
[14] 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.
[15] D.N.W. Cooper, Galectinomics: finding themes in complexity, Biochim. Biophys. Acta 1572 (2002) 209-231.
[16] 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.
[17] Y.D. Lobsanov, J.M. Rini, Galectin structure, Trends Glycosci. Glycotechnol. 9 (1997) 145-154.
[18] A.T. Ogden, I. Nunes, K. Ko, S. Wu, C.S. Hines, A.F. Wang, R.S. Hegde, R.A. Lang, GRIFIN, a novel lens-specific protein related to the galectin family, J. Biol. Chem. 273 (1998) 28889-28896.
[19] Y. Ueda, C. Fukiage, M. Shih, T.R. Shearer, L.L. David, Mass measurements of C-terminally truncated α-crystallins from two-dimensional gels identify Lp82 as a major endopeptidase in rat lens, Mol. Cell. Proteomics 1 (2002) 357-365.
[20] K.A. Barton, C.D. Hsu, J.M. Petrash, Interactions between small heat shock protein α-crystallin and galectin-related interfiber protein (GRIFIN) in the ocular lens, Biochemistry 48 (2009) 3956-3966.
[21] J. Hirabayashi, K.-i. Kasai, Effect of amino acid substitution by sited-directed mutagenesis on the carbohydrate recognition and stability of human 14-kDa β-galactoside-binding lectin, J. Biol. Chem. 266 (1991) 23648-23653.
[22] H. Ahmed, G.R. Vasta, Unlike mammalian GRIFIN, the zebrafish homologue (DrGRIFIN) represents a functional carbohydrate-binding galectin, Biochem. Biophys. Res. Commun. 371 (2008) 350-355.
[23] 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.
[24] Y. Oda, K.-i. Kasai, Purification and characterization of β-galactoside-binding lectin from chick embryonic skin, Biochim. Biophys. Acta 761 (1983) 237-245.
[25] S.L. Crittenden, C.F. Roff, J.L. Wang, Carbohydrate-binding protein 35: identification of the galactose-specific lectin in various tissues of mice, Mol. Cell. Biol. 4 (1984) 1252-1259.
[26] 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.
[27] 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.
[28] 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.

[29] A. Göhler, C. Buchner, S. Doose, H. Kaltner, H.-J. Gabius, Analysis of homodimeric avian and human galectins by two methods based on fluorescence spectroscopy: different structural alterations upon oxidation and ligand binding, Biochimie 94 (2012) 2649-2655.

[30] H. Kaltner, T. Singh, J.C. Manning, A.-S. Raschta, S. André, F. Sinowatz, H.-J. Gabius, Network monitoring of adhesion/growth-regulatory galectins: localization of the five canonical chicken proteins in embryonic and maturing bone and cartilage and their introduction as histochemical tools, Anat. Rec. 298 (2015) 2051-2070.

[31] 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.

[32] 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.

[33] D. Maglott, J. Ostell, K.D. Pruitt, T. Tatusova, Entrez Gene: gene-centered information at NCBI, Nucleic Acids Res. 39 (2011) D52-57.

[34] M. Krzeminski, T. Singh, S. André, M. Lensch, A.M. Wu, A.M.J.J. Bonvin, H.-J. Gabius, Human galectin-3 (Mac-2 antigen): defining molecular switches of affinity to natural glycoproteins, structural and dynamic aspects of glycan binding by flexible ligand docking and putative regulatory sequences in the proximal promoter region, Biochim. Biophys. Acta 1810 (2011) 150-161.

[35] H.-J. Gabius, Influence of type of linkage and spacer on the interaction of □-galactoside-binding proteins with immobilized affinity ligands, Anal. Biochem. 189 (1990) 91-94.

[36] 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.

[37] 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.

[38] K. Diepold, K. Bomans, M. Wiedmann, B. Zimmermann, A. Petzold, T. Schlothauer, R. Mueller, B. Moritz, J.O. Stracke, M. Molhoj, D. Reusch, P. Bulau, Simultaneous assessment of Asp isomerization and Asn deamidation in recombinant antibodies by LC-MS following incubation at elevated temperatures, PLoS One 7 (2012) e30295.

[39] E. Folzer, K. Diepold, K. Bomans, C. Finkler, R. Schmidt, P. Bulau, J. Huwyler, H.C. Mahler, A.V. Koulov, Selective oxidation of methionine and tryptophan residues in a therapeutic IgG1 molecule, J. Pharm. Sci. 104 (2015) 2824-2831.

[40] M. Haberger, A.K. Heidenreich, T. Schlothauer, M. Hook, J. Gassner, K. Bomans, M. Yegres, A. Zwick, B. Zimmermann, H. Wegele, L. Bonnington, D. Reusch, P. Bulau, Functional assessment of antibody oxidation by native mass spectrometry, MAbs 7 (2015) 891-900.

[41] H. Kaltner, K. Seyrek, A. Heck, F. Sinowatz, H.-J. Gabius, Galectin-1 and galectin-3 in fetal development of bovine respiratory and digestive tracts. Comparison of cell type-specific expression profiles and subcellular localization, Cell Tissue Res. 307 (2002) 35-46.

[42] H.-C. Siebert, J. Rosen, K. Seyrek, H. Kaltner, S. André, N.V. Bovin, P.-G. Nyholm, F. Sinowatz, H.-J. Gabius, α2,3/α2,6-Sialylation of *N*-glycans: non-synonymous signals with marked developmental regulation in bovine reproductive tracts, Biochimie 88 (2006) 399-410.

[43] H.-J. Gabius, R. Engelhardt, F. Cramer, R. Bätge, G.A. Nagel, Pattern of endogenous lectins in a human epithelial tumor, Cancer Res. 45 (1985) 253-257.

[44] 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.

[45] 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. USA 101 (2004) 17033-17038.

[46] 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.

[47] A. Moise, S. André, F. Eggers, M. Krzeminski, M. Przybylski, H.-J. Gabius, Toward bioinspired galectin mimetics: identification of ligand-contacting peptides by proteolytic-excision mass spectrometry, J. Am. Chem. Soc. 133 (2011) 14844-14847.

[48] M.F. Simpanya, G. Wistow, J. Gao, L.L. David, F.J. Giblin, K.P. Mitton, Expressed sequence tag analysis of guinea pig (*Cavia porcellus*) eye tissues for NEIBank, Mol. Vis. 14 (2008) 2413-2427.

[49] P.A. Poland, C. Rondanino, C.L. Kinlough, J. Heimburg-Molinaro, C.M. Arthur, S.R. Stowell, D.F. Smith, R.P. Hughey, Identification and characterization of endogenous galectins expressed in Madin Darby canine kidney cells, J. Biol. Chem. 286 (2011) 6780-6790.

[50] Y. Sakakura, J. Hirabayashi, Y. Oda, Y. Ohyama, K.-i. Kasai, Structure of chicken 16-kDa β-galactoside-binding lectin. Complete amino acid sequence, cloning of cDNA and production of recombinant lectin, J. Biol. Chem. 265 (1990) 21573-21579.

[51] E.M. Rapoport, V.K. Matveeva, H. Kaltner, S. André, O.A. Vokhmyanina, G.V. Pazynina, V.V. Severov, I.M. Ryzhov, E.Y. Korchagina, I.M. Belyanchikov, H.-J. Gabius, N.V. Bovin, Comparative lectinology: delineating glycan-specificity profiles of the chicken galectins using neoglycoconjugates in a cell assay, Glycobiology 25 (2015) 726-734.

[52] G. Levi, V.I. Teichberg, Isolation and physicochemical characterization of electrolectin, a β-*D*-galactoside-binding lectin from the electric organ of *Electrophorus electricus*, J. Biol. Chem. 256 (1981) 5735-5740.

[53] 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.

[54] S. Barcelo-Batllori, H. Corominola, M. Claret, I. Canals, J. Guinovart, R. Gomis, Target identification of the novel antiobesity agent tungstate in adipose tissue from obese rats, Proteomics 5 (2005) 4927-4935.

[55] C. D'Anna, C. Cascio, D. Cigna, G. Galizzi, I. Deidda, L. Bianchi, D. Russo, R. Passantino, L. Bini, P. Guarneri, A retinal proteomics-based study identifies αA-crystallin as a sex steroid-regulated protein, Proteomics 11 (2011) 986-990.

[56] L.F. Castagna, C.A. Landa, Isolation and characterization of a soluble lactose-binding lectin from postnatal chicken retina, J. Neurosci. Res. 37 (1994) 750-758.

[57] C.A. Maldonado, L.F. Castagna, G.A. Rabinovich, C.A. Landa, Immunocytochemical study of the distribution of a 16-kDa galectin in the chicken retina, Invest. Ophthalmol. Vis. Sci. 40 (1999) 2971-2977.
[58] U. Schlötzer-Schrehardt, S. André, C. Janko, H. Kaltner, J. Kopitz, H.-J. Gabius, M. Herrmann, Adhesion/growth-regulatory galectins in the human eye: localization profiles and tissue reactivities as a standard to detect disease-associated alterations, Graefes Arch. Clin. Exp. Ophthalmol. 250 (2012) 1169-1180.
[59] T. Gonen, P. Donaldson, J. Kistler, Galectin-3 is associated with the plasma membrane of lens fiber cells, Invest. Ophthalmol. Vis. Sci. 41 (2000) 199-203.
[60] R. Dahm, S. Branmke, J. Dawczynski, R.H. Nagaraj, M. Kasper, Developmental aspects of galectin expression in the lens, Histochem. Cell Biol. 119 (2003) 219-226.
[61] J.R. Kuszak, J. Alcala, H. Maisel, Biochemical and structural features of chick lens gap junctions, Exp. Eye Res. 33 (1981) 157-166.
[62] M.A. Thompson, J.W. Hawkins, J. Piatigorsky, Complete nucleotide sequence of the chicken αA-crystallin gene and its 5' flanking region, Gene 56 (1987) 173-184.
[63] J.B. McDermott, C.A. Peterson, J. Piatigorsky, Structure and lens expression of the gene encoding chicken βA3/A1-crystallin, Gene 117 (1992) 193-200.
[64] M.K. Duncan, H.J. Roth, M. Thompson, M. Kantorow, J. Piatigorsky, Chicken βB1 crystallin: gene sequence and evidence for functional conservation of promoter activity between chicken and mouse, Biochim. Biophys. Acta 1261 (1995) 68-76.
[65] S.P. Bhat, R.E. Jones, M.A. Sullivan, J. Piatigorsky, Chicken lens crystallin DNA sequences show at least two δ-crystallin genes, Nature 284 (1980) 234-238.
[66] L.S. Musil, E.C. Beyer, D.A. Goodenough, Expression of the gap junction protein connexin43 in embryonic chick lens: molecular cloning, ultrastructural localization, and post-translational phosphorylation, J. Membr. Biol. 116 (1990) 163-175.
[67] D.M. Rup, R.D. Veenstra, H.Z. Wang, P.R. Brink, E.C. Beyer, Chick connexin-56, a novel lens gap junction protein. Molecular cloning and functional expression, J. Biol. Chem. 268 (1993) 706-712.
[68] J.X. Jiang, T.W. White, D.A. Goodenough, D.L. Paul, Molecular cloning and functional characterization of chick lens fiber connexin 45.6, Mol. Biol. Cell 5 (1994) 363-373.
[69] S.G. Remington, Chicken filensin: a lens fiber cell protein that exhibits sequence similarity to intermediate filament proteins, J. Cell Sci. 105 (1993) 1057-1068.
[70] M. Ireland, H. Maisel, A cytoskeletal protein unique to lens fiber cell differentiation, Exp. Eye Res. 38 (1984) 637-645.
[71] H. Rüdiger, H.-J. Gabius, Plant lectins: occurrence, biochemistry, functions and applications, Glycoconjugate J. 18 (2001) 589-613.
[72] M. Wenzel, H. Gers-Barlag, A. Schimpl, H. Rüdiger, Time course of lectin and storage protein biosynthesis in developing pea (*Pisum sativum*) seeds, Biol. Chem. Hoppe-Seyler 374 (1993) 887-894.
[73] G. Schecher, H. Rüdiger, Interaction of the soybean (*Glycine max*) seed lectin with components of the soybean protein body membrane, Biol. Chem. Hoppe-Seyler 375 (1994) 829-832.
[74] D. Gupta, H. Kaltner, X. Dong, H.-J. Gabius, C.F. Brewer, Comparative cross-linking activities of lactose-specific plant and animal lectins and a natural lactose-binding immunoglobulin G fraction from human serum with asialofetuin, Glycobiology 6 (1996) 843-849.
[75] V. Percec, P. Leowanawat, H.J. Sun, O. Kulikov, C.D. Nusbaum, T.M. Tran, A. Bertin, D.A. Wilson, M. Peterca, S. Zhang, N.P. Kamat, K. Vargo, D. Moock, E.D. Johnston, D.A. Hammer, D.J. Pochan, Y. Chen, Y.M. Chabre, T.C. Shiao, M. Bergeron-Brlek, S. André, R. Roy, H.-J. Gabius, P.A. Heiney, Modular synthesis of amphiphilic Janus glycodendrimers and their self-assembly into glycodendrimersomes and other complex

architectures with bioactivity to biomedically relevant lectins, J. Am. Chem. Soc. 135 (2013) 9055-9077.
[76] S. Zhang, R.-O. Moussodia, C. Murzeau, H.J. Sun, M.L. Klein, S. Vértesy, S. André, R. Roy, H.-J. Gabius, V. Percec, Dissecting molecular aspects of cell interactions using glycodendrimersomes with programmable glycan presentation and engineered human lectins, Angew. Chem. Int. Ed. 54 (2015) 4036-4040.
[77] S.K. Patnaik, P. Stanley, Lectin-resistant CHO glycosylation mutants, Methods Enzymol. 416 (2006) 159-182.
[78] 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.
[79] J. Lee, S.H. Park, S. Sundaram, T.S. Raju, N.L. Shaper, P. Stanley, A mutation causing a reduced level of expression of six β4-galactosyltransferase genes is the basis of the Lec19 CHO glycosylation mutant, Biochemistry 42 (2003) 12349-12357.
[80] 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.
[81] 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.
[82] 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.
[83] 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.
[84] 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.
[85] F.M. Ruiz, B.A. Scholz, E. Buzamet, J. Kopitz, S. André, M. Menendez, 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.
[86] T. Gonen, A.C. Grey, M.D. Jacobs, P.J. Donaldson, J. Kistler, MP20, the second most abundant lens membrane protein and member of the tetraspanin superfamily, joins the list of ligands of galectin-3, BMC Cell Biol. 2 (2001) 17.
[87] C.F. Louis, K.C. Hur, A.C. Galvan, E.M. TenBroek, L.J. Jarvis, E.D. Eccleston, J.B. Howard, Identification of an 18,000-dalton protein in mammalian lens fiber cell membranes, J. Biol. Chem. 264 (1989) 19967-19973.
[88] L.A. Ervin, L.E. Ball, R.K. Crouch, K.L. Schey, Phosphorylation and glycosylation of bovine lens MP20, Invest. Ophthalmol. Vis. Sci. 46 (2005) 627-635.
[89] R.Y. Yang, D.K. Hsu, F.-T. Liu, Expression of galectin-3 modulates T-cell growth and apoptosis, Proc. Natl. Acad. Sci. USA 93 (1996) 6737-6742.
[90] S. Akahani, P. Nangia-Makker, H. Inohara, H.R. Kim, A. Raz, Galectin-3: a novel antiapoptotic molecule with a functional BH1 (NWGR) domain of Bcl-2 family, Cancer Res. 57 (1997) 5272-5276.

[91] 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.

[92] R.M. Gray, M.J. Davis, K.M. Ruby, P.G. Voss, R.J. Patterson, J.L. Wang, Distinct effects on splicing of two monoclonal antibodies directed against the amino-terminal domain of galectin-3, Arch. Biochem. Biophys. 475 (2008) 100-108.

[93] C. Boscher, Y.Z. Zheng, R. Lakshminarayan, L. Johannes, J.W. Dennis, L.J. Foster, I.R. Nabi, Galectin-3 protein regulates mobility of N-cadherin and GM1 ganglioside at cell-cell junctions of mammary carcinoma cells, J. Biol. Chem. 287 (2012) 32940-32952.

[94] M.L. Chiu, J.C. Jones, E.J. O'Keefe, Restricted tissue distribution of a 37-kD possible adherens junction protein, J. Cell Biol. 119 (1992) 1689-1700.

[95] M.L. Chiu, D.A.D. Parry, S.R. Feldman, D.G. Klapper, E.J. O'Keefe, An adherens junction protein is a member of the family of lactose-binding lectins, J. Biol. Chem. 269 (1994) 31770-31776.

[96] Y. Nemoto-Sasaki, K. Hayama, H. Ohya, Y. Arata, M.K. Kaneko, N. Saitou, J. Hirabayashi, K.-i. Kasai, *Caenorhabditis elegans* galectins LEC-1–LEC-11: structural features and sugar-binding properties, Biochim. Biophys. Acta 1780 (2008) 1131-1142.

[97] Y. Nemoto-Sasaki, S. Takai, T. Takeuchi, Y. Arata, K. Nishiyama, A. Yamada, H. Takahashi, H. Natsugari, K.-i. Kasai, The *DC2.3* gene in *Caenorhabditis elegans* encodes a galectin that recognizes the galactose β1→4 fucose disaccharide unit, Biol. Pharm. Bull. 34 (2011) 1635-1639.

[98] D. Houzelstein, I.R. Goncalves, A. Orth, F. Bonhomme, P. Netter, *Lgals6*, a 2-million-year-old gene in mice: a case of positive Darwinian selection and presence/absence polymorphism, Genetics 178 (2008) 1533-1545.

4.2 Publication 2

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,*], Andrea Flores-Ibarra[b,*], Malwina Michalak[c,*], 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*

[*]Contributed equally

[#]Corresponding author: Tel.: +49(0)89-2180-2290; FAX: +49(0)89-2180-992290, e-mail address: gabius@tiph.vetmed.uni-muenchen.de or gabius@lectins.de

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 among this family of six proteins due to absence of binding lactose. The documented high degree of sequence conservation among vertebrate orthologues confers the status of a model for delineating an assumed 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.

Keywords: adhesion · crystallography · lectin · phylogenesis · proliferation

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 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 was 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´ (*BamHI* restriction site underlined) for full-length C-GRP or 5´-CTGGGATCCTGCGGGCACATCAAAGGAGGG-3´ (*BamHI* restriction site underlined) for C-GRP-C, in both cases using the antisense primer 5´-

CGTACGGTCGACTCAGCCAAGTTTTGTAAGCTGAAG-3´ (*SalI* 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 *BamHI*/*SalI* endonucleases. Insertions into *BamHI/SalI*-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 *BamHI* 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 transferred 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 were 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^{TM} 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 mL 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 642 glycocompounds 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-hydroxysuccimide ester derivative (Sigma, Munich, 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 10220 $M^{-1}cm^{-1}$ (http://web.expasy.org/protparam/).

Systematic screening for conditions to crystallize C-GRP(-C) were 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 (4,882)
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 (2,285)
Testing reflections	1,593 (101)
Protein atoms (non H)	1,064
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]$R_{merge} = S_{hkl}S_i\ |I_i(hkl) - \langle I(hkl)\rangle| / S_{hkl}S_iI_i(hkl)$, [b]$R_{meas} = \Sigma_{hkl}\ (N-1)^{-1/2}\ \Sigma_i|I_i(hkl) - \langle I(hkl)\rangle| / \Sigma_{hkl}\ \Sigma_i\ I_i(hkl)$, 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 [83]

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.

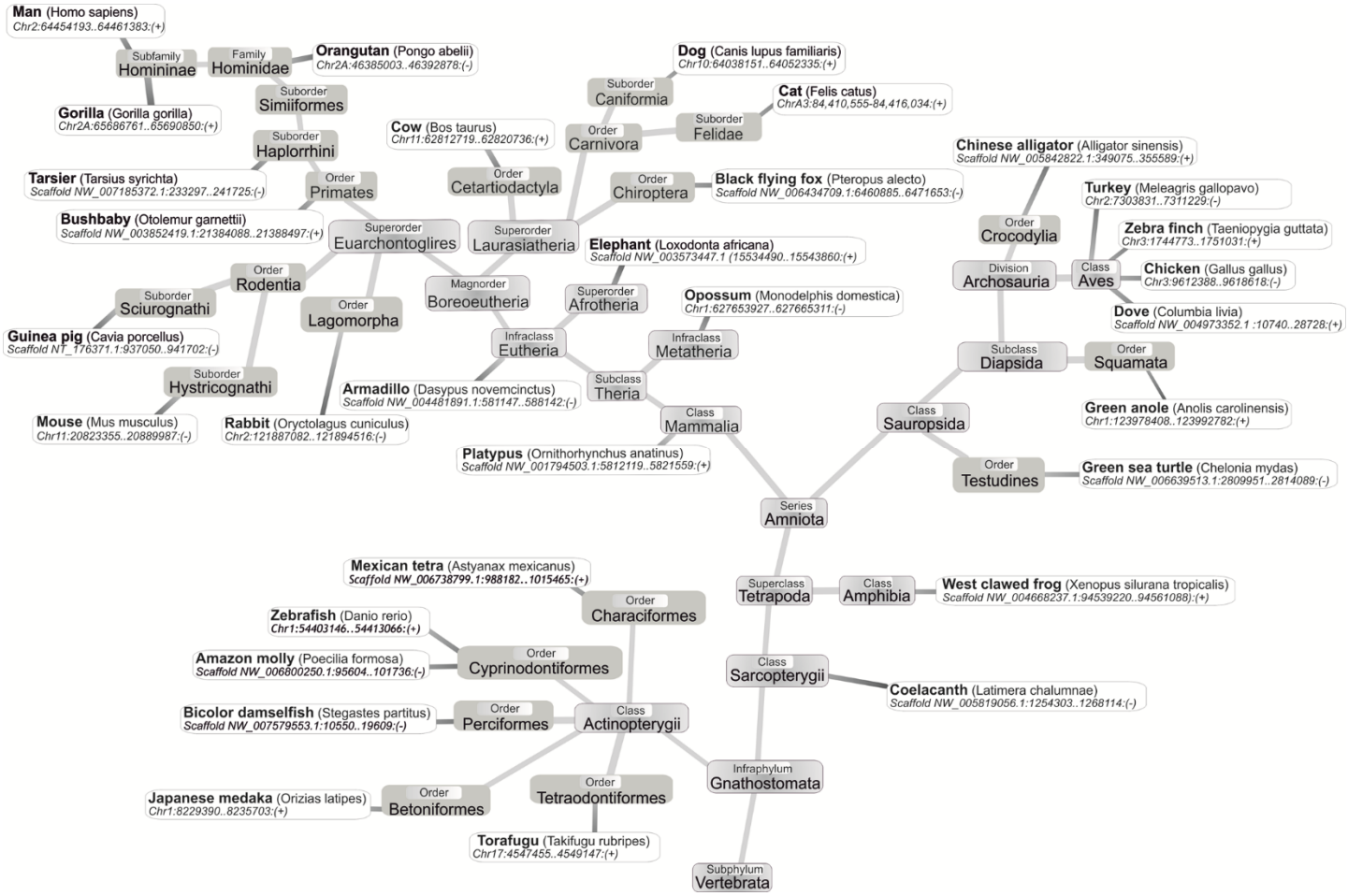


Fig. 1. Compilation of information on GRP gene in man and 31 vertebrate species from different branches of the taxonomic tree available in databases. Database-retrieved information on the genes, present either in forward (+) or in reverse (-) direction, is listed as chromosome or scaffold number together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

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 proto-type 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 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.

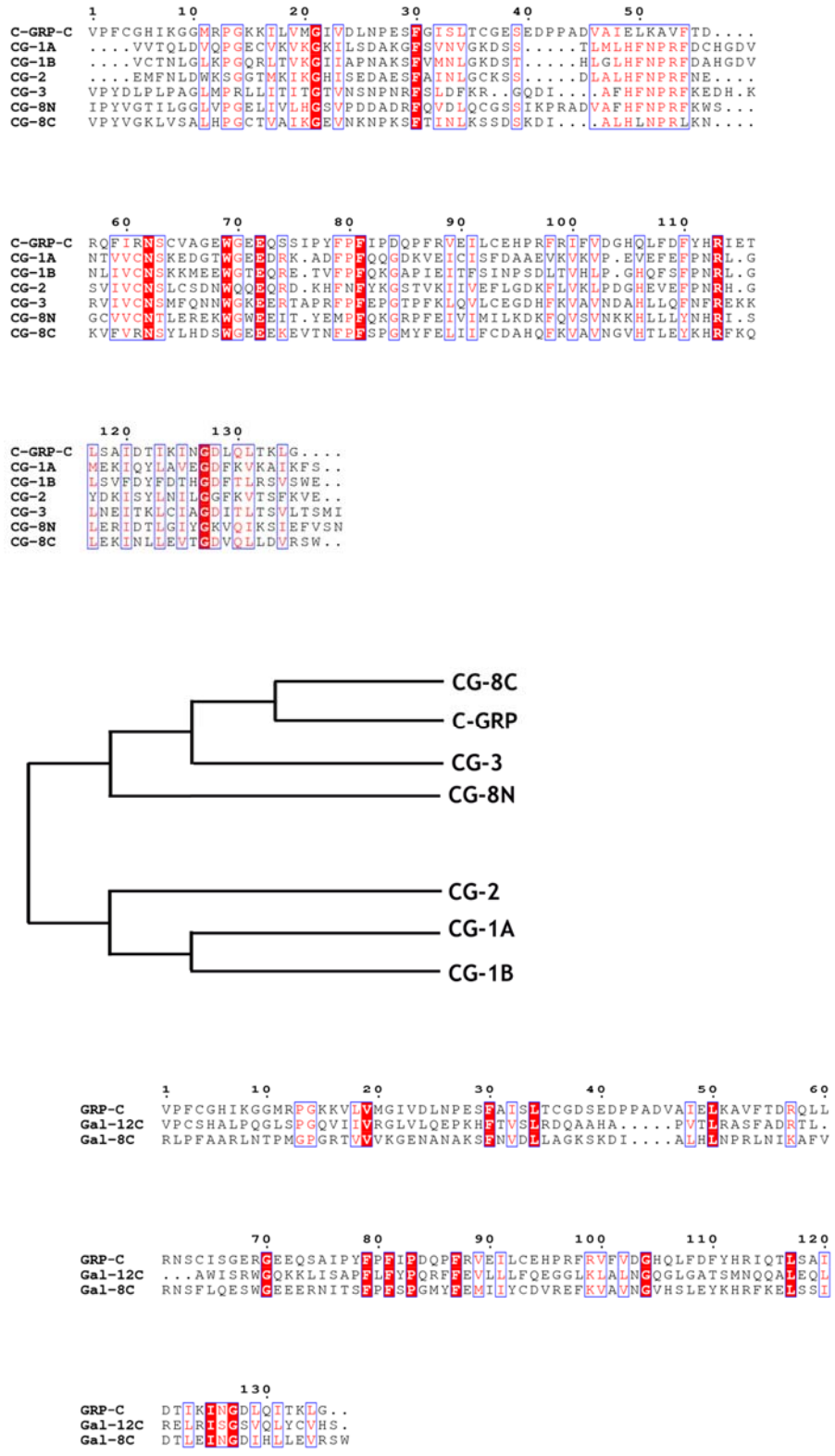


Fig. 2. Comparison of amino acid sequences of the CRDs for the shortened version of C-GRP, i.e. C-GRP-C (amino acid 37 is set to position 1 for alignment), versus the five canonical CGs (top), the respective phylogenetic family-tree diagram (middle) and the alignment of the amino acid sequence of the 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.

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).

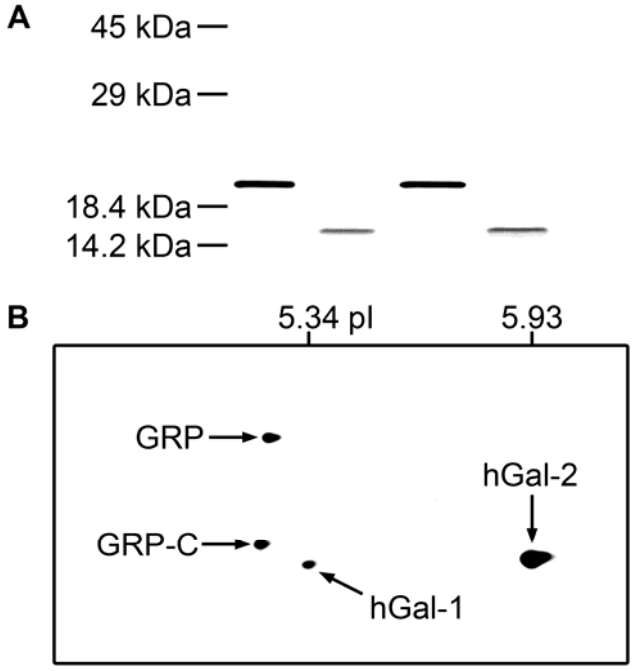


Fig. 3. Documentation of one- and two-dimensional gel electrophoretic analysis of recombinant C-GRP and C-GRP-C. (A) Electrophoretic mobility of purified C-GRP and C-GRP-C in one-dimensional polyacrylamide gel electrophoresis under denaturing conditions in the presence of 2-mercaptoethanol in a 15% running 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).

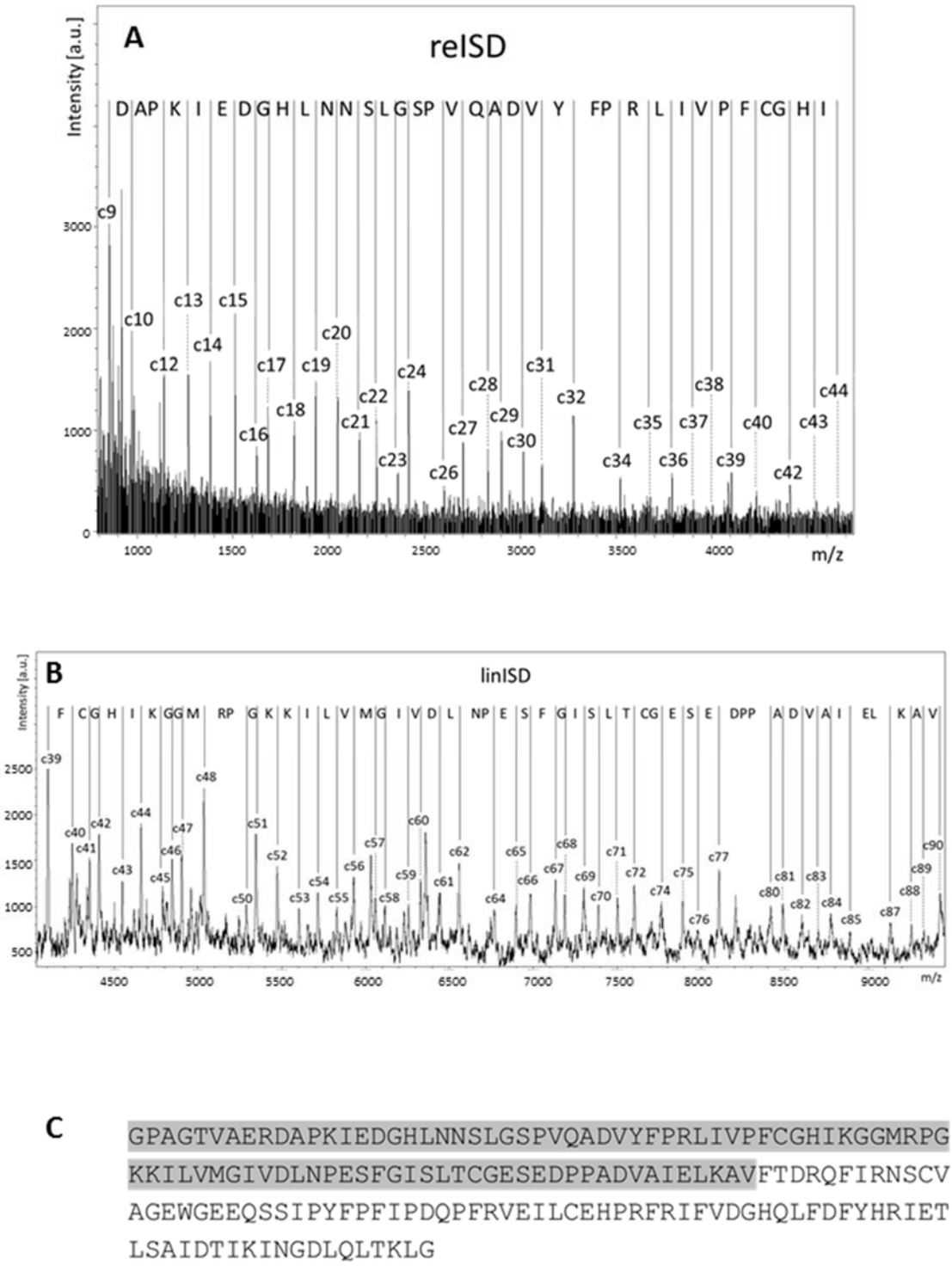


Fig. 4. Sequencing of the N-terminus of C-GRP by reISD and linISD. (A) reISD mass spectrum of C-GRP shows the smallest detectable peptide at the c9/c10 and the ensuing stepwise peptide ladder. (B) linISD spectrum shows the smallest detectable peptide at c39/c40 and the ensuing stepwise peptide ladder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

To 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.

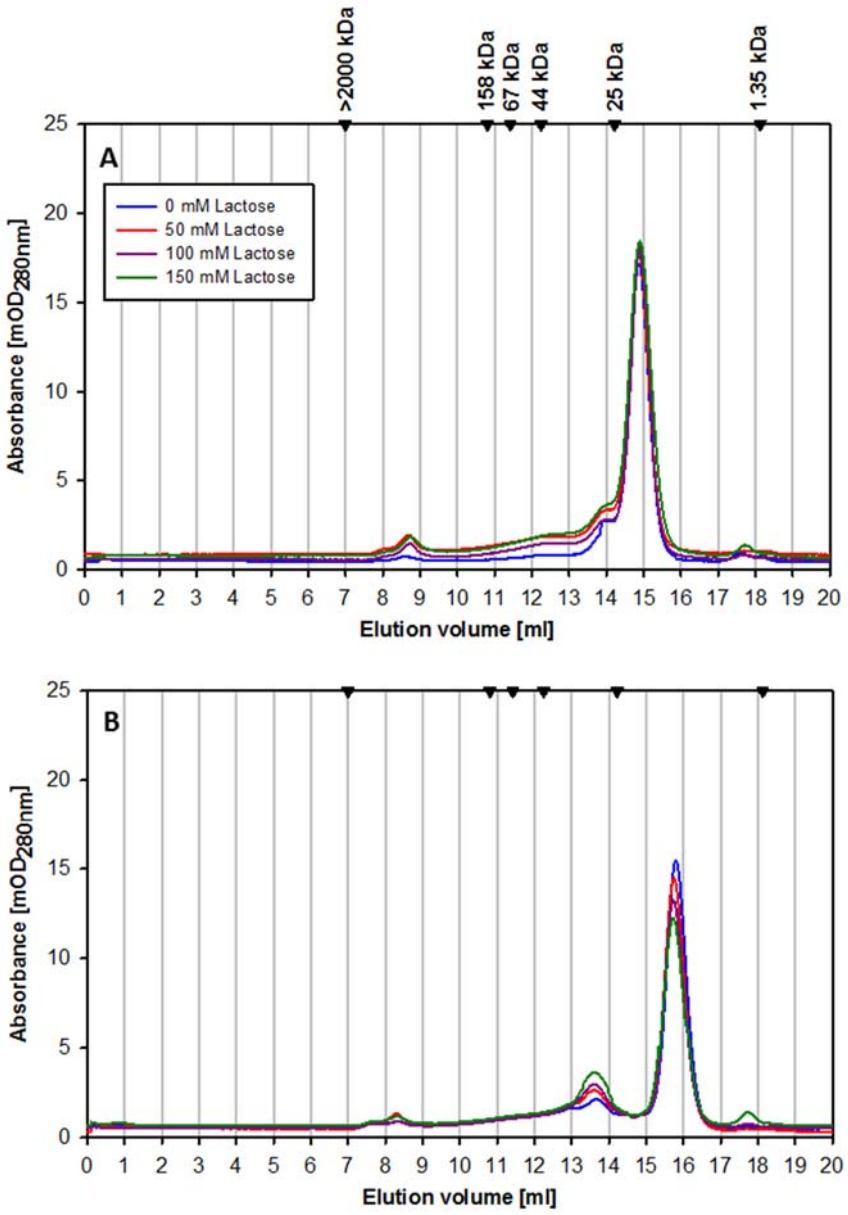


Fig. 5. Gel filtration analysis of C-GRP and C-GRP-C. (A) Elution profiles of the physiological protein (C-GRP with molecular mass of 19069 Da; please see Fig. S5A for MALDI TOF MS-based information) and (B) its shortened version (C-GRP-C with molecular mass of 15339 Da; please see Fig. S6A for respective 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.

Considering its monomeric status in solution and its sequence addition to the CRD, C-GRP resembles rat galectin-5 from the group of proto-type 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 fluosrescein 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 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-inhhibitable 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.

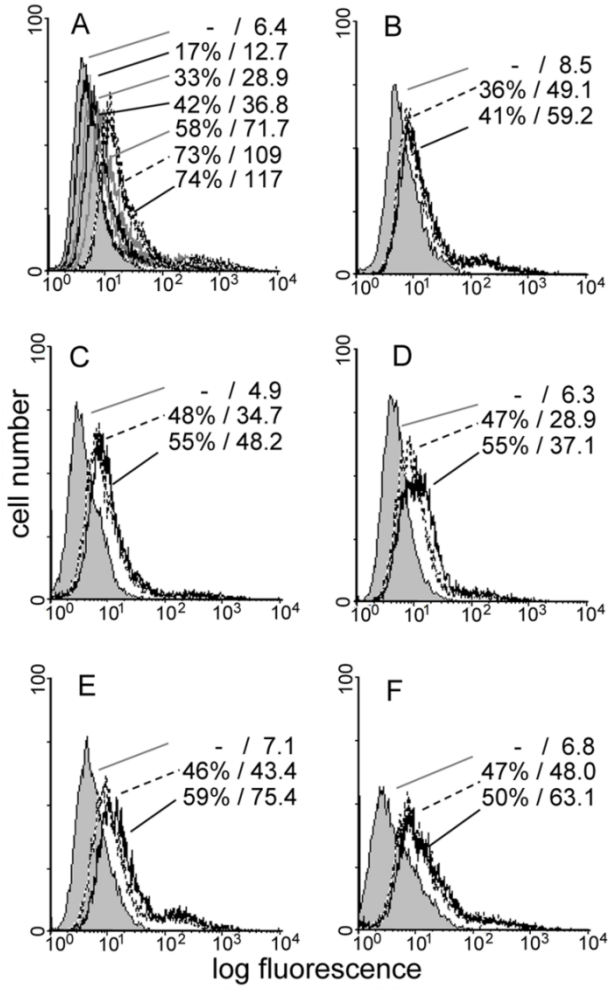


Fig. 6. Semilogarithmic representation of fluorescent cell surface staining by labeled C-GRP. Quantitative data on percentage of positive cells (%) and mean fluorescence intensity are given for each curve in each panel. The control value (background) is drawn as grey-shaded area. (A) Staining of parental 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 overexpressing α2,6-sialyltransferase I (C), Lec2 (decreased level of sialylation) (D), Lec8 (decreased level of galactosylation) (E) and Lec19 (decreased level of galactosylation) (F).

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.

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 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 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].

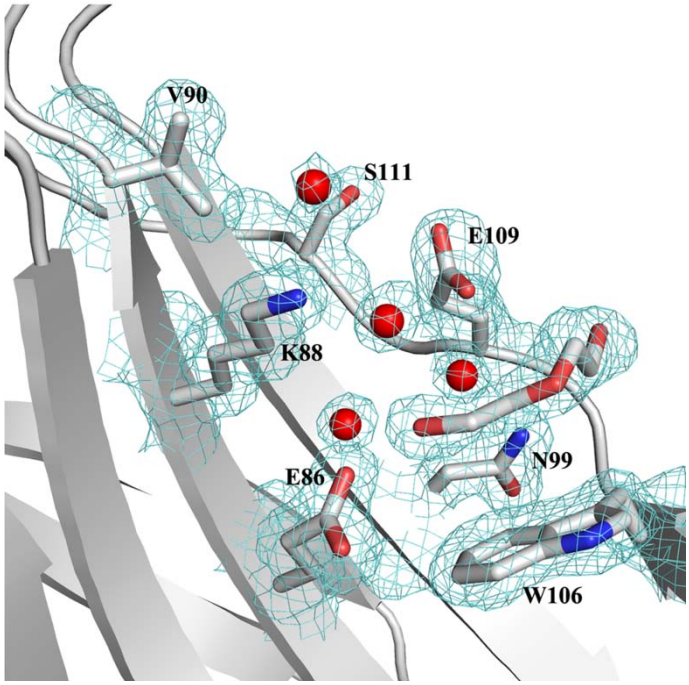


Fig. 7. Close-up view of the putative contact site for a β-galactoside in C-GRP-C. Respective residues in C-GRP-C and an ethylene glycol molecule are show in ball-and-stick mode, water molecules are depicted as red spheres. The observed electron density map 2Fo-Fc is contoured at 1.0 σ.

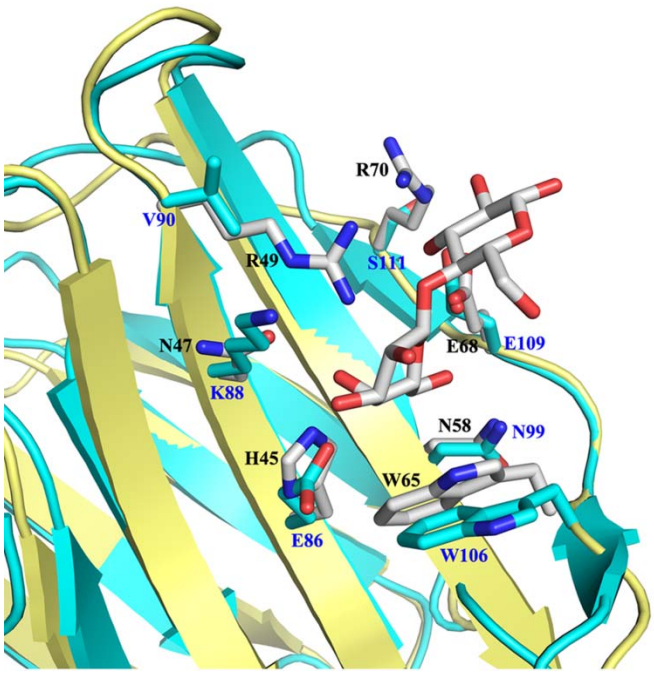


Fig. 8 Superposition of the structure of the contact site for a β-galactoside in CG-2 with the structurally equipositioned of C-GRP-C shown in Fig. 7. Amino acids of CG-2 are colored in yellow and those of C-GRP-C in cyan. Residues making contact to lactose of CG-2 and those of equivalent positions of 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.

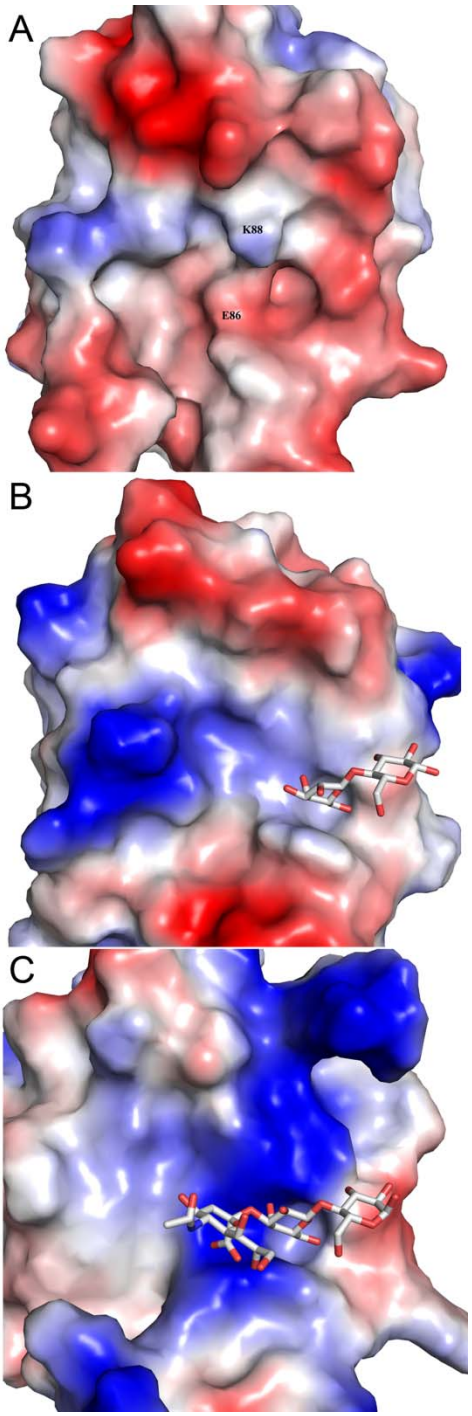


Fig. 9. Electrostatic surface potential maps of C-GRP-C (A), CG-2 (B) and CG-8N (C), contoured from -10 kT/e (intense red) to +10 kT/e (intense blue). Differences in distribution of acidic residues between C-GRP-C and the two canonical CRDs as well as absence of a tunnel-like cavity present in CG-2 and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

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].

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.

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), 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) 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) D52-57.
[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. USA 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. USA 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. Kwon do, 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. USA 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] P.A. Karplus, K. Diederichs, Linking crystallographic model and data quality, Science 336 (2012) 1030-1033.

4.3 Publication 3

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

Herbert Kaltner[a], Gabriel García Caballero[a], Fred Sinowatz[b], Sebastian Schmidt[a], Joachim C. Manning[a], Sabine André[a], Hans-Joachim Gabius[a,#]

[a]*Institute of Physiological Chemistry and* [b]*Institute of Anatomy, Histology and Embryology, Faculty of Veterinary Medicine, Ludwig-Maximilians-University Munich, Veterinärstr. 13, 80539 Munich, Germany*

[#]Corresponding author: Tel.: +49-(0)89-2180-2290; FAX: +49-(0)89-2180-992290

e-mail address: gabius@tiph.vetmed.uni-muenchen.de or gabius@lectins.de

ABSTRACT

Background: Galectin-related protein (GRP), present in vertebrates, is special within this family of adhesion/growth-regulatory proteins due to its strong positive selection and loss of canonical lectin activity.

Methods: RT-PCR and Western blotting together with flow cytofluorimetry and immunocyto- and histochemistry monitor expression and localization of chicken GRP. The promoter sequence of the GRP gene is processed computationally to detect putative sites for binding transcription factors. The labeled protein is applied as probe to detect binding sites on cells and in sections, along with glycocompounds to test inhibition of the association.

Results: Expression of GRP in chicken is limited to bursa of Fabricius, immunohistochemically found in B cells, also in bursal epithelium and vessels. Presence in B cells is shared with only one canonical galectin, i.e. CG-8. Binding to a chicken lymphoma line was specific and saturable, not affected by lactose but completely blocked by heparin, as also seen in sections.

Conclusions: Expression monitoring initiated for GRP reveals a distinct site of localization in chicken, much more restricted than for any of its canonical galectins.

Keywords: B cells · bursa · lectin · lymphocytes · promoter

1. Introduction

Their increasingly emerging physiological significance justifies efforts to detect, to structurally characterize and correspondingly classify animal and human lectins [1,2]. The resulting formation of families gives direction to aim first at accomplishing a complete listing of the members of each family on the genomic level, next at defining biochemical properties of each protein. Measuring their expression and localization profiles and then performing a detailed network analysis are the following steps. This work will relate the evolutionary course of gene/sequence diversifications to characteristics of the individual homologous proteins from their structure to their cell/tissue presence and localization. When applying this concept to the adhesion/growth-regulatory galectins, to bring the status of knowledge for the galectin-related protein (GRP), an integral part of the phylogeny of this class of lectins [3,4], to the same level as already attained for canonical proteins was an obvious necessity. In this case, the indications for an exceptionally high level of sequence conservation among vertebrates [3], the loss of binding to the common ligand lactose for human GRP [5,6] and the general status of the five canonical chicken galectins (CGs) as model suited for comprehensive network analysis of general relevance [7] had prompted detailed biochemical characterization of chicken GRP (C-GRP) [8]. Except for the occurrence of C-GRP-specific mRNA in bursal lymphocytes [3], no information on expression of this gene was available for this sixth member of the galectin family in chicken or for any other vertebrate.

In this report, monitoring of C-GRP expression in 4-week-old chickens is documented on the level of mRNA (RT-PCR) and of protein (Western blotting) as well as of its localization in tissue sections and isolated/cultured cells. Using the full set of non-cross-reactive anti-CG antibodies, C-GRP positivity is set into relation to the corresponding properties for each canonical CG. A computational inspection of the promoter region flanked the experimental work to track down putative regulatory sites for a tissue-specific expression. In addition to the immunocyto- and histochemical monitoring, labeled C-GRP, along with the labeled forms of the other family members, is introduced as tool to detect sites in the C-GRP-positive tissue/cells, which are reactive with this protein. The combined results of the biochemical [8] and cyto- and immunohistochemical [this study] investigations shape the notion for C-GRP to be an acquisition to the galectin family with strong positive selection, a unique binding capacity and its own, very particular profile of expression. The documented data initiate GRP localization and detection of cellular binding sites as a step toward functional characterization. Equally important, the addition of C-GRP to the network of canonical CGs is assumed to have relevance

beyond this model organism, prompting detailed comparative analysis for GRP among vertebrates.

2. Materials and methods

2.1. Materials

The monoclonal antibody AV20 against the chicken B cell marker chB6 [9,10] was kindly provided by S. Härtle (Chair of Animal Physiology, Faculty of Veterinary Medicine, Ludwig-Maximilians-University, Munich, Germany), FITC-labeled anti-mouse IgG1 was a generous gift from H. Ammer (Institute of Veterinary Pharmacology and Toxicology, Faculty of Veterinary Medicine, Ludwig-Maximilians-University, Munich, Germany). CGs and C-GRP, which were purified after recombinant production using affinity chromatography on lactosylated Sepharose 4B as crucial step (CGs) or as fusion protein (C-GRP) as described [8,11,12], were labeled under activity-preserving conditions with the N-hydroxysuccinimidyl derivative of biotin, fluorescein isothiocyanate or the succinimidyl ester derivative of the Alexa Fluor® 488 dye (Invitrogen, Darmstadt, Germany), as described previously [13]. The fungal (*Polyporus squamosus*) lectin (PSL) was purified, controlled for purity and labeled as described [14-16]. Extent of labeling was routinely determined by mass spectrometry [17], labeled lectins were rigorously checked for activity and maintained specificity by haemagglutination (CGs) and by binding assays using asialofetuin as matrix or mammalian cells, especially Chinese hamster ovary (CHO) parental cells and mutant lines altered in distinct aspects of glycosylation (for details, please see [8]). The solid-phase assay for assessment of binding of surface-presented C-GRP (using 0.1 µg/mL, 0.25 µg/mL and 0.5 µg/mL for coating) to biotinylated CGs (titrations with 5-40 µg/mL) was performed under identical conditions. Biotinylated plant lectins (MAA-I, *Maackia amurensis* agglutinin-I; *Phaseolus vulgaris* erythroagglutinin, PHA-E; *Phaseolus vulgaris* leukoagglutinin, PHA-L; Jacalin, JAC; peanut agglutinin, PNA) and Texas-Red Avidin D® for visualizing their binding were purchased from Vector Laboratories (distributed by Enzo Life Sciences, Lörrach, Germany). DT40 B lymphoma cells (avian leukosis-induced bursal lymphoma line; ATCC® CRL-2111™) [18,19] were kindly provided by S. Härtle and B. Kaspers (Chair of Animal Physiology, Faculty of Veterinary Medicine, Ludwig-Maximilians-University, Munich, Germany).

2.2. Tissue preparation and isolation of bursal B cells

Bursae of Fabricius were dissected from the proctadael region of the cloaca of five 4-week-old chickens, snap frozen in liquid nitrogen and then further processed for RNA isolation, extract preparation for Western blotting and cryo-sectioning. Bursal B cells were routinely obtained from freshly dissected tissue that was cut into small pieces in ice-cold phosphate-buffered saline, pH 7.2 (PBS). These pieces were placed on a stainless steel mesh screen (pore size: 75 μm) and gently pressed through the screen using the plunger of a 10 mL syringe into a petri dish, whose surface was covered with ice-cold PBS. The resulting suspension was transferred to a 50 mL tube and filled up to 50 mL with PBS. After 5-10 min on ice, the supernatant was gently poured into a new 50 mL tube and centrifuged for 10 min at 225 x g, the resulting cell pellet was resuspended in 10 mL PBS and this suspension layered on top of an equal volume of Ficoll solution (density: 1.077 mg/mL; Biocoll Separating Solution, Biochrom AG, Berlin, Germany). Following a centrifugation step at room temperature for 12 min at 400 x g, the cell-containing interphase was collected and washed twice with ice-cold PBS. Thereafter, the cell pellet was suspended in RPMI 1640 medium (containing 8% fetal calf serum (obtained from Life Technologies, Darmstadt, Germany), 2% chicken serum (Sigma-Aldrich, Munich, Germany) and antibiotics) to give 2.5 x 10^7 cells/mL, 10 μL of this suspension were carefully pipetted onto microscope slides, cells were fixed in cold acetone for 10 s and used for cytochemical staining. Cells were also pelleted, snap frozen and used for RNA isolation and Western blotting, respectively.

2.3. Expression profiling by RT-PCR

Isolation of total RNA from bursa of Fabricius tissue and purified B cells as well as the ensuing preparation of cDNA were done, as described previously for tissues of 6-month-old chickens [12,20]. PCR amplification of C-GRP-specific mRNA was directed by the sense primer 5′-TCTAGAATGGCGGGGACCGTG-3′ and the antisense primer 5′-AAGCTTTCAGCCAAGTTTTGTAAG-3′ as described [8]. The mRNAs of the canonical CGs used primer sets as in previous studies [12,20]. In detail, CG-1A-specific mRNA was amplified with the sense primer 5′-GGATCCATGGAGCAAGGACTG-3′ and the antisense primer 5′-GAATTCAATTTTCTCCATGCCCAGCCG-3′, CG-1B-specific cDNA by the sense primer 5′-ATGTCTTGTCAGGGACCA-3′ and the antisense primer 5′-TTACTCCCAGCTGACAGACCG-3′, CG-2-specific cDNA by the sense primer 5′-

AGAATGTTTGAAATGTTCAAC-3′ and the antisense primer 5′-TCACTCCACCTTGAAGGAG-3′, CG-3-specific cDNA by the sense primer 5′-CCCGGCGTACCCTGGATA-3′ and the antisense primer 5′-AAATCATGGAGGTCAAAACAC-3′, and CG-8-specific cDNA by the sense primer 5′-ATGATGTCCTTGGATGGA-3′ and the antisense primer 5′-CTACCAGCTCCTCACATC-3′. The lengths of amplified cDNAs for C-GRP, CG-1A, CG-1B, CG-2, CG-3, and CG-8S/L are 516 bp, 351 bp, 408 bp, 399 bp, 627 bp, and 891/948 bp, respectively. The reaction mixture for the PCR reactions was prepared, as recommended by the distributor of the *Taq* DNA polymerase (Qiagen). In general, the analyses were carried out in a volume of 20 µL containing 2.5 µg cDNA, 0.5 U Taq polymerase, 50 µM dNTPs, 1×PCR-buffer (commercial mixture of Tris-HCl, KCl, $(NH_4)_2SO_4$, 1.5 mM $MgCl_2$, pH 8.7), at 20 °C, 1×Q-Solution and 0.1 µM of sense and antisense primers. Amplification of the galectin-specific cDNAs was started with an initial denaturation step at 94 °C for 4 min, then 29 consecutive cycles of the following series of steps were performed: denaturation at 94 °C for 45 s, annealing at 52 °C (CG-1B, CG-2, CG-3, CG-8, actin) or 60 °C (C-GRP, CG-1A) for 45 s and extension at 72 °C for 1 min. The final extension step was carried out at 72 °C for 10 min. PCR products and reagents were separated in 2% agarose gel electrophoresis at 80 V for 40 min. The loading control with chicken β-actin-specific mRNA was established with the sense primer 5′-ATGGCTCCGGTATGTGC-3′ and the antisense primer 5′-ACCGTGTTCAATGGGGT-3′. The length of amplified cDNA was 185 bp.

2.4. Extract preparation and Western blotting

Tissue pieces from bursae of Fabricius and isolated bursal B cells were processed in lysis buffer (50 mM Tris-HCl, pH 8.0, 150 mM NaCl and 1% (w/v) NP-40). Polyclonal antibody preparations were obtained using recombinant C-GRP (full length)/CGs as antigens in rabbits as described [20,21]. In detail, immunization was started with 200 µg recombinant C-GRP in an emulsion of 0.5 mL PBS mixed with Freund's complete adjuvant (0.5 mL; Sigma-Aldrich) and pursued with booster injections in three-to-six-week intervals using Freund's incomplete adjuvant (Sigma-Aldrich). The titer was regularly monitored by ELISAs with the antigen. The immunoglobulin G (IgG) fraction obtained by affinity chromatography over protein A-Sepharose 4B was rigorously examined for cross-reactivity by Western blotting/ELISAs using each of the canonical CGs/C-GRP as bait, followed by removal of respective activity by chromatographic depletion on resin presenting the respective CG and subsequent controls to

verify complete removal [22]. The content of protein after coupling to divinyl sulfone-activated Sepharose 4B varied for the CGs between 4.5 mg/mL and 12 mg/mL resin. Western blot analyses were performed following an optimized protocol previously applied to CG-8 detection in extracts of bursal tissue from 6-month-old chickens [20]. Briefly, proteins (100 µg of cell and tissue lysates, 12 ng of purified recombinant C-GRP as positive control and a 25fold excess of each of the canonical CGs (300 ng)) were separated on a discontinuous SDS polyacrylamide gel (4% stacking gel, 10% running gel), and proteins were transferred by tank blotting onto a nitrocellulose membrane (0.2 µm pore size; Schleicher & Schuell, Dassel, Germany). Residual sites on the matrix for binding protein were saturated by incubation with a solution of Tris-buffered saline (TBS, pH 7.5) containing 0.1% Tween 20 (TBS-T) and 5% powdered milk. Membranes were routinely incubated overnight at 4 °C with IgG fractions against C-GRP/canonical CGs (anti-C-GRP: 0.5 µg/mL; anti-CG-1A: 1.0 µg/mL; anti-CG-1B: 0.25 µg/mL; anti-CG-2: 1.0 µg/mL; anti-CG-3: 0.25 µg/mL; anti-CG-8: 2.0 µg/mL) in blocking solution. Following a washing step with TBS-T, membranes were incubated with goat anti-rabbit IgG-horseradish peroxidase conjugate (0.5 µg/mL; Sigma-Aldrich) in blocking solution for 1 hour at room temperature and developed using enhanced chemiluminescence detection, driven by a substrate mixture with 1.25 mM sodium salt of luminol in 2 mL 0.1 M Tris-HCl (pH 8.6), 0.2 mL of a 6.7 mM p-coumaric acid solution in dimethyl sulfoxide and 0.6 µL H_2O_2 (30% v/v) at room temperature. Exposure time of the processed blots to CL-XPosure™ X-ray film (Pierce, Bonn, Germany) was set from 2 to 10 minutes to yield optimal signal intensity at minimal background.

2.5. Immuno- and (ga)lectin cyto- and histochemistry

Snap-frozen bursae of Fabricius were cut into serial sections of about 5-10 µm thickness using a CryoStar NK70 (Thermo Fisher Scientific, Dreieich, Germany). They were mounted on Superfrost® Plus slides (Menzel, Braunschweig, Germany), fixed for 10 min in ice-cold acetone and then air-dried. For immunostaining of tissue sections and fixed B cells (please see 2.2), processing started with washing in 10 mM PBS followed by saturation of sites for non-specific protein binding by an incubation step with a solution of 1% (w/v) carbohydrate-free bovine serum albumin (BSA) in PBS also containing 5% (v/v) normal goat serum (Enzo Life Sciences). Incubation with antibody was done in PBS/1% BSA solution containing one of the six anti-C-GRP/CG IgG fractions (anti-C-GRP: 0.5-2.0 µg/mL; anti-CG-1A: 0.25-2.0 µg/mL; anti-CG-1B: 0.5-2.0 µg/mL; anti-CG-2: 1.0-2.0 µg/mL; anti-CG-3: 0.25-1.0 µg/mL; anti-CG-8: 1.0-4.0

µg/mL) overnight at 4 °C in a humid chamber. The sections were carefully rinsed with PBS and thereafter incubated in the dark with Alexa Fluor® 568-labeled goat anti-rabbit IgG (whole molecule) conjugate (1.0 µg/mL; Sigma-Aldrich) for one hour at room temperature in a humid chamber, also containing 4′,6-diamidine-2′-phenylindole dihydrochloride (DAPI). When visualizing presence of C-GRP/CG-8 together with the chB6 marker by double staining, incubations with the polyclonal anti-C-GRP/CG-8 IgG fraction and second-step reagent (without adding DAPI) were performed as described above. The PBS/1% BSA solution containing the monoclonal antibody AV20 (2.5-5.0 µg/mL) was applied to the sections overnight at 4 °C in a humid chamber. After a washing step, the sections were incubated with a solution of PBS/1% BSA containing FITC-labeled anti-mouse IgG_1 (1.0 µg/mL) and DAPI (1.0 µg/mL). Stringent controls including omission of first-step reagent and antibody blocking (up to 20-fold molar ratio of antigen to antibody) were carried out with tissue sections and fixed B cells together with processing sections from organs known for positivity/negativity for canonical CGs to exclude any antigen-independent staining.

For lectin histochemistry with biotinylated plant/fungal lectins, tissue sections or fixed B cells were treated for 1 hour at room temperature in a humid chamber with a 2% (w/v) solution of BSA in 10 mM 4-(2-hydroxyethyl)-1-piperazine-ethanesulfonic acid buffer (HEPES, pH 7.5) to saturate non-specific protein-binding sites. Subsequently, after carefully rinsing the sections with HEPES buffer, incubation with the respective biotinylated plant/fungal lectin (PSL: 0.125-2.0 µg/mL; MAA-I: 0.5-1.0 µg/mL; JAC: 1.0-4.0 µg/mL; PNA: 0.25-1.0 µg/mL; PHA-E: 0.5-1.0 µg/mL; PHA-L: 0.125-4.0 µg/mL) was performed in HEPES/2% BSA solution overnight at 4 °C. Thereafter, washing steps with HEPES buffer to remove unbound lectin and with 10 mM PBS, pH 7.2, preceded the incubation step with a solution of PBS/1% BSA containing Texas-Red Avidin D® (30 µg/mL; Enzo Life Sciences) and DAPI (1.0 µg/mL). Controls included haptenic inhibition with mono- or disaccharides (JAC: 800 mM galactose; PNA: 200 mM galactose) to ascertain carbohydrate dependence of binding.

Processing for double staining to visualize presence of C-GRP/CGs and of accessible sites reactive with C-GRP/CGs in sections was started with exposing them for 1 hour at room temperature in a humid chamber to a 1% (w/v) solution of BSA in PBS containing 5% goat serum to saturate sites for non-specific binding of protein. Subsequently, the sections were incubated with PBS/1% BSA containing anti-C-GRP or anti-CG IgG fractions (with concentrations, as given above) in a humid chamber overnight at 4 °C. After three thorough washing steps with PBS, Alexa Fluor® 568-labeled goat anti-rabbit IgG (whole molecule) antibody conjugate (1.0 µg/mL; Sigma-Aldrich) was applied for 1 hour at room temperature in

a humid chamber in the dark. Then, after a washing step with PBS, sections were incubated with solutions of Alexa Fluor® 488-labelled C-GRP or CG (C-GRP: 2.5-5.0 µg/mL; CG-1A and -1B: 1.0-5.0 µg/mL; CG-2: 2.0-8.0 µg/mL; CG-3: 0.5-5.0 µg/mL; CG-8: 2.0-10.0 µg/mL) adding DAPI overnight at 4 °C in a darkroom. Controls included haptenic inhibition of CG binding with 200 mM lactose. For inhibition of binding of Alexa Fluor® 488-labelled C-GRP to tissue sections and fixed B cells, heparin (Biomol, Hamburg, Germany) and further polysaccharides were applied as tested for a human heparin-binding protein [23].

In all cases, processed sections were mounted in antifade medium (Vectashield; Enzo Life Sciences), and microphotographs during fluorescence microscopy were recorded using an AxioImager.M1 microscope (Carl Zeiss MicroImaging, Göttingen, Germany) equipped with an AxioCam MRm digital camera and the software AxioVision version 4.6.

2.6. Transmission electron microscopy

Immediately after removal, small pieces (1 mm side length) of bursa of Fabricius tissue were fixed by immersion in Karnovsky´s solution (2.5% glutaraldehyde and 2% paraformaldehyde) for 24 hours at 4 °C. Thereafter, tissue blocks were treated with 1% osmium tetroxide (Plano, Wetzlar, Germany)/1.5% potassium ferrocyanide (Sigma-Aldrich), dehydrated in a graded series of ethanol (50%, 70%, 90% and absolute ethanol, each step 30 minutes) and embedded in Polyembed 812 BDMA (Polysciences, Eppelheim, Germany). Semithin sections of 1 µm were prepared and stained with Richardson's solution [24]. Ultrathin sections were cut with an ultramicrotome equipped with a diamond knife (Reichert-Labtec, Wolfratshausen, Germany), placed on 150-mesh copper grids (Polysciences), stained with uranyl acetate (Scientific Ltd., Stansted, UK) and lead citrate (Agar Aids, Stansted, UK), then examined by a transmission electron microscope (EM 902; Zeiss) to record microphotographs.

2.7. Flow cytofluorimetry

Quantitative determination of carbohydrate-dependent lectin binding was performed by flow cytofluorimetry in a FACS Calibur instrument (Becton-Dickinson, Heidelberg, Germany). Solutions with 2 x 10^5 cells per sample were carefully washed to remove any interfering serum compounds, the extent of non-specific protein binding was reduced by using a solution of 0.1% carbohydrate-free BSA in Dulbecco's PBS. Fluorescent C-GRP/CGs or anti-C-GRP/CG antibodies (2.0 µg/mL) were incubated with the cells for 30 min at 4 °C to minimize endocytic

uptake. Fluorescent anti-rabbit IgG (1:100 dilution; Sigma-Aldrich) was used for detection of binding of the primary antibody, as described previously [25]. For intracellular staining, cells were fixed with 4% paraformaldehyde for 25 min at room temperature, followed by an incubation step in the presence of 0.5% saponin (Roth, Karlsruhe, Germany) for 30 min at room temperature to permeabilize the cell membrane. Control experiments without the incubation step with labeled antibody/lectin or in the presence of glycocompounds to block carbohydrate-dependent lectin binding as well as systematic titrations to define optimal concentrations were run in each case, routinely using aliquots of the same cell batch in comparative analysis.

2.8. Promoter analysis

Computational identification of sequence motifs that satisfy criteria of a binding site for transcription factors were performed, as described previously [26,27], using the program MatInspector, which minimizes redundant matches by arranging similar and/or functionally related transcription-factor-binding sites into matrix families [28]. The MatInspector software (http://www.genomatix.de/index.html) identifies matches by comparing the proximal promoter sequence with weighted matrix descriptions of functional binding sites based on the TRANSFAC® database and on the MatInspector-inherent 634 matrices grouped into 279 families (Matrix Library 9.3, released March 2015). The optimized threshold of a weight matrix was defined as the matrix similarity threshold that allows a maximum of three matches in 10 kb of non-regulatory test sequences. By scanning the imported sequence against the relative frequency of each nucleotide at a particular position in the program-based distribution profile, the matrix similarity was calculated on-the-run. Only sites with a core similarity of 1.0, which is reached only if the sequence analyzed is completely identical to matrix-inherent sequence sections with highest degree of conservation, were considered to be of putative significance.

With the algorithms provided by this program, the manually edited proximal promoter sequences (-2500 bp upstream of the translation start, applying the sequence text view tool implemented in NCBI Gene [29]) of five genes for B cell markers, i.e. activation-induced cytidine deaminase, (AICDA, Gene ID: 418257) [30], B cell-activating factor receptor (BAFF receptor)/tumor necrosis factor receptor superfamily, member 13C (TNFRSF13C, Gene ID: 417983) [31], CD40 (Gene ID: 395385) [32], CD79B (Gene ID: 419940) [33], chloride intracellular channel 2 (CLIC2, Gene ID: 422178) [34], and genes for C-GRP (Gene ID: 421278), and CG-8 (Gene ID: 430948) were used for analysis. In the case of the transmembrane activator and CAML interactor (TACI)/tumor necrosis factor receptor superfamily, member

13B (TNFRSF13B, Gene ID: 770275) [35], the processed promoter sequence was -2500 bp upstream of the transcription start.

3. Results and discussion

3.1. Expression profiling

Screening of C-GRP expression on the level of mRNA was performed with cDNA preparations of a panel of organs from 2-day-old, 4-week-old and 6-month-old chickens. Flanked by a loading control, positivity was defined as appearance of a band at the calculated size, as exemplarily shown in Fig. 1. In the documented case, this signal was detectable, and negativity after processing of material from frozen liver extracts served as control to exclude false-positive data. cDNAs from bursa, heart, kidney and thymus were consistently positive, cDNAs from other organs such as small intestine led to signals in up to two stages of development, while cDNA from liver was always negative. A similar picture with broad positivity had emerged for expression of the human and murine GRP genes based on entries to expressed sequence tag listings [3]. Extracts from chicken bursa had previously been reported to produce a strong signal for the mRNA specific for CG-8L (the isoform of the tandem-repeat-type protein with long (L) linker maintaining an exon relative to the short (S) form) and also for CG-3 [12,20]. This evidence was confirmed and extended to positivity for all five canonical CGs (not shown).

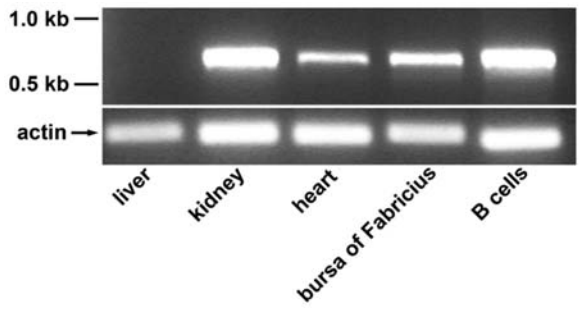


Fig. 1. Expression profiling of C-GRP by RT-PCR. The presence of C-GRP-specific mRNA was detected in samples from tissues and isolated bursal B cells of 4-week-old chickens as band at 516 bp. Probing of mRNA for actin served as positive control. Positions of markers are indicated (left).

In order to bring the analysis to the level of the protein, an antibody preparation against C-GRP was raised and ensured to be devoid of cross-reactivity to any canonical CG by affinity

chromatography using resin-immobilized CGs. When then probing for protein presence in extracts of organs from 4-week-old animals by Western blotting, occurrence of positivity was found to be restricted to bursa and isolated bursal B cells (Fig. 2A). Evidently, C-GRP was present in this organ at its stage of full maturation. In contrast to the canonical CGs, which are expressed in various adult organs with individual distribution profiles (for overview, please see [7,36]), immunohistochemistry with this antibody preparation detected no evidence for C-GRP positivity in any other tested tissue. Having obtained signals for the mRNAs of the five canonical CGs, the next question to address was the representation of canonical CGs in this tissue and its B cells. Presence of CG-8L, already described previously in whole-tissue extracts [20], was confirmed so that the respective analysis found both CG-8L and C-GRP in extracts of bursal B cells (Fig. 2). Systematic investigation by Western blotting disclosed signals for all canonical CGs in bursa extracts but, except for CG-8L, not in extracts of bursal B cells (Fig. 2). When examining the signals for CG-1A, staining was observed at the positions of the mono- and dimers under denaturing conditions, as for example reported for murine regulatory T cells [37]. Evidently, i) presence of C-GRP in amounts detectable by Western blotting appears specific for bursa of Fabricius at its full stage of maturation and ii) this organ contains the full set of canonical CGs, making it to an ideal study object to define the localization profiles of the six family members in one organ.

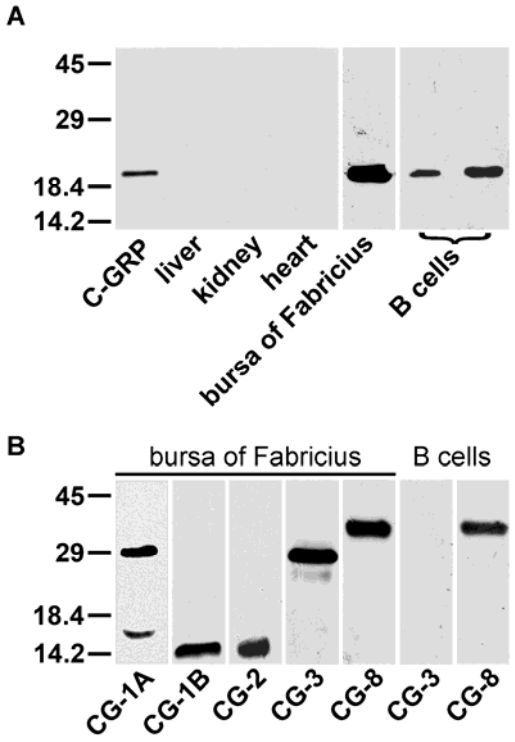


Fig. 2. Expression profiling of C-GRP and the canonical CGs by Western blotting. Detection of C-GRP (A)/CGs (B) in extracts from tissue samples (100 µg total protein; liver, kidney, heart, bursa of Fabricius for C-GRP, bursa of Fabricius for canonical CGs) and isolated bursal B cells (50 µg and 100 µg total protein) of 4-week-old chickens using polyclonal IgG fractions against C-GRP/CGs (anti-C-GRP: 0.5 µg/mL; anti-CG-1A: 1.0 µg/mL; anti-CG-1B: 0.25 µg/mL; anti-CG-2: 1.0 µg/mL; anti-CG-3: 0.25 µg/mL; anti-CG-8: 2.0 µg/mL), each free of any cross-reactivity to the other five proteins. Positions of molecular weight markers are indicated, a control with purified C-GRP (12 ng) is included (A, left), origin of extract (A) and the type of the detected CG (B) given.

3.2. Immunocyto- and histochemical and cytofluorimetric profiling

Initial immunohistochemical experiments with different organs were carried out to prove specificity and to preclude missing a specific signal by titrations. In detail, sections of an organ panel including bursa, heart and kidney were processed to ascertain correlation of Western blotting data with positivity in immunohistochemistry. Antibodies against canonical CGs served as positive control to exclude false-negative results. Systematic experiments were run with a series of concentrations of the antibody preparations for C-GRP and the canonical CGs. In full agreement to the data from Western blotting, C-GRP positivity was present in sections of bursa but no other tissue, when testing of 4-week-old animals. Obtaining the known regional selectivity for canonical CGs, as documented for example in sections of kidney previously [7,38], ensured the reliability of the technical procedure. These immunohistochemical series thus solidified the focus on this tissue type. In order to attain an optimal signal-to-background

ratio, titrations were performed for each antibody preparation on serial sections, setting the stage for the network monitoring. A scheme of the bursa's morphological organization along with its actual appearance in a section is presented in Fig. 3 to provide an orientation for readers. As follows, localization profiles with individually distinct characteristics will be presented for C-GRP and each of the canonical CGs in Figs. 4A-C, G-I, M-R, together with electron microscopical documentation on relevant cell types with CG positivity (Figs. 4D-F, J-L) (for detailed explanations to each illustration, please see legend to this figure; for overview of staining profiles, please see Table 1).

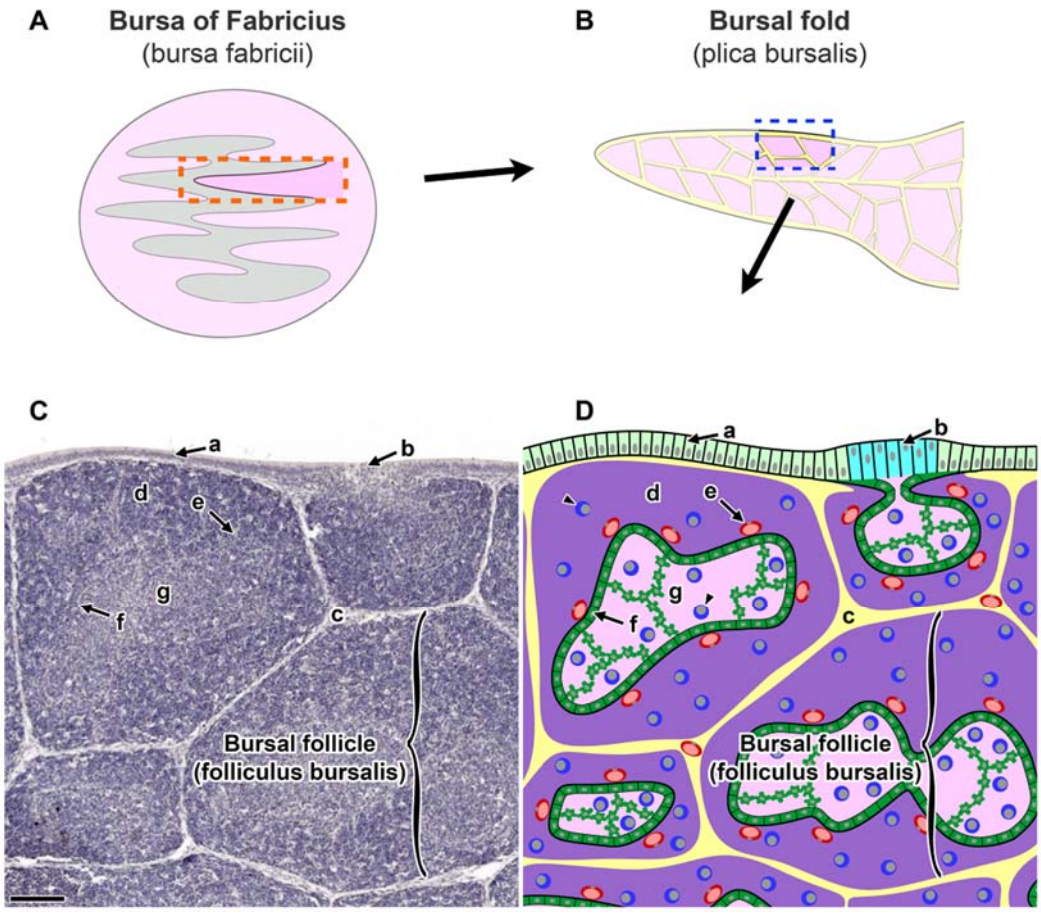


Fig. 3. The bursa of Fabricius (A) is a blind sac-like diverticulum in the dorsal wall of the cloaca. The mucosal layer in the inner surface forms multiple folds (one such fold or plica bursalis, as marked by a dotted orange line is shown enlarged (B)). These folds contain lymphoid tissue organized into nodules, also called bursal follicles. The area depicted inside the blue dotted line is shown enlarged as a microphotograph of a hemalaun-stained tissue section (C; *scale bar*: 100 µm) and as a scheme of the same section (D). The follicles which contains B cells (*arrowheads*) are covered by interfollicular surface epithelium (IFE, a) or, when in direct contact with the epithelium, by follicular-associated epithelium (FAE, b). The tissue between the follicles is made up of interfollicular connective tissue (c). The follicles consist of a cortex or pars lymphoreticularis (d), a cortico-medullary border which is formed by numerous capillaries (e) and a cortico-medullary epithelium (f), and a medulla or pars lymphoepithelialis (g).

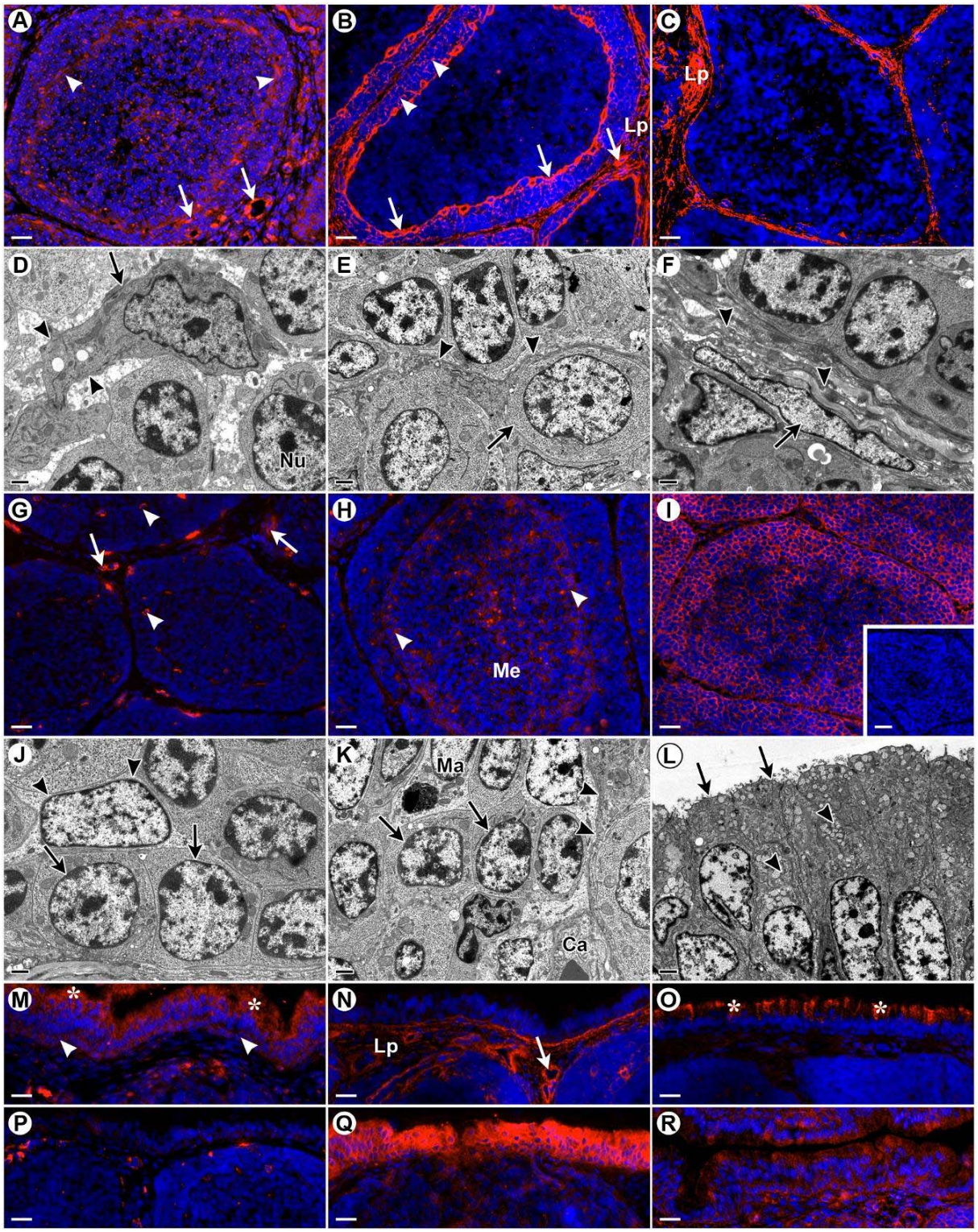


Fig. 4. Immunohistochemical localization of C-GRP and canonical CGs by fluorescence microscopy and identification of specific cell types on the electron microscopical level in sections through follicles (A-L) as well as follicle-associated epithelium (FAE) and interfollicular epithelium (IFE) (M-R) of bursae of Fabricius from 4-week-old chickens. Because these illustrations establish the basis for the summary given in Table 1, each microphotograph is explained in thorough detail. (A, M) C-GRP presence in reticular epithelial cells (*arrowheads,* staining intensity prominent at the cortico-medullary border), in B cells (at different stages of their lineage, dot-like staining pattern in the Golgi apparatus) both in medulla and cortex. Strong staining intensity is observed in

the endothelium of blood vessels (*arrows*) located in the capillary layer at the cortico-medullary junction and in the lamina propria (A). In the FAE, strong reactivity was observed supranuclearly *(asterisks)*, whereas basal cells (*arrowheads*) were stained only weakly (M). (B, N) Intense reactivity of CG-1A-specific antibodies in the basal lamina and in the endothelium of capillaries (*arrows*) located at the cortico-medullary border and between follicles, in the lamina propria (Lp) and in mesenchymal reticular cells of the cortex (B, *arrowheads*). FAE/IFE were negative (N). (C, O) CG-1B presence in the lamina propria (Lp, C) and supranuclearly (*asterisks*) in epithelial cells (O). On the electron microscopical level, micrographs D-F illustrate details of C-GRP/CG-positive cells and structures. (D) Reticular cells (*arrow*) showed their typical irregular shape with filopodia (*arrowheads*), and nucleoli (Nu) of lymphoblasts were prominent. (E) Separation of medulla and cortex by the cortico-medullary border, which consists of the basal lamina (*arrowheads*, CG-1A positive, please see B) and the cortico-medullary epithelial cells (*arrow*, CG-3 positive, please see H). (F) Ultrastructure of interfollicular tissue (positive for CG-1A, CG-1B, please see B, C). Collagen fibers (*arrowheads*) and a fibroblast (*arrow*) in its typical shape are shown. (G-I; P-R) Fluorescent staining patterns after applying anti-CG-2 (G, P), anti-CG-3 (H, Q) and anti-CG-8 (I, R), respectively. CG-2 presence is confined to endothelial cells in capillaries (*arrowheads*) at the cortico-medullary border and in blood vessels (*arrows*) embedded in the lamina propria (G). Cells lining the FAE/IFE were negative (P). (H, Q) Reactivity for anti-CG-3 was observed in cortico-medullary epithelial cells (*arrowheads*) at the cortico-medullary border and in reticular epithelial cells disseminated throughout the whole medulla (Me, H). Strong immunopositivity for CG-3 was found in the cells of the FAE/IFE (Q). (I, R) CG-8 presence is confined to medullary and cortical B cells (I; inset: control without first-step antibody), and with weaker intensity to the cytoplasm of basal cells in the FAE/IFE (R). (J, K) Electron micrographs of reticular epithelial cells (*arrowheads*, CG-3 positive, please see H) surrounded by B lymphoblasts (*arrows*, with prominent nucleoli; C-GRP and CG-8 positive, please see A, I) or a capillary vessel on the bottom right edge (Ca, K; CG-1A-, CG-2-positive, please see B, G) and in the vicinity of a macrophage (Ma) containing a large amount of phagocytosed material (K). (L) Ultrastructure of cells in the FAE (positive for C-GRP, CG-1B, CG-3, CG-8, please see M, O, Q, R). Thin microvilli (*arrows*) on the cells´ surface and supranuclear secretory granules (*arrowheads*) are clearly visible. Concentrations of antibodies used were 1.0 µg/mL for anti-CG-1A IgG, anti-CG-1B IgG as well as for anti-CG-3 IgG, 2.0 µg/mL for anti-C-GRP IgG and anti-CG-2 IgG, and 4.0 µg/mL for anti-CG-8 IgG. Nuclei were counterstained with DAPI. The scale bars are 1 µm (D-F; J-L), 20 µm (A-C; G-I; M-R) or 50 µm (*inset* to I).

Immunomonitoring for C-GRP reveals a distinct pattern with positivity in the follicle-associated and interfollicular epithelia (FAE/IFE) as well as in B cells, capillary vessels and reticular epithelial cells (Figs. 4A, M), in close correspondence with the electron microscopical characterization of cell features. As summarized in Table 1, staining was cytoplasmic, here often supranuclear. In this respect, the protein can share intracellular functionality with other members of this family, for example in counterreceptor routing, growth regulation, or splicing [39,40]. The parallel testing of the specific antibody preparations for each canonical CG disclosed non-redundant expression profiles. As could be expected due to the positive Western blot data (Fig. 2), each CG could be localized: CG-1A in basal lamina and vessels (Figs. 4B, N), CG-1B in lamina propria and the IFE (Figs. 4C, O), CG-2 mostly in capillaries (Figs. 4G,

P), CG-3 strongest in epithelial cells (Figs. 4H, Q) and CG-8, in contrast to CG-3 and the three proto-type CGs but similar to C-GRP, in medullary and cortical B cells (Figs. 4I, R). Pinpointing cell types/regions of overlapping positivity, C-GRP shares positivity in vessels with CG-1A, in IFE/FAE with CG-3 and CG-1B (IFE) and in B cells with CG-8 (Table 1). Because solid-phase assays (but not gel filtration) revealed a tendency for self-association of C-GRP and for association of this protein with CG-3, less with CG-8 and proto-type CGs (not shown), as first noted for surface-presented human galectin-9 in self-association and association with galectins-3 and -8 but not galectin-1 [41,42], an in situ interaction between C-GRP and CG-3 in a suited structural context could be possible in epithelia. Turning to the case of C-GRP/CG-8, Cell-type-specific co-expression of C-GRP with CG-8 had already been indicated when probing B cell extracts by Western blotting (Fig. 2). Owing to the potential of functional significance of this co-expression, considering that the two murine galectins-1 and -8 are both expressed by B cells and cooperate to promote plasma cell formation [43], we further validated this result by colocalization of C-GRP/CG-8 and a B cell marker, hereby also examining the cellular distribution in more detail.

Table 1
Immunocyto- and histochemical profiles of presence of C-GRP and the five canonical CGs

cell type	C-GRP	CG-1A	CG-1B	CG-2	CG-3	CG-8
epithelium						
interfollicular epithelium (IFE)	+[a]/++[b]	-	++[b]	-	+++[f]	+[f]
follicle-associated epithelium (FAE)	+[a]/++[b]	-	-	-	+++[f]	-
basal lamina	-	+/++	-	-	-	-
lamina propria						
interfollicular tissue	-	+	++	-	-	-
blood vessels	++[c]	++[e]	-	++[c]	-	-
cortex						
B cells	+[d]	-	-	-	-	++[g]
mesenchymal reticular cells	-	+	-	-	-	-
macrophages	-	-	-	-	-	-
cortico-medullary border						
epithelial cells	+	-	-	-	-	-
capillary vessels	++[c]	+++[e]	-	+[c]	-	-
medulla						
B cells	+[d]	-	-	-	-	++[g]
reticular epithelial cells	-	-	-	-	+	-
dendritic cells	-	-	-	-	-	-
isolated B cells	++[d]	-	-	-	-	++

Signal intensity was semiquantitatively grouped into the categories: - negative, + weak but significant, ++ medium, +++ strong; [a]cytoplasm of basal cells, [b]supranuclear positivity, [c]endothelium, [d]dot-like staining pattern at the Golgi apparatus, [e]endothelium and smooth muscle layer, [f]cytoplasmic, [g]surface associated.

The staining patterns of the antibodies against C-GRP and CG-8 (red; Figs. 5A, D) were separately superimposed with the distribution of green signals obtained by applying a monoclonal antibody against a pan B cell marker, i.e. the cell surface chB6 epitope (Figs. 5B, E). Intriguingly, the resulting overlay for the sections through follicular medulla revealed no apparent overlap between C-GRP and chB6, due to cytoplasmic (C-GRP) and cell surface (chB6) presence (Fig. 5C). In contrast, staining patterns for CG-8 and the marker clearly show an overlap in the sections (Fig. 5F). The same procedure was performed for isolated B cells from bursal tissue and provided evidence for a lower extent of colocalization in the pairing of the B cell marker with C-GRP (Figs. 5G-I) than with CG-8 (Figs. 5J-L). The other CGs were not detectable immunocytochemically under these conditions (not shown). These results enabled to add this information on isolated B cells to Table 1 and to conclude that the localization profiles of C-GRP and CG-8 are different in B cells.

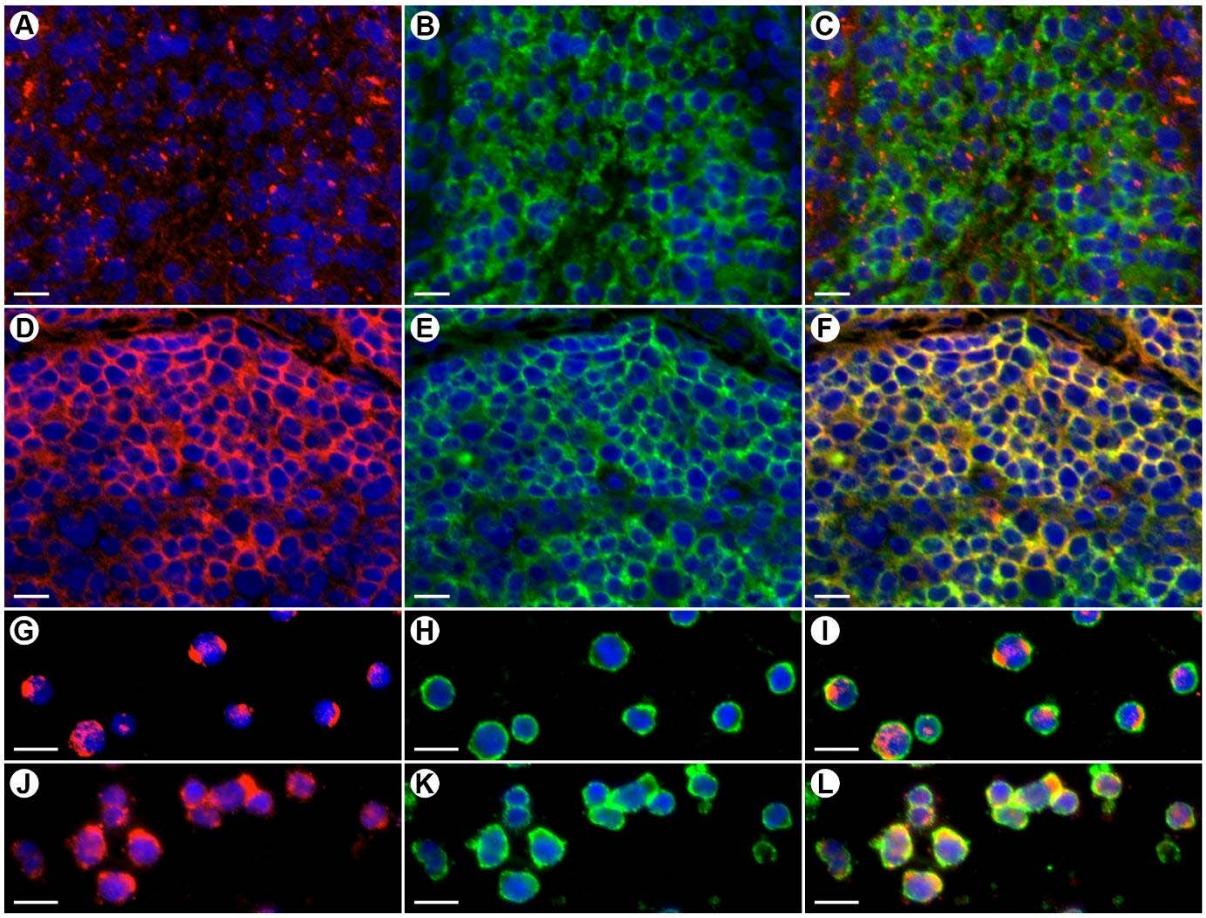


Fig. 5. Colocalization of C-GRP (red) and CG-8 (red) with the pan B cell marker, the chB6 (Bu-1) (green) surface molecule, in sections through the medulla of a follicle (A-F) and in isolated B cells (G-L). Anti-C-GRP/CG-8-specific first-step antibodies were applied in combination with Alexa Fluor® 568-labeled second-step goat anti-rabbit IgG. Mouse monoclonal antibody AV20 to detect chB6 was combined with FITC-labeled goat anti-mouse IgG1 as a second-step reagent. (A-F) Positivity for C-GRP and the membrane-associated staining pattern of the chB6 surface molecule in follicle-associated (A, B) and isolated B cells (G, H) were not colocalized (C, I). In contrast, staining for CG-8 overlaps clearly with the pattern of the marker both in the follicular tissue (D-F) and in isolated B cells (J-L), indicating CG-8 presence on the cell surface. For concentrations of antibodies used, please see legend to Fig. 4; monoclonal antibody AV20 was applied at 2.5 µg/mL. Nuclei were counterstained with DAPI. The scale bars are 10 µm.

As a means to further study this aspect on cells, we carried out flow cytofluorimetric measurements on native and permeabilized bursal lymphoma (DT40) cells. Under these conditions and testing CG-1B as control, cells were only very weakly positive for presence of C-GRP on the surface, in comparison more frequently and more intensely seen in the cytoplasm after treatment of cells with detergent for membrane permeabilization (Fig. 6). As seen in tissue sections and B cells, this protein is present within cells also for the cultured line. In comparison, the contributions to positivity indicated a relative shift of staining to the cell surface in the case

of CG-8 (Fig. 6). Thus, these two members of the galectin family are present in virus-transformed chicken B cells with differences in quantitative aspects of sites of their localization. Having hereby revealed where C-GRP can be detected in bursa, identifying B cells as major site, and expression in normal and malignant B cells, questions on putative regulatory sequences in the promoter arise and are addressed.

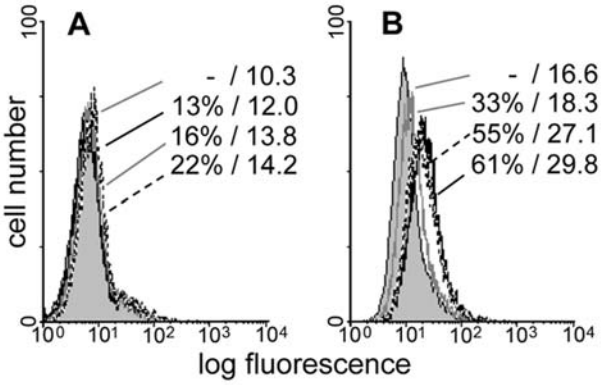


Fig. 6. Semilogarithmic representation of fluorescent staining of cells of the chicken DT40 cell line by rabbit polyclonal antibodies directed against C-GRP (black line), CG-1B (grey line) and CG-8 (dashed line). Antibodies were used at a concentration of 2 µg/mL. (A) Cell surface staining and (B) staining of cells after permeabilization by addition of 0.5% saponin to the incubation buffer. The control value without antibody addition is given as shaded area. Numbers characterizing staining (percentage of positive cells/mean florescence intensity) are given in each panel.

3.3. Computational profiling of the promoter region

Applying search and assessment algorithms to the promoter region of the C-GRP gene (-2500 bp upstream of the translation start), the set of putative binding sites for transcription factors (TFs) can be established. The respective list, given in Table S1, is the basis for comparative considerations, as a guideline for experimental work. When applied to the question on C-GRP expression in B cells, a starting point for tracing functionally relevant sequences is a comparison to a respective compilation for the CG-8 gene, based on the disclosed co-expression in B cells. In addition, spotting identities to promoter sequences of genes of chicken B cell-associated products can provide further guidance. Therefore, a set of six genes meeting this criterion was selected and their promoter sequences processed accordingly. As shown in Table 2, a total of 11 structural motifs that are classified as binding sites for transcription factors (TFBS) is encountered in all these promoter regions, and, moreover, further sequence elements with putative reactivity to TFs were found, if one, two or three genes were not part of the

comparison (Table 2). Obviously, these sequences are candidates for detailed experimental testing of their impact on actual expression. That this search uncovered a series of similar motifs also encourages to examine promoter sequences of the GRP gene in other vertebrates, keeping the high-level sequence conservation in mind.

Table 2
Shared sequences in the promoters of the genes for C-GRP, the canonical CG-8 and six B cell-associated proteins (AICDA, BAFF-R, CD40, CD79B, CLIC2, and TACI)

	TFBS in 8/8 promoters	TFBS in 7/8 promoters (incl. C-GRP + CG-8)	TFBS in 6/8 promotors (incl. C-GRP + CG-8)	TFBS in 5/8 promoters (incl. C-GRP + CG-8)
Transcription factor (TF)	ETV1[a], Lactoferrin[a], MYOD[a], MZF1[a], NFAT5[a], NF-κB[a], OLIG2[a], PAX5[a], PRDM1[a], SPI1[a], TCFE2A[a]	MAZ[a], MAZR[a], NFAT[a], OCT1[a], TLX1[b], VMYB[a], WHN[a], ZBTB7[a], ZNF202[a], ZNF219[b]	ARID5A[b], BCL6[a], FOXP1_ES[a], GATA1[a], JUNB[a], MEL1[a], MYBL1[a], RBPJK[a], STAT3[a]	EVI1[a], HHEX[b], HOX1-3[a], BLIMP1[b], STAT1[a], RFX4[a]

[a] unmarked TFs show a common distribution within analyzed promoter regions of B cell-associated and CG genes
[b] TFs marked in grey are putative interesting factors, because their binding site is mainly detectable in B cell-associated promoter regions

The distribution of putative transcription factor binding sites (TFBS) for TFs expressed in cells of the immune system are shown for C-GRP, CG-8 and genes, that are known to be expressed in B-cells, such as AICDA, BAFF-R, CD40, CD79B, CLIC2, and TACI. Names of transcription factors: ARID5A: AT rich interactive domain 5A (MRF1-like); BCL6: B-cell CLL/lymphoma 6; BLIMP1: Transcriptional repressor B lymphocyte-induced maturation protein-1; ETV1: Ets variant 1; EVI1: EVI1-myeloid transforming protein; FOXP1_ES: Alternative splicing variant of FOXP1 activated in ESCs; GATA1: GATA-binding factor 1; HHEX: Haematopoietically expressed homeobox, proline-rich homeodomain protein; HOX1-3: Hox-1.3, vertebrate homeobox protein; JUNB: Transcription factor Jun-B; Lactoferrin: Lactotransferrin and delta-lactoferrin; MAZ: Myc-associated zinc finger protein; MAZR: MYC-associated zinc finger protein related transcription factor; MEL1: MEL1 (MDS1/EVI1-like gene 1) DNA-binding domain 2; MYBL1: V-myb avian myeloblastosis viral oncogene homolog-like 1 (AMYB); MYOD: Myogenic regulatory factor MyoD (myf3); MYOGENIN: Myogenic bHLH protein myogenin (myf4); MZF1: Myeloid zinc finger protein; NFAT: Nuclear factor of activated T-cells; NFAT5: Nuclear factor of activated T-cells 5; NF-κB: Nuclear factor κB; OCT1: Octamer-binding transcription factor-1, POU class 2 homeobox 1 (POU2F1); OLIG2: Oligodendrocyte lineage transcription factor 2; PAX5: Paired box protein 5, B-cell-specific activator protein; PRDM1: PRDI (positive regulatory domain I element) binding factor 1; RBPJK: Mammalian transcriptional repressor RBP-Jκ/CBF1; RFX4: Regulatory factor X, 4; SPI1: SPI-1 proto-oncogene; hematopoietic transcription factor PU.1; STAT1: Signal transducer and activator of transcription 1; STAT3: Signal transducer and activator of transcription 3; TCFE2A: transcription factor E2a (E12/E47); TLX1: T cell leukemia homeobox 1; VMYB: v-Myb, variant of AMV v-myb; WHN: Winged helix protein; ZBTB7: Zinc finger and BTB domain containing 7A; ZNF202: Zinc finger 202; ZNF219: Kruppel-like zinc finger protein 219.

Of course, expression profiling and GRP localization will need to be defined in other vertebrates, before these comparisons should be performed. Experimentally, a salient aspect of GRP activity can next be characterized due to the availability of the protein. Having determined the loss of the typical lectin activity in C-GRP and seen binding to mammalian cells without inhibitability by a panel of glycans [8], the availability of labeled C-GRP enables to use the protein as tool to monitor sections of chicken bursa and isolated B cells for reactivity. As routinely done by lectin histochemistry, the labeled protein can then associate to accessible binding partners. This technique enables a comparison between staining profiles of C-GRP and CG-8, which can be extended to all canonical CGs, thus establishing a comprehensive

fingerprinting of reactivity to the six proteins, and it can additionally characterize aspects of the glycome, relevant for CG-8 binding.

3.4. (Ga)lectin cyto- and histochemical and cytofluorimetric profiling

The respective experiments with C-GRP and CG-8 were performed by double labeling visualizing the protein (in red) and reactive sites (in green), and this pair of cyto- and histochemical illustrations is presented (Figs. 7A, D, G, J, for immunolocalization, Figs. 7B, E, H, K for localization of C-GRP/CG-8 reactivity and Figs. 7C, F, I, L for overlay). As done for antibodies, systematic titrations had been performed for the labeled proteins to define a concentration that will generate an optimal signal-to-background ratio. In addition, applying lactose as additive to the lectin-containing solution proved the canonical-galectin-dependent staining to be inhibitable (insets to Figs. 7E, K). C-GRP reactivity was found in distinct sites, as summarized in Table 3. Whereas capillary vessels were immunopositive (Figs. 4A, 7A), weak reactivity for the labeled protein was seen at these sites (Fig. 7B). Labeled C-GRP was found to bind to the IFE/FAE more strongly, yielding a similar intensity as reached in immunohistochemistry (Fig. 7B). Yellow staining in superimposition was thus present in the apical part of cells in the IFE and also in capillary vessels of the cortico-medullary border (Fig. 7C). The reactivity for CG-8 was rather similar in the epithelium, matching immunopositivity in IFE (Table 3, Fig. 4R, Fig. 7D) so that the overlap with sites of lectin presence mostly resided in B cells (Fig. 7F). In isolated B cells, signal overlap, too, was apparent, here more on the cell surface for CG-8 than C-GRP (Figs. 7G-L). These data showed that both labeled proteins are sensors for binding sites in the tissue and encouraged to extend the analysis to the other canonical CGs.

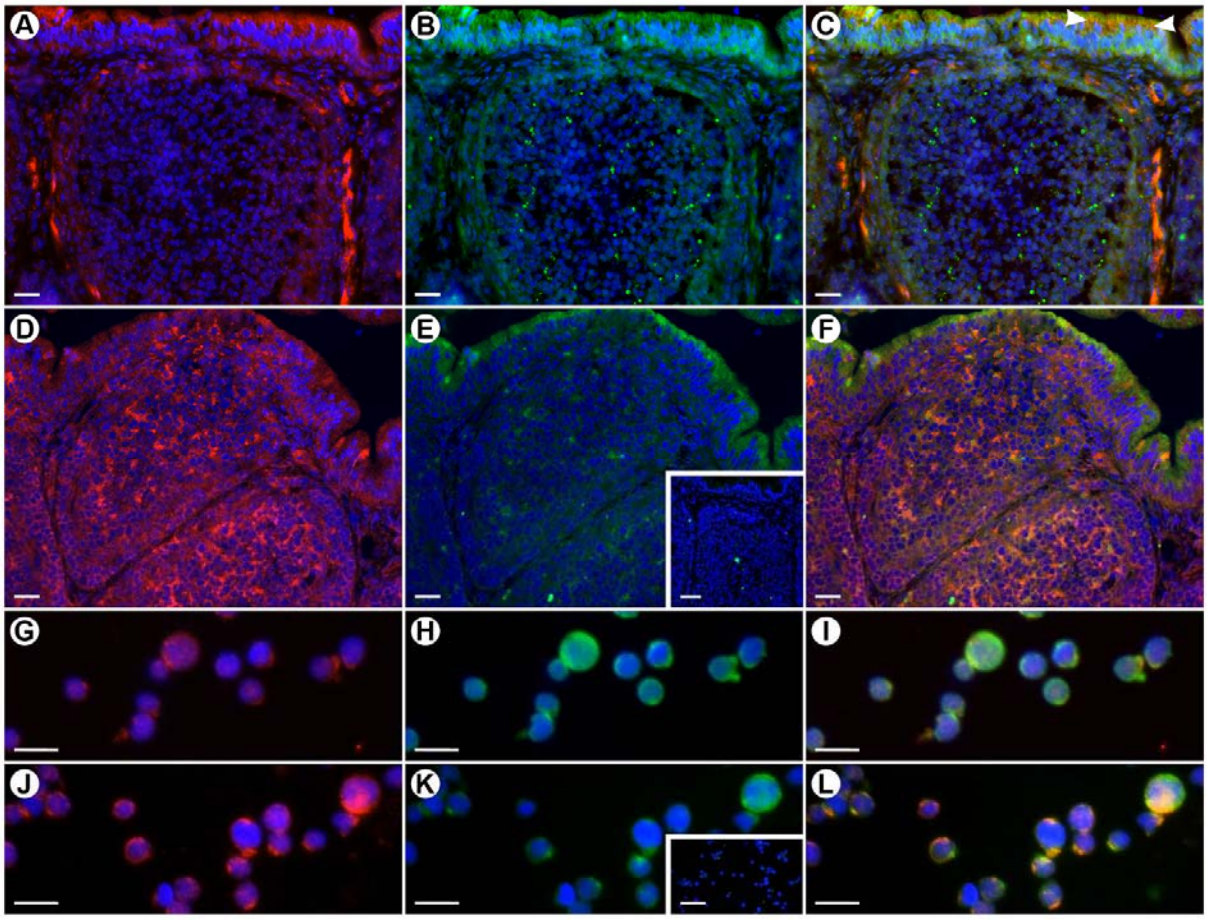


Fig. 7. Combined histo- and cytochemical staining using anti-C-GRP IgG/fluorescent C-GRP and anti-CG8 IgG/fluorescent CG-8, respectively, in sections of follicles (A-F) and in isolated bursal B cells (G-L). Staining with Alexa Fluor® 568 (red)-labeled second-step antibody to detect C-GRP (A, G)/CG-8 (D, J) presence and Alexa Fluor® 488 (green)-labeled C-GRP (B, H)/CG-8 (E, K; *insets*: controls by inhibition with cognate sugar) to detect accessible binding sites was performed. Positivity for C-GRP (A, G) and C-GRP-binding sites (B, H) led to overlap (yellow) in double staining in the apical part of cells of the IFE (C, *arrowheads*). A weak overlap was detected in capillary vessels (C) and also in isolated bursal B cells (I). Overlap of CG-8 presence and accessible binding sites was found in bursal B cells (in tissue sections and isolated cells) to various degrees (F, L), and also in the apical part of the IFE (F), which is reactive with labeled CG-8 (E). For concentrations of antibodies used, please see legend to Fig. 4. Concentrations of fluorescent C-GRP and CG-8 used were 10 µg/mL. Nuclei were counterstained with DAPI. The scale bars are 10 µm (G-L), 20 µm (A-F, *inset* to K) or 50 µm (*inset* to E).

Table 3
Cyto- and histochemical profiles of binding of fluorescent C-GRP and the five canonical CGs

cell type	C-GRP	CG-1A	CG-1B	CG-2	CG-3	CG-8
epithelium						
interfollicular epithelium (IFE)	+/++[a]	+[c]	++[d]	-	++/+++[c]	+/++[d]
follicle-associated epithelium (FAE)	+/++[a]	+[c]	-/+[a]	-	++[c]	+/++[d]
basal lamina	-	-	-	-	-	-
lamina propria						
interfollicular tissue	-/+	-	+	-	-	-
blood vessels	-/+[b]	-	-	-	-	-
cortex						
B cells	+	-	-	-	-	+
mesenchymal reticular cells	-	-	-	-	-	-
macrophages	-	-	-	-	-	-
cortico-medullary border						
epithelial cells	+	-	-	-	-	-
capillary vessels	+[b]	-	-	-	-	-
medulla						
B cells	+	-	-	-	-	+
reticular epithelial cells	-	-	-	-	-	-
dendritic cells	-	-	-	-	-	-
isolated B cells	+/++	-	-	-	-	+/++

Signal intensity was semiquantitatively grouped into the categories: - negative, + weak but significant, ++ medium, +++ strong; [a]cytoplasmic, [b]endothelium, [c]supranuclear, [d]apical.

These experiments revealed different patterns of reactivity, as also summarized in Table 3. Weak binding of CG-1A was detected in the FAE/IFE (Fig. 8A), without apparent overlap to its immunolocalization, as documented by seeing no yellow signal in superimposition (Fig. 8B). In contrast, such an overlap was strong for CG-1B in the IFE but not in the FAE (Figs. 8C, D), and CG-2 served as negative specificity control (Figs. 8E, F). The two types of CG-3-dependent stainings led to marked overlap in IFE/FAE (Figs. 8G, H). Tested on isolated bursal B cells, application of none of these canonical galectins yielded a significant staining (not shown). B cells, positive for C-GRP and CG-8, are yet reactive with both C-GRP and CG-8, arguing in favor of a functional pairing in situ. Looking at mammalian B cells, the fact that they express also other types of lectins and bind them gives reason to add a monitoring of the glycophenotype to our study. As a relevant example, the siglec CD22 (siglec-2) or siglec-G, a B-cell-restricted surface protein in man and mouse, binds α2,6-sialylated N-glycan chains [1,44-48]. The emerging functionality of certain glycans by lectin recognition (for recent reviews, please see [49,50]) prompted us to perform lectin histochemistry, in addition to galectin histochemistry. As tools, plant and fungal lectins with specificities for common N- and O-glycans were selected, facilitating to map profiles of a series of biorelevant glycan epitopes given in Table 4 [51,52].

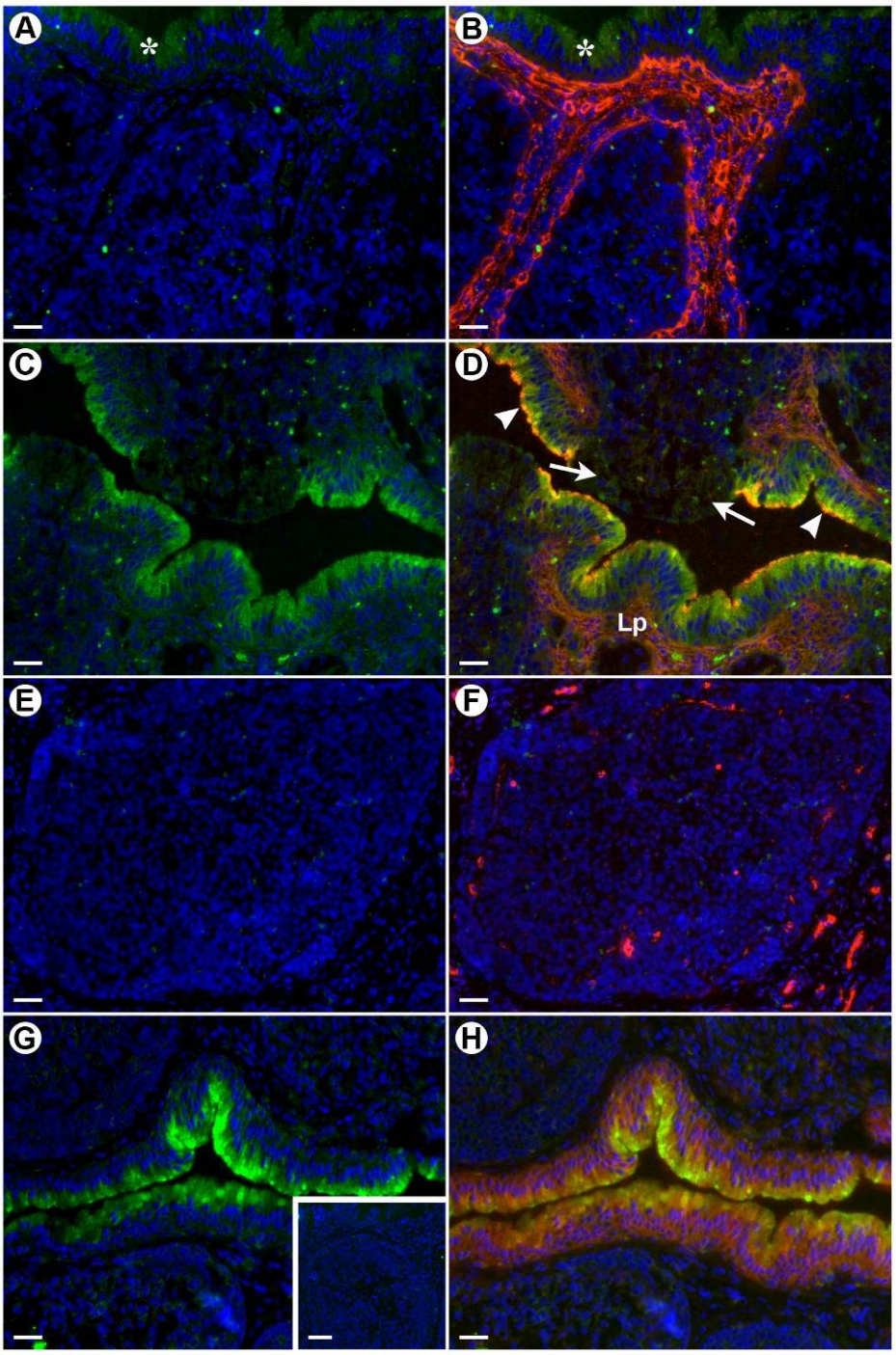


Fig. 8. Galectin histochemical staining using labeled CGs and its overlay with the immunohistochemical profile. CG-specific binding sites were visualized using ALEXA Fluor® 488-labeled CG-1A, CG-1B, CG-2 and CG-3 (A, C, E, G), double staining with respective immunohistochemical profile are given in (B, D, F, H). In the case of fluorescent CG-1A no overlap to CG-1A presence (please see also Fig. 4B) in basal lamina, in the endothelium of capillaries at the cortico-medullary border and between follicles, in the lamina propria and in mesenchymal reticular cells of the cortex was seen with CG-1A-binding sites, present supranuclearly (*asterisks*) in FAE/IFE (A, B). Fluorescent CG-1B (C) showed distinct overlap (in orange/yellow) apically in the IFE (*arrowheads*, D) and to a limited extent (orange) in the lamina propria (Lp). Weak reactivity for labeled CG-1B was seen in the FAE

(*arrows*). (E, F) CG-2 presence was found in capillary vessels, whereas no accessible binding sites were detected. (G, H) Binding sites for fluorescent CG-3 were detected supranuclearly in IFE/FAE (G; *inset*: control by inhibition with cognate sugar) leading to marked overlap (H). Concentrations of fluorescent CGs used were 10 μg/mL for CG-1A, 2.0 μg/mL for CG-1B, and 8.0 μg/mL for CG-3. Nuclei were counterstained with DAPI. The scale bars are 20 μm or 50 μm (*inset* to G).

The set of six plant/fungal lectins to map aspects of glycosylation was thus tested under identical conditions. Their glycan specificities, starting with α2,6-sialylated N-glycans as compiled in Table 4, enable to probe for presence of epitopes relevant for reactivity to galectins and other types of lectin. For example, α2,6-sialylation of N-glycans, besides being a CD22 ligand, abolishes galectin reactivity, unless repeats of N-acetyllactosamine are present (in the case of galectin-3) [53-55]. The lectin from the fungus *Polyporus squamosus* is exquisitely specific for this branch-end modification of N-glycans [16,56].

Table 4
Specificity profile of fungal/plant agglutinins used for glycophenotyping

Species	Abbreviation	Monosaccharide specificity	Potent glycan ligands
Polyporus squamosus (polypore mushroom)	PSL	[a]	Neu5Acα6Galβ4Glc(NAc) (over 300-fold more active than LacNAc, not reactive with free Neu5Ac); 6´-sulfation tolerated; 6'-sialyl T_n not reactive
Maackia amurensis-I (leukoagglutinin)	MAA-I	[a]	Neu5Acα3Galβ4GlcNAc/Glc, 3´-sulfation tolerated
Phaseolus vulgaris erythroagglutinin (kidney bean)	PHA-E	[a]	Bisected complex-type N-glycans: Galβ4GlcNAcβ2Manα6(GlcNAcβ2-Manα3)(GlcNAcβ4)Manβ4GlcNAc
Phaseolus vulgaris leukoagglutinin (kidney bean)	PHA-L	[a]	Tetra- and tri-antennary N-glycans with β6-branching
Artocarpus integrifolia (jack fruit)	Jacalin (JAC)	Gal/GalNAc	Galβ3GalNAcα; sialylation of T/ T_n antigens tolerated
Arachis hypogaea (peanut)	PNA	Gal	Galβ3GalNAcα/β

[a]no monosaccharide known as ligand.

The staining of B cells by the fungal lectin in sections and cell preparations, documented in Figs. 9A, G (for a detailed description of staining properties, please see legend to this figure), indicates presence of α2,6-sialylation. The MAA-I binding in the FAE, shown in Fig. 9B,

reveals presence of the branch-end isomer, i.e. α2,3-sialylated N-glycans. The N-terminal domain of CG-8 could target this epitope [57,58], and also mammalian siglecs such as sialoadhesin (siglec-1) target this epitope. Isolated B cells were negative (Fig. 9H). Complex-type N-glycans with core modifications, e.g. allowing branching to reach tri- and tetraantennary status or modulating lectin reactivity (Table 4), are also candidates for galectin binding, presence of these epitopes presenting regional specificity (Figs. 9C, D, I, J). Concerning core 1 O-glycans, α2,3-sialylation shows similar regional selectivity, and B cells lack this reactivity, as they also were shown not to bind MAA-I, which is specific for α2,3-sialylated (type II) N-glycans (Figs. 9E, H, K). The T antigen, in contrast, is present on B cells (Figs. 9F, L), indicating a regulation in the status of terminal glycosylation for mucin-type O-glycans. This aspect has been thoroughly studied for mammalian T cells, where changes occur upon activation [59]. As technical aspect, the illustrated combination of applying tissue and plant/fungal lectins reveals the chicken proteins to serve as versatile probes exactly as the commonly used plant proteins do, encouraging standard application.

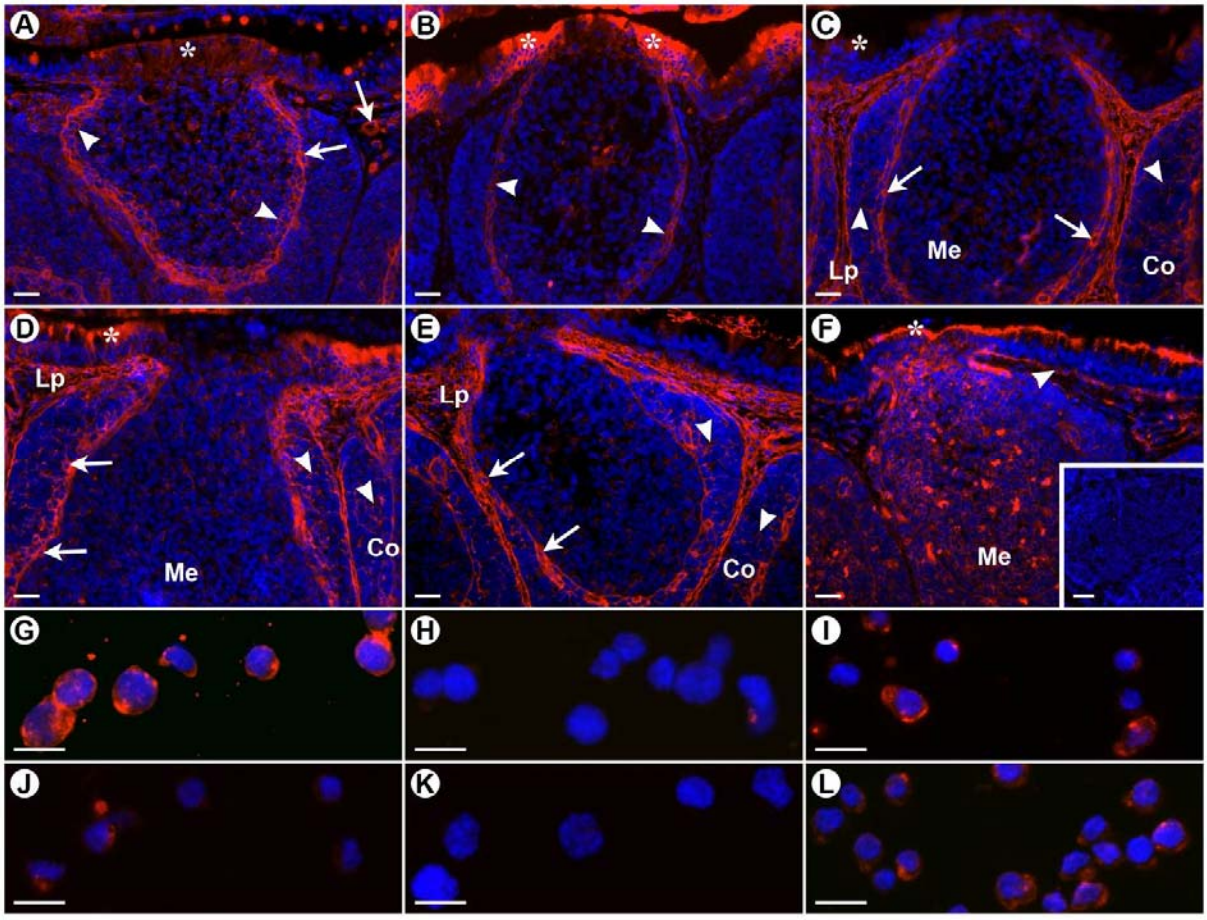


Fig. 9. Lectin histo- and cytochemical localization of selected glycan determinants applying biotinylated plant/fungal lectins and Texas-Red Avidin D® as second-step reagent in sections through follicles (A-F) and in isolated B cells of bursae of Fabricius (G-L) from 4-week-old chickens. (A, G) Staining by *Polyporus squamosus* lectin (PSL) in B cells (different stages of their lineage) of medulla and cortex, cortico-medullary epithelial cells (*arrowheads*), basal lamina and endothelial cells (vessels in the capillary layer at the cortico-medullary border and in the lamina propria, *arrows*). Comparatively reduced intensity of staining for PSL was found in the FAE (*asterisk*) (A). Isolated B cells were strongly positive (G). (B, H) Positivity for *Maackia amurensis-I* (leukoagglutinin) (MAA-I) was observed in cortico-medullary epithelial cells (*arrowheads*, B) and with strong intensity in the FAE (*asterisks*, B), whereas isolated B cells were negative (H). (C, I; D, J) Binding sites for *Phaseolus vulgaris* erythroagglutinin (PHA-E)/leukoagglutinin (-L) were found in B cells located in the medulla (Me), in endothelial cells (capillary layer at the cortico-medullary border, *arrows*), in the lamina propria (Lp), mesenchymal reticular cells (*arrowheads*) in the cortex (Co) and, especially for PHA-L, supranuclearly (*asterisk*) in the IFE (C, D). Isolated B cells showed a heterogeneous profile (I, J). (E, K) Reactivity for *Artocarpus integrifolia* agglutinin (jacalin, JAC) was mainly confined to the basal lamina and endothelial cells of the capillary layer at the cortico-medullary border (*arrows*, E), to the lamina propria (Lp) and mesenchymal reticular cells (*arrowheads*) in the cortex (Co), isolated B cells were negative (K). (F, L) Peanut agglutinin (PNA) positivity was found in B cells and, based on their morphological appearance, in bursal secretory dendritic cells, mainly of the medulla (Me), as well as in the intrafollicular capillary layer, also in the apical part (*asterisk*) and basal cell layer (*arrowhead*) of the IFE (F; inset: control by inhibition with cognate sugar). Isolated B cells were positive to variable extents (L). Concentrations of lectins used were 4.0 µg/mL for PHA-L, 2.0 µg/mL for PSL, 1.0 µg/mL

for MAA-I, JAC as well as PNA, and 0.5 µg/mL for PHA-E. Nuclei were counterstained with DAPI. The scale bars are 10 µm (G-L), 20 µm (A-F) or 50 µm (*inset* to F).

Considering reactivity of chicken B lymphocytes from blood to the canonical galectins CG-1A and CG-2 in flow cytofluorimetry [60] and the detection of C-GRP/CG-8 in the DT40 line (Fig. 6), probing for reactive sites with the chicken proteins was performed also in this cell system. Binding of C-GRP and the canonical CGs was saturable, and no difference was detectable when using C-GRP/C-GRC-C as probes (not shown). The observation that the N-terminal section is apparently not relevant for cell binding, as noted for staining of mammalian cells [8], may indicate a role of the tail in C-GRP routing, as reported for the N-terminal section and for its serine phosphorylation in the case of the chimera-type galectin-3 [61,62]. Similar functionality can be assumed in the cases of tailed proto-type *C. elegans* galectins, which have extensions of 37-50 amino acids at their C-termini [63,64]. As given in detail in the accompanying report [8], special care was taken to exclude a non-specific (dye-dependent) staining. Susceptibility of the signal (measured as percentage of positive cells/mean fluorescence intensity) to presence of lactose was only seen for the canonical CGs, here reaching a pronounced level of reduction (Fig. 10). The cultured malignant B cells, in contrast to isolated bursal B cells in cytochemistry, have obviously gained reactivity to canonical CGs. This may be linked to viral transformation. In fact, genetic alterations are known to affect the glycome [65]. When testing a series of glycocompounds, extent of GRP-C (and C-GRP-C) binding to the chicken cells was reduced by the presence of heparin, less by other proteoglycans or sulfated polysaccharides (Figs. 11A-D). Heparin activity to diminish cell surface binding was significant already at µg/mL concentrations (Fig. 11D). Remarkably, heparin also proved inhibitory to C-GRP binding in sections of chicken bursa and isolated B cells (Figs. 11E-H), whereas no effect on CG-8 binding was seen (not shown). These results imply differences in the biochemical nature of binding site(s) for C-GRP and CG-8. Also noteworthily, C-GRP binding to mammalian cells was not sensitive to heparin presence [8]. In contrast, binding to normal and malignant chicken B cells was, without an influence of the presence of the N-terminal extension of C-GRP. These data reveal a species-dependent difference and underscore to give preference to work in the homologous system. If secreted by a non-classical route, as is common for galectins [66], C-GRP may therefore exert functionality via specific cell surface binding.

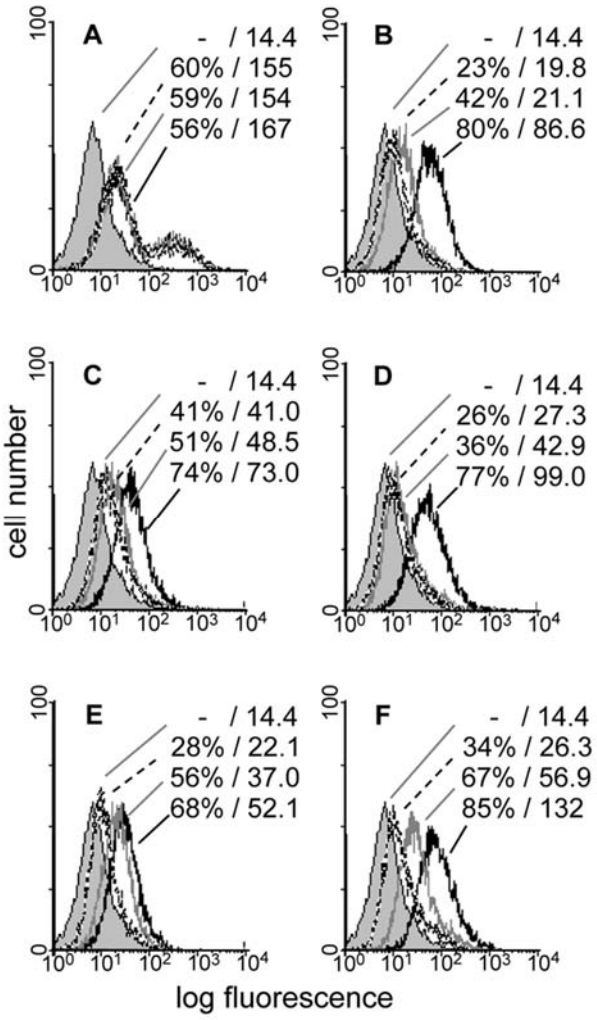


Fig. 10. Staining of cells of the chicken DT40 line with fluorescent C-GRP or canonical CGs in the absence (black line) or presence of 10 mM (grey line) and 100 mM (dashed line) lactose, respectively. The control value without lectin addition is given as shaded area. Panels recorded with aliquots of the same cell batch illustrate data for: (A) C-GRP (10 µg/mL), (B) CG-1A (2 µg/mL), (C) CG-1B (10 µg/mL), (D) CG-2 (10 µg/mL), (E) CG-3 (10 µg/mL) and (F) CG-8S (2 µg/mL).

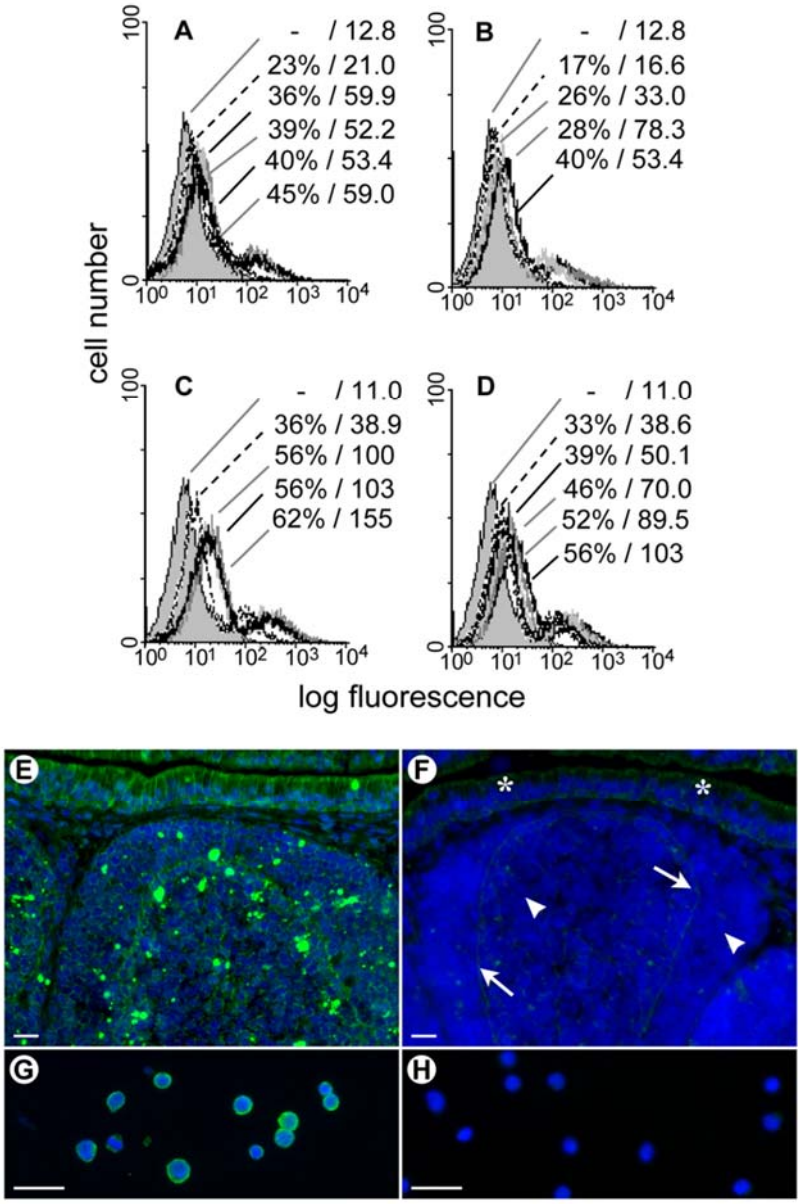


Fig. 11. Inhibition of cell surface binding of C-GRP (10 µg/ml, black line) to chicken DT40 cells by glycosaminoglycans and anionic sugar/polysaccharides. The control value without lectin addition is given as shaded area. Numbers characterizing staining (percentage of positive cells/mean fluorescence intensity) in each panel are always given in the order of the given listing from top to bottom. (A) Testing of fucoidan (1 mg/mL), D-glucuronic acid (50 mM), dextran sulfate (1 mg/mL) and dextran (1 mg/mL), (B) 0.25 mg/mL heparin, chondroitin sulfate and carrageenan, (C) same substances as in (B) at a concentration of 0.05 mg/mL, (D) concentration-dependent inhibition of C-GRP binding by decreasing concentrations of heparin (25, 12.5, 6 and 1.5 µg/mL). (E-H) Histochemical detection of accessible binding sites for fluorescent C-GRP in sections through a follicle and in isolated B cells in the absence (E, G) or presence (F, H) of heparin (50 µg/mL). Staining of follicle associated (E) and isolated (G) B cells by Alexa Fluor® 488 (green)-labeled C-GRP (5 µg/mL), was completely abolished in the presence of heparin both in tissue sections (F, *arrowheads*) and in isolated B cells (H), whereas IFE (F, *asterisks*) and epithelial cells of the cortico-medullary border (F, *arrows*) showed a strongly reduced but still significant signal intensity. Nuclei were counterstained with DAPI. The scale bars are 10 µm (G, H) or 20 µm (E, F).

4. Conclusions and perspectives

The high-level sequence conservation of GRP among vertebrates nourishes assumptions for a not yet defined functionally fundamental role. As a step toward its elucidation we here initiated the characterization of GRP distribution in vivo. At the same time, this work on C-GRP enabled to set this information in relation to the complete galectin network. Compared to all canonical CGs, C-GRP's profile of expression is special in being restricted to a single organ, that is to the bursa of Fabricius. Presence in and binding to B cells is shared by CG-8, an activity modulator of diverse blood cells such as plasma cells, T cells, megakaryocytes and neutrophils [43,67-69]. In view of the network concept for tissue lectins, the hypothesis of a functional cooperation between these two members of the galectin family deserves to be tested. That activated human T cells are responsive to three proto-type galectins via different mechanisms of caspase recruitment [70] and that a functional cooperation between galectins, then causing severe degeneration, appears to be operative in osteoarthritis [71,72] supports this assumption. Involvement of distinct counterreceptors is likely in these cases. Owing to the detection of α2,6-sialylated N-glycans attention should then be given, too, to characterizing expression of the siglec with respective specificity, i.e. CD22, for which an expressed sequence tag is known in chicken (EST 3109241). On the phylogenetic level, taking monitoring expression of GRP from chicken to mammals can lead to shape ideas for functional correlations in order to explain why this protein's sequence is kept constant in vertebrates to the extraordinary extent documented in [8].

Acknowledgements

Insightful discussions with Drs. B. Friday, M. Ilsüm and A. Leddoz and the valuable comments of the reviewers are greatly appreciated. The generous support by EC funding for the ITN Glycopharm (contract no. 317297) is gratefully acknowledged.

References

[1] 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.

[2] 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.

[3] D.N.W. Cooper, Galectinomics: finding themes in complexity, Biochim. Biophys. Acta 1572 (2002) 209-231.

[4] 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.

[5] 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.

[6] 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.

[7] 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.

[8] G. García Caballero, A. Flores-Ibarra, M. Michalak, N. Khasbiullina, N.V. Bovin, S. André, J.C. Manning, S. Vértesy, F.M. Ruiz, H. Kaltner, J. Kopitz, A. Romero, H.-J. Gabius, 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, submitted.

[9] C.A. Tregaskes, N. Bumstead, T.F. Davison, J.R. Young, Chicken B-cell marker chB6 (Bu-1) is a highly glycosylated protein of unique structure, Immunogenetics 44 (1996) 212-217.

[10] C.J. Rothwell, L. Vervelde, T.F. Davison, Identification of chicken Bu-1 alloantigens using the monoclonal antibody AV20, Vet. Immunol. Immunopathol. 55 (1996) 225-234.

[11] 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.

[12] 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.

[13] H.-J. Gabius, B. Wosgien, M. Hendrys, A. Bardosi, Lectin localization in human nerve by biochemically defined lectin-binding glycoproteins, neoglycoprotein and lectin-specific antibody, Histochemistry 95 (1991) 269-277.

[14] R. Kadirvelraj, O.C. Grant, I.J. Goldstein, H.C. Winter, H. Tateno, E. Fadda, R.J. Woods, Structure and binding analysis of *Polyporus squamosus* lectin in complex with the Neu5Acα2-6Galβ1-4GlcNAc human-type influenza receptor, Glycobiology 21 (2011) 973-984.

[15] 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.

[16] H. Mo, H.C. Winter, I.J. Goldstein, Purification and characterization of a Neu5Acα2-6Galβ1-4Glc/GlcNAc-specific lectin from the fruiting body of the polypore mushroom *Polyporus squamosus*, J. Biol. Chem. 275 (2000) 10623-10629.
[17] D. Kübler, C.-W. Hung, T.K. Dam, J. Kopitz, S. André, H. Kaltner, M. Lohr, J.C. Manning, L. He, H. Wang, A. Middelberg, C.F. Brewer, J. Reed, W.D. Lehmann, H.-J. Gabius, Phosphorylated human galectin-3: facile large-scale preparation of active lectin and detection of structural changes by CD spectroscopy, Biochim. Biophys. Acta 1780 (2008) 716-722.
[18] T.W. Baba, B.P. Giroir, E.H. Humphries, Cell lines derived from avian lymphomas exhibit two distinct phenotypes, Virology 144 (1985) 139-151.
[19] J.M. Buerstedde, S. Takeda, Increased ratio of targeted to random integration after transfection of chicken B cell lines, Cell 67 (1991) 179-188.
[20] 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.
[21] H. Kaltner, K. Seyrek, A. Heck, F. Sinowatz, H.-J. Gabius, Galectin-1 and galectin-3 in fetal development of bovine respiratory and digestive tracts. Comparison of cell type-specific expression profiles and subcellular localization, Cell Tissue Res. 307 (2002) 35-46.
[22] I. Saal, N. Nagy, M. Lensch, M. Lohr, J.C. Manning, C. Decaestecker, S. André, R. Kiss, I. Salmon, H.-J. Gabius, Human galectin-2: expression profiling by RT-PCR/immunohistochemistry and its introduction as histochemical tool for ligand localization, Histol. Histopathol. 20 (2005) 1191-1208.
[23] B. Kohnke-Godt, H.-J. Gabius, Heparin-binding lectin from human placenta: purification, partial molecular characterization and its relationship to basic fibroblast growth factors, Biochemistry 28 (1989) 6531-6538.
[24] K.C. Richardson, L. Jarett, E.H. Finke, Embedding in epoxy resins for ultrathin sectioning in electron microscopy, Stain Technol. 35 (1960) 313-323.
[25] S. André, S. Kojima, N. Yamazaki, C. Fink, H. Kaltner, K. Kayser, H.-J. Gabius, Galectins-1 and -3 and their ligands in tumor biology, J. Cancer Res. Clin. Oncol. 125 (1999) 461-474.
[26] M. Krzeminski, T. Singh, S. André, M. Lensch, A.M. Wu, A.M.J.J. Bonvin, H.-J. Gabius, Human galectin-3 (Mac-2 antigen): defining molecular switches of affinity to natural glycoproteins, structural and dynamic aspects of glycan binding by flexible ligand docking and putative regulatory sequences in the proximal promoter region, Biochim. Biophys. Acta 1810 (2011) 150-161.
[27] 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.
[28] K. Cartharius, K. Frech, K. Grote, B. Klocke, M. Haltmeier, A. Klingenhoff, M. Frisch, M. Bayerlein, T. Werner, MatInspector and beyond: promoter analysis based on transcription factor binding sites, Bioinformatics 21 (2005) 2933-2942.
[29] D. Maglott, J. Ostell, K.D. Pruitt, T. Tatusova, Entrez Gene: gene-centered information at NCBI, Nucleic Acids Res. 39 (2011) D52-57.
[30] T.H. Tran, M. Nakata, K. Suzuki, N.A. Begum, R. Shinkura, S. Fagarasan, T. Honjo, H. Nagaoka, B cell-specific and stimulation-responsive enhancers derepress Aicda by overcoming the effects of silencers, Nat. Immunol. 11 (2010) 148-154.
[31] K. Schneider, S. Kothlow, P. Schneider, A. Tardivel, T. Gobel, B. Kaspers, P. Staeheli, Chicken BAFF - a highly conserved cytokine that mediates B cell survival, Int. Immunol. 16 (2004) 139-148.

[32] S. Kothlow, I. Morgenroth, C.A. Tregaskes, B. Kaspers, J.R. Young, CD40 ligand supports the long-term maintenance and differentiation of chicken B cells in culture, Dev. Comp. Immunol. 32 (2008) 1015-1026.
[33] C.E. Sayegh, S.L. Demaries, K.A. Pike, J.E. Friedman, M.J. Ratcliffe, The chicken B-cell receptor complex and its role in avian B-cell development, Immunol. Rev. 175 (2000) 187-200.
[34] J. Korte, T. Fröhlich, M. Kohn, B. Kaspers, G.J. Arnold, S. Härtle, 2D DIGE analysis of the bursa of Fabricius reveals characteristic proteome profiles for different stages of chicken B-cell development, Proteomics 13 (2013) 119-133.
[35] S.K. Reddy, T. Hu, R. Gudivada, K.A. Staines, K.E. Wright, L. Vickerstaff, S. Kothlow, L.G. Hunt, C. Butter, B. Kaspers, J.R. Young, The BAFF-Interacting receptors of chickens, Dev. Comp. Immunol. 32 (2008) 1076-1087.
[36] H. Kaltner, T. Singh, J.C. Manning, A.-S. Raschta, S. André, F. Sinowatz, H.-J. Gabius, Network monitoring of adhesion/growth-regulatory galectins: localization of the five canonical chicken proteins in embryonic and maturing bone and cartilage and their introduction as histochemical tools, Anat. Rec. 298 (2015) 2051-2070.
[37] J. Wang, Z.H. Lu, H.-J. Gabius, C. Rohowsky-Kochan, R.W. Ledeen, G. Wu, Cross-linking of GM1 ganglioside by galectin-1 mediates regulatory T cell activity involving TRPC5 channel activation: possible role in suppressing experimental autoimmune encephalomyelitis, J. Immunol. 182 (2009) 4036-4045.
[38] S. André, H. Kaltner, J.C. Manning, P.V. Murphy, H.-J. Gabius, Lectins: getting familiar with translators of the sugar code, Molecules 20 (2015) 1788-1823.
[39] F.-T. Liu, R.J. Patterson, J.L. Wang, Intracellular functions of galectins, Biochim. Biophys. Acta 1572 (2002) 263-273.
[40] J.L. Wang, R.M. Gray, K.C. Haudek, R.J. Patterson, Nucleocytoplasmic lectins, Biochim. Biophys. Acta 1673 (2004) 75-93.
[41] N. Miyanishi, N. Nishi, H. Abe, Y. Kashio, R. Shinonaga, S. Nakakita, W. Sumiyoshi, A. Yamauchi, T. Nakamura, M. Hirashima, J. Hirabayashi, Carbohydrate-recognition domains of galectin-9 are involved in intermolecular interaction with galectin-9 itself and other members of the galectin family, Glycobiology 17 (2007) 423-432.
[42] D. Solís, M.J. Maté, M. Lohr, J.P. Ribeiro, L. López-Merino, S. André, E. Buzamet, F.J. Cañada, H. Kaltner, M. Lensch, F.M. Ruiz, G. Haroske, U. Wollina, M. Kloor, J. Kopitz, J.L. Sáiz, M. Menéndez, J. Jiménez-Barbero, A. Romero, H.-J. Gabius, N-Domain of human adhesion/growth-regulatory galectin-9: preference for distinct conformers and non-sialylated *N*-glycans and detection of ligand-induced structural changes in crystal and solution, Int. J. Biochem. Cell Biol. 42 (2010) 1019-1029.
[43] C.M. Tsai, C.H. Guan, H.W. Hsieh, T.L. Hsu, Z. Tu, K.J. Wu, C.H. Lin, K.I. Lin, Galectin-1 and galectin-8 have redundant roles in promoting plasma cell formation, J. Immunol. 187 (2011) 1643-1652.
[44] L.D. Powell, R.K. Jain, K.L. Matta, S. Sabesan, A. Varki, Characterization of sialyloligosaccharide binding by recombinant soluble and native cell-associated CD22. Evidence for a minimal structural recognition motif and the potential importance of multisite binding, J. Biol. Chem. 270 (1995) 7523-7532.
[45] T. Angata, E.C.M. Brinkman-van der Linden, I-type lectins, Biochim. Biophys. Acta 1572 (2002) 294-316.
[46] N. Kimura, K. Ohmori, K. Miyazaki, M. Izawa, Y. Matsuzaki, Y. Yasuda, H. Takematsu, Y. Kozutsumi, A. Moriyama, R. Kannagi, Human B-lymphocytes express α2-6-sialylated 6-sulfo-N-acetyllactosamine serving as a preferred ligand for CD22/Siglec-2, J. Biol. Chem. 282 (2007) 32200-32207.
[47] J.A. Walker, K.G. Smith, CD22: an inhibitory enigma, Immunology 123 (2008) 314-325.

[48] J.C. Poe, T.F. Tedder, CD22 and Siglec-G in B cell function and tolerance, Trends Immunol. 33 (2012) 413-420.
[49] H.-J. Gabius, The magic of the sugar code, Trends Biochem. Sci. 40 (2015) 341.
[50] H.-J. Gabius, J.C. Manning, J. Kopitz, S. André, H. Kaltner, Sweet complementarity: the functional pairing of glycans with lectins, Cell Mol. Life Sci. 73 (2016) 1989-2016.
[51] A. Danguy, F. Akif, B. Pajak, H.-J. Gabius, Contribution of carbohydrate histochemistry to glycobiology, Histol. Histopathol. 9 (1994) 155-171.
[52] J. Roth, Lectins for histochemical demonstration of glycans, Histochem. Cell Biol. 136 (2011) 117-130.
[53] C. Sparrow, H. Leffler, S.H. Barondes, Multiple soluble β-galactoside-binding lectins from human lung, J. Biol. Chem. 262 (1987) 7383-7390.
[54] N. Ahmad, H.-J. Gabius, H. Kaltner, S. André, I. Kuwabara, F.-T. Liu, S. Oscarson, T. Norberg, C.F. Brewer, Thermodynamic binding studies of cell surface carbohydrate epitopes to galectins-1, -3 and -7. Evidence for differential binding specificities, Can. J. Chem. 80 (2002) 1096-1104.
[55] J. Hirabayashi, T. Hashidate, Y. Arata, N. Nishi, T. Nakamura, M. Hirashima, T. Urashima, T. Oka, M. Futai, W.E.G. Müller, F. Yagi, K.-i. Kasai, Oligosaccharide specificity of galectins: a search by frontal affinity chromatography, Biochim. Biophys. Acta 1572 (2002) 232-254.
[56] V. Toma, C. Zuber, H.C. Winter, I.J. Goldstein, J. Roth, Application of a lectin from the mushroom *Polysporus squamosus* for the histochemical detection of the NeuAcα2,6Galβ1,4Glc/GlcNAc sequence of N-linked oligosaccharides: a comparison with the *Sambucus nigra* lectin, Histochem. Cell Biol. 116 (2001) 183-193.
[57] E.M. Rapoport, V.K. Matveeva, H. Kaltner, S. André, O.A. Vokhmyanina, G.V. Pazynina, V.V. Severov, I.M. Ryzhov, E.Y. Korchagina, I.M. Belyanchikov, H.-J. Gabius, N.V. Bovin, Comparative lectinology: delineating glycan-specificity profiles of the chicken galectins using neoglycoconjugates in a cell assay, Glycobiology 25 (2015) 726-734.
[58] 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.
[59] M.C. Clark, L.G. Baum, T cells modulate glycans on CD43 and CD45 during development and activation, signal regulation, and survival, Ann. N. Y. Acad. Sci. 1253 (2012) 58-67.
[60] M. Schneller, S. André, J. Cihak, H. Kaltner, H. Merkle, G.J. Rademaker, J. Haverkamp, J. Thomas-Oates, U. Lösch, H.-J. Gabius, Differential binding of two chicken β-galactoside-specific lectins to homologous lymphocyte subpopulations and evidence for inhibitory activity of the dimeric lectin on stimulated T cells, Cell. Immunol. 166 (1995) 35-43.
[61] H.C. Gong, Y. Honjo, P. Nangia-Makker, V. Hogan, N. Mazurak, R.S. Bresalier, A. Raz, The NH_2 terminus of galectin-3 governs cellular compartmentalization and functions in cancer cells, Cancer Res. 59 (1999) 6239-6245.
[62] R.P. Menon, R.C. Hughes, Determinants in the N-terminal domains of galectin-3 for secretion by a novel pathway circumventing the endoplasmic reticulum-Golgi complex, Eur. J. Biochem. 264 (1999) 569-576.
[63] J. Hirabayashi, Y. Arata, K. Hayama, K.-i. Kasai, Galectins from the nematode *Caenorhabditis elegans* and the glycome project, Trends Glycosci. Glycotechnol. 13 (2001) 533-549.
[64] Y. Nemoto-Sasaki, K. Hayama, H. Ohya, Y. Arata, M.K. Kaneko, N. Saitou, J. Hirabayashi, K.-i. Kasai, *Caenorhabditis elegans* galectins LEC-1-LEC-11: structural features and sugar-binding properties, Biochim. Biophys. Acta 1780 (2008) 1131-1142.

[65] H.-J. Gabius, Glycobiomarkers by glycoproteomics and glycan profiling (glycomics): emergence of functionality, Biochem. Soc. Trans. 39 (2011) 399-405.
[66] R.C. Hughes, Secretion of the galectin family of mammalian carbohydrate-binding proteins, Biochim. Biophys. Acta 1473 (1999) 172-185.
[67] N. Nishi, H. Shoji, M. Seki, A. Itoh, H. Miyanaka, K. Yuube, M. Hirashima, T. Nakamura, Galectin-8 modulates neutrophil function via interaction with integrin aM, Glycobiology 13 (2003) 755-763.
[68] M.V. Tribulatti, M.G. Figini, J. Carabelli, V. Cattaneo, O. Campetella, Redundant and antagonistic functions of galectin-1, -3, and -8 in the elicitation of T cell responses, J. Immunol. 188 (2012) 2991-2999.
[69] C. Zappelli, C. van der Zwaan, D.C. Thijssen-Timmer, K. Mertens, A.B. Meijer, Novel role for galectin-8 protein as mediator of coagulation factor V endocytosis by megakaryocytes, J. Biol. Chem. 287 (2012) 8327-8335.
[70] A. Sturm, M. Lensch, S. André, H. Kaltner, B. Wiedenmann, S. Rosewicz, A.U. Dignass, H.-J. Gabius, Human galectin-2: novel inducer of T cell apoptosis with distinct profile of caspase activation, J. Immunol. 173 (2004) 3825-3837.
[71] 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.
[72] S. Toegel, D. Weinmann, S. André, S.M. Walzer, M. Bilban, S. Schmidt, C. Chiari, R. Windhager, C. Krall, I.M. Bennani-Baiti, H.-J. Gabius, Galectin-1 couples glycobiology to inflammation in osteoarthritis through the activation of an NF-κB-regulated gene network, J. Immunol. 196 (2016) 1910-1921.

5. DISCUSSION

Two non-canonical galectins, namely C-GRP and C-GRIFIN, complete the galectin protein family in chicken. Thorough database mining (NCBI nucleotide BLAST browser, Ensembl genome browser) revealed no further ESTs, partial or complete nucleotide sequences (on mRNA and genomic level) encoding further galectin-like or galectin-related proteins in chicken. The amino acid sequences of GRP are highly conserved in vertebrates (18 amino acid exchanges between chicken and man, only one amino acid exchange between mouse and man, please see phylogenetic analysis in Publication 2) and, due to deviations from the key residues of the characteristic sequence signature of galectins (two to five depending on the species; chicken possesses four deviations), sugar binding is abolished (loss of hydrogen bonds to 6´-hydroxyl- and 4´-hydroxyl-group as present in D-galactose). Such a loss is also reported for rat GRIFIN [123] and human galectin-10 (Charcot-Leyden crystal protein) [121]. Interestingly, and in contrary to canonical galectins, GRP was not found in the genomes of invertebrates.

Despite the presence of mRNA detected from total RNA isolated from a series of embryonic and adult chicken organs (heart, kidney, small intestine and thymus) the protein was only detected by Western Blotting and localized immunohistochemically (visualization with fluorophores) in sections of the bursa of Fabricius (please see Publication 3). In spite of the lack of binding to lactose (affinity chromatography on a lactose-Sepharose 4B gel matrix is the crucial step in the purification of recombinant canonical galectins) – the purification of recombinant C-GRP depends on a GST-Tag procedure – and the deviations of strictly conserved amino acid residues in the contact site which result in the loss of characteristic hydrogen bonding, we were able to inhibit binding of fluorescently labeled C-GRP to the cell surface of the B-cell line DT40 (a chicken B cell line derived from an avian leucosis virus (ALV)-induced bursal lymphoma [128]) by the glycosaminoglycan (GAG) heparin (a polymer consisting of repetitions of the basic disaccharide unit [-4GlcNAcα1-4GlcAβ1-]n, with diverse presence of sulfations, among them, the rare 3-O-sulfation [129].

To get a proper idea about the mode of interaction, one of the next experiments could be the tryptic fragmentation of C-GRP bound to heparin which is for its part coupled to Sepharose 4B (affinity matrix) with subsequent identification of heparin-binding peptides by mass-spectrometric identification [130]. Inhibition by heparin in FACScan and lectin cytochemical analysis does not necessarily mean that this GAG is a counterreceptor. It may, for example, associate to a positively charged and (somehow) matching region of C-GRP and cover the real site for counterreceptor interaction by spatial blocking not competitive inhibition. Noticeably,

the purification of a heparin-binding lectin [131] from human placenta revealed a protein whose properties were not in line with galectins. Interactions with GAGs are not unique for C-GRP. The potential of GAGs to operate as ligands for galectins was already assessed using the technique of frontal affinity chromatography [129,132]. Galectins (especially -3, -7 and -9) bound to 6´-O-desulfated β-Gal/GalNAc containing GAGs, such as keratan sulfate, chondroitin and dermatan presenting the typical hydroxyl groups constellation (4´-OH, 6´-OH from Gal/GalNAc, 3-OH from Glc/GlcNAc). No retardation with resin-immobilized galectins was seen with N-acetylheparosan and hyaluronan, which lack the hydroxyls' display.

The sequence alterations in GRPs can also have other consequences: the shape of the binding site changes in a way that other sugars than β-galactosides act as bioactive ligands. This was shown for human galectin-10 (four deviations in key residues of the contact site) with mannose as ligand [121] and for the fungal (*Coprinopsis cinerea*) galectin (-like protein) CGL3 (R81W), in which the central tryptophan was changed to arginine (as it is in mammalian GRP), where binding activities were proven for chitooligosaccharides or also for LacdiNAc [133].

Only for C-GRP, localization and expression in the epithelium and intracellularly in B cells of the bursa of Fabricius and of the B cell line DT40 could be clearly described. The next step should be an immunocyto- and histochemical study on human material (tumor and carcinoma cell lines; sections of cancer specimen) as done for C-GRP (surprisingly, sequence conservation did not ensure immunological cross-reactivity: so we have to start raising an antibody to human GRP), it is first necessary to see GRP's expression profile in mammalian tissue. GRP, being monomeric in solution such as rat galectin-5, may thus have a special function at a special site such as this rat protein has for exosomal sorting and routing during rat reticulocyte maturation [134]. The biochemical nature of the underlying recognition is a fully open issue.

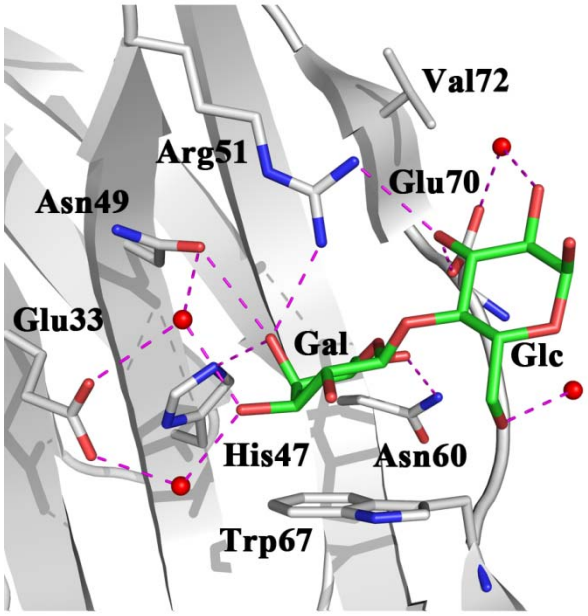


Fig. 14: Detailed part of the CRD of C-GRIFIN (provided by F. Ruiz and A. Romero, Madrid, Spain). Hydrogen bonds are depicted in magenta dashed lines.

In C-GRIFIN (Fig. 14) which is – on the contrary to GRIFIN in mammals – a β-galactoside-binding lectin, valine replaces the key residue arginine at position 72 (Arg72Val) (Fig. 14). Compared to chicken galectin-2 [20] this substitution results in a loss of hydrogen bonding to the 3-hydroxylgroup of Glc/GlcNAc. Obviously, hydrogen bonding with Arg51 and Glu70 (Fig. 14) is sufficient to fix energy-minimized conformers of the ligand in the contact site. Why do mammalian GRIFINs not bind to a sugar ligand? Here, an additional exchange at position 49 from asparagine to lysine exists. An explanation could be, that the binding of lactose is prevented through spatial hindrance by the lysine side chain impairing the formation of hydrogen bonding to the 4`-hydroxyl group of galactose.

The only site of localization of C-GRIFIN (also mammalian GRIFINs) is the eye lense, as seen by RT-PCR, Western blotting and immunohistochemistry, the latter showing a characteristic expression pattern different from the ones obtained when localizing the canonical galectins in this tissue. This highly restricted presence of GRIFIN, confined to a single tissue type, indicates that expression of GRIFIN genes should be under a strict regulation mechanism.

The function of these non-canonical galectins is far from clear. As galectins exert their effects after pairing with counterreceptors (please see Table 1 in section 2.4), a reasonable approach to obtain an indication of the functional capacities of C-GRP and C-GRIFIN is to purify and structurally characterize bioactive counterreceptors. Initially, recombinant protein (high amounts of both proteins can be isolated from *E. coli* cultures) can be coupled with high density to gel matrices. Thus, we will be able to selectively enrich (glyco)proteins from extracts

of bursal B cells and of cells of the B cell line DT40 (in the case of C-GRP) or from extracts of embryonic/adult lenses (in the case of C-GRIFIN). Also, magnetic separation of C-GRP/C-GRIFIN-associated complexes by tosyl-activated dynabeads coated with recombinant C-GRP or C-GRIFIN can be a valuable option to serve this purpose. (Glyco)proteins with reactivity to C-GRP or C-GRIFIN will be separated by two-dimensional gel electrophoresis and identified by mass spectrometric fingerprinting (including Orbitrap analysis [46]).

When performing the search on bioactive counterreceptors for GRIFIN and GRP we have to consider that our approach should also address the question whether both proteins are engaged in protein-protein interaction. That applies all the more so since both proteins displayed intracellular (cytoplasmic) expression in line with other e.g. galectin-1/-3 which cooperate intracellularly in tumor proliferation with cytoplasmic proteins such as oncogenic H-/K-Ras [135, 136] or in spliceosome assembly in vivo with nuclear Gemin4 [109] (please see also Table 1 in section 2.4). More recently, it was shown that galectin-8 recruits NDP52, an autophagy cargo receptor [137, 138] and that galectin-3 associates with Alix forming a complex which negatively regulates $CD4^+$ T-cell activation [139]. A more in depth functional characterization of both proteins could be achieved when *in silico* performed analyses to map transcription factor binding sites in the nucleotide sequences of the promoter sections (please see Publications 1 and 3) will be implemented in experimental approaches such as luminescence-based reporter assays and/or chromatin immunoprecipitation (ChIP) experiments. First results, obtained performing reporter assays, in the case of the C-GRIFIN's promoter sequence showed that lens-specific transcription factors such as MafB and L-Maf direct the transcription of C-GRIFIN (personal communication by Dr. Sebastian Schmidt), presumably in an early period of lens development. This is supported by recent detection of presence of C-GRIFIN in the developing lens of 3-day chicken embryos. L-Maf [140] activates downstream in the process of lens development, c-MAF (early in lens development) and MafB (later in lens development, in differentiating lens fibers) and the genes for crystalline proteins [141]. In turn, the crystalline proteins interact presumably with C-GRIFIN, CG-1A as well as with CG-3 (both canonical CGs expressed in lens), as part of a network whose components have lens-specific transcription factor binding sites in common.

In chicken, only up to five canonical and two non-canonical galectins can participate in such networks, whereas in man evidently 16 galectin genes [142], several gene duplications/triplications [80] and splice variants [19,80] arrange in more complex network which makes it difficult to unravel the functions of the individual galectins. The comparatively small avian "galectinome" enables comprehensive galectin and immunohistochemical mapping in

embryogenesis to connect galectins to developmental processes in a non-mammalian organism. Of note, recent observations of the involvement of the avian galectin-1 ortholog (CG-1A) in cell self-organizing dynamics within cartilage maturation [143] and limb skeletal morphogenesis, together with spatially and temporally restricted galectin reactivity, supports the concept of this CGs' network involved in developmental processes. CG-1A/CG-8 reciprocally increase their expression, CG-1A promotes pre-cartilage condensations, CG-8 antagonizes it. Monitoring the CGs' network in progressing limb development (beyond developmental day five) additionally, CG-1B/-3 became part of the network [80] in these tissues (long bone, cartilage).

Not only in embryonic/fetal development but also in the development and progression of diseases galectin networks appear to be involved. Exemplarily, networks contribute to the regulation of processes driving cancer growth and progression such as adhesion (pro-/anti-adhesive galectins), homotypic aggregation of tumor cells, metastasis [144], invasion, angiogenesis [145] and apoptosis. Main features of such an orchestrated cooperation in these processes are:

(i) More than one galectin participates, the predominant galectin depends on tumor type and stage (please see [146] for review). Also it must be defined whether the interactions between galectins are synergistic or antagonistic. The main task in preclinical and clinical studies is to perform galectin-fingerprinting and then to dissect the role of each individual galectin (canonical/non-canonical) in the network. For example, in prostate cancer, galectin-3 strongly influences metastasis and angiogenesis, galectin-8 controls anoikis due to a reorganization of the cytoskeleton and supports homotypic aggregation of prostate cancer cells, thus ensuring the survival of the cancer cells in the blood circulation, an important requirement for invasion and forming of metastases [147]. A further example for a galectin network was unraveled in several clinical studies with patients suffering from hepatocellular carcinoma. Overexpressing galectin-1 [148] and/or nuclear galectin-3 [149] led to a poor prognosis and the overall survival was definitely reduced, whereas upregulation of galectin-9 in the same tumor type served as positive prognostic factor with a comparably enhanced survival rate [150].

(ii) Glycans presented by counterreceptors often act as molecular switches for galectin reactivity, e. g. α-2,6 sialylation functions as a switch <u>off</u> signal in the pancreatic carcinoma cell-line Capan-1 reconstituted with the tumor suppressor $p16^{INK4a}$ and in the glioblastoma cell line U87, preventing binding of galectin-1, which acts either pro-apoptotic (Capan-1 cells) or pro-invasive (U87 cells) (please see section 2.4).

(iii) Intracellular pathways participate in these networks either by outside-inside signaling triggered through association of galectins with surface-associated counterreceptors and/or cytoplasmic/nuclear interaction of galectins with molecules of such activated pathways [135,151]. Genetically engineered model cell lines with glycomic profiles deliberately shifted in degree of sialylation either resistant to pro-malignant galectin properties or susceptible to pro-apoptotic galectin activities will be valuable tools to perform experiments in this context.

The involvement of galectins in diseases as described above and in section 2.4 raises the question on possible therapeutic options to inhibit the activities of galectins. Genetic knock-down (e.g. shRNA, CRISPR/Cas9) of e.g. galectin-1 is a possible approach as shown in the case of human glioblastoma (U87) cells when interactions of the cells with galectin-1 promotes invasiveness, and in clinical terms account for a poor patients' prognosis. Up-regulation of ST6Gal-I (please see section 2.4) in the same cell line by genetic engineering in order to enhance α2,6-sialylation may result in a reduced susceptibility to galectin-1 and a loss of invasiveness. Here, the influence of α2,6-sialylation as switch-off signal for galectin binding as potential therapeutic approach to prevent invasiveness will be defined. Further possibilities to prevent binding of galectins to surface-associated counterreceptors may be application of anti-galectin antibodies, galectin-derived blocking peptides or defective galectin variants which cannot elicit post-binding effects. Employing tailor-made synthetic inhibitory bi- to oligovalent glycoclusters with lactose as headgroups [152] is a further option to interfere with galectin/carbohydrate pairing. In a preclinical histochemical test system, resembling physiological conditions, sections of paraffin-embedded tissues were used to test the inhibitory potencies of such glycoclusters taking as a measure the reduction of staining intensities [153]. Remarkably, besides genetic engineering, an appealing idea to get new therapeutics is to use glycodendrimersomes (GDs) (please see section 2.4). Applying GDSs with appropriate sugar headgroups accordingly, they can be used as nanoparticle-sized decoy receptor to scavenge (harmful) galectin, e.g. in osteoarthritis, where Gal-1 is a driving force of the pathogenesis by inducing a pro-degradative/inflammatory gene signature [154] or in the above described examples of cancer. Loading the dendrimers with sugar ligands with special affinity to a galectin, e.g. LacdiNAc or A-tetrasaccharide, will obviously make scavenging selective. Further preclinical and clinical studies are undoubtedly necessary to ascertain future therapeutic application of these procedures.

6. SUMMARY

Chicken (*Gallus gallus*) as Model for Network Analysis of Adhesion-/Growth-regulatory Galectins.

Biochemical characterization of C-GRIFIN/C-GRP and first complete histochemical analysis for the galectin family in bursa of Fabricius

Lectins (carbohydrate-binding proteins) translate glycan-encoded information of cellular glycoconjugates into effects, hereby being involved in many physiological and pathological processes. A class of adhesion-/growth-regulatory lectins, the galectins, share binding specificity for β-galactosides, β-sandwich folding and key amino acid residues in the contact site (including a central tryptophan), forming a characteristic sequence signature. In addition to their carbohydrate specificity, the topological mode of presentation of their carbohydrate recognition domain(s) (CRD) is assumed to play a key role for their bioactivities. Three types of structural design are known, classified as proto-type (homodimeric), chimera-type (CRD in combination with an N-terminal proline-rich tandem repeat domain) and tandem-repeat-type (two CRDs connected by a linker peptide). To proceed to a comprehensive study of galectin expression, a model with complete representation of these types of design and a minimal number of individual proteins is most suited. Chicken fulfills the given prerequisites with a set of five canonical galectins. In detail, it comprises three proto-type chicken galectins (CGs) (the paralogue pair CG-1A/B and CG-2), chimera-type CG-3 and tandem-repeat-type CG-8. Moreover, on the genomic level we found (partial) sequences of two non-canonical (deviations in the sequence signature) members, not yet characterized, namely C-GRIFIN and C-GRP. Both proteins are encoded by a single copy gene, are highly conserved and their presence is restricted to vertebrates. Recombinant expression of both proteins, mass spectrometric (MS) analysis and gel filtration identified C-GRIFIN as a homodimeric, C-GRP as a monomeric protein. In the case of C-GRIFIN, the key amino acid residue in the contact site arginine at position 72 was replaced by valine. In spite of this exchange, C-GRIFIN is able to bind β-galactosides, especially histoblood group B-tetrasaccharide. Two lactose binding peptides after tryptic digestion on lactose-Separose 4B beads were identified by MS. Analysis of mRNA and protein expression demonstrated presence restricted to the lens, immunohistochemically the protein was detected in the semicircular fibers of the main lens body. C-GRP exhibited four deviations (e.g. asparagine at position 51 was exchanged by lysine) and did not bind β-galactosides, the crystal structure indicates that the side chain of lysine at position 51 prevents a snug fit of

galactose in the binding pocket. RT-PCR demonstrated the occurrence of C-GRP's mRNA in numerous tissues whereas protein bands in Western blots were only observed in tissue extracts from bursa of Fabricius (also immunopositive for the five canonical CGs). Histochemically, C-GRP presence was limited to bursa of Fabricius and localized in bursal B cells, epithelial cells of the interfollicular epithelium (IFE)/follicle-associated epithelium (FAE) and blood vessels. The full set of canonical galectins could be localized in the bursa of Fabricius: CG-1A in basal lamina and vessels, CG-1B in lamina propria and IFE, CG-2 in endothelium of capillaries, CG-3 in epithelial cells, CG-8 in medullary and cortical B cells. C-GRP presence overlaps with CG-1A (endothelium of vessels), CG-1B (IFE), CG-3 (IFE/FAE) and CG-8 (B cells). The binding of fluorescently labeled C-GRP to B cells can be blocked by heparin, not necessarily meaning that the contact site(s) is the putative CRD. Future perspectives are arising now from the availability of the full set of canonical/non-canonical galectins in chicken, allowing us to proceed with network analyses in embryonic development. The isolation of bioactive counterreceptors for human/chicken GRIFIN/GRP is mandatory to clearly define functions and to clarify their involvement in physiological and pathological processes.

7. ZUSAMMENFASSUNG

Analyse eines Netzwerks von Adhäsions- und Wachstum-regulierender Galektine im Modellsystem Huhn (*Gallus gallus*).

Biochemische Charakterisierung von C-GRIFIN/C-GRP und erste vollständige histochemische Analyse einer Galektinfamilie in der Bursa Fabricii

Lektine sind kohlenhydratbindende Proteine, die die in Glykanen zellulärer Glykokonjugate festgelegten Informationen dazu nutzen, um eine Vielzahl physiologischer und pathologischer Vorgänge zu beeinflussen. Eine Klasse dieser Lektine, stellen die Galektine, die Adhäsion und Wachstum regulieren, dar. Sie haben die spezifische Bindung von β-Galaktosiden, die β-sandwichartige Anordnung von 5-6 β-Faltblättern, und eine charakteristische Anordnung bestimmter Aminosäurereste in der Bindungsstelle (Sequenzsignatur) gemeinsam. Dabei ist nicht nur die Bindung zu spezifischen Zuckern von Bedeutung, sondern auch die Anordnung der kohlenhydratbindenden Domänen (CRD). Die drei möglichen Anordnungen der CRDs sind der Proto-Typ (Homodimer), der Chimera-Typ (CRD kombiniert mit N-terminaler Prolin-reicher Tandem-Repeat Domäne), und der Tandem-Repeat-Typ (zwei CRDs verbunden durch einen Linkerpeptid). Um eine umfassende Studie zur Expression von Galektinen anzufertigen wird ein Modell benötigt, das die komplette Anzahl der möglichen Anordnungen der CRDs mit einer möglichst geringen Anzahl individueller Proteine vereint. Das Huhn erfüllt die entsprechenden Voraussetzungen. Im Hühnergenom finden sich die Gene für lediglich fünf kanonische Galektine. Im einzelnen sind drei Galektine vom Proto-Typ (die paralogen Proteine CG-1A/1B und CG-2), jeweils eines vom Chimera-Typ (CG-3) und eines vom Tandem-Repeat-Typ (CG-8) vorhanden. Darüber hinaus konnten auf genomischer Ebene Sequenzabschnitte von zwei noch nicht charakterisierten nicht-kanonischen (Abweichungen bei den Aminosäuren, die für die Sequenzsignatur relevant sind) Galektinen, ermittelt werden.

Beide Proteine werden jeweils von einem single-copy-Gen kodiert, sind hoch konserviert und kommen nur bei Vertebraten vor. Rekombinante Expression beider Proteine, Massenspektrometrie und analytische HPLC-Gelfiltration identifizierten C-GRIFIN als homodimeres, C-GRP als monomeres Protein. Im Falle von C-GRIFIN liegt ein Austausch der Aminosäure Arginin gegen Valin an Position 72 vor. Trotz dieser Abweichung kann C-GRIFIN β-Galaktoside, insbesonders das Blutgruppen-Tetrasaccharid B binden. Mithilfe der massen-spektrometrischen Analyse konnten zwei Lactose-bindende Peptide identifiziert werden. Die

Untersuchung der mRNA- und der Proteinexpression zeigte, dass C-GRIFIN lediglich in der Augenlinse vorhanden ist. Immunhistochemisch konnte das Protein in den halbkreisförmigen Fasern des Linsenkörpers nachgewiesen werden. Bei C-GRP sind vier Abweichungen von der Sequenzsignatur der Galektine (u.a. ein Austausch von Asparagin durch Lysin an Position 51) festzustellen. Das Protein ist nicht β-Galaktosid-bindend. Die Kristallstrukturanalyse zeigte, dass die hervorstehende Seitenkette des Lysinrestes an Position 51 verhindert, dass Galaktose in die Bindungsstelle eingepasst werden kann. Die RT-PCR Analyse ergab, dass die spezifische mRNA weitverbreitet in zahlreichen Geweben vorkommt, der Proteinnachweis gelang jedoch nur in Organextrakten der Bursa Fabricii (ebenso konnte hier der Proteinnachweis auch für die fünf anderen kanonischen CGs geführt werden). Auch in der histochemischen Untersuchung konnte C-GRP nur in der Bursa Fabricii, und zwar in den B-Zellen, den Epithelzellen des interfollikulären Epithels (IFE) bzw. des follikel-assoziierten Epithels (FAE) und in Endothelien der Blutgefässe lokalisiert werden. Ebenso treten die fünf kanonischen CGs in der Bursa Fabricii auf: CG-1A in der Basallamina und Endothelien der Blutgefässe, CG-1B in der Lamina propria und den Epithelzellen des IFE, CG-2 in Kapillaren, CG-3 in Epithelzellen und CG-8 in B-Zellen des Cortex und der Medulla. Ein gemeinsames Vorkommen von C-GRP und jeweils eines kanonischen CGs erfolgte in Endothelien der Blutgefässe (CG-1A), in den Epithelzellen des IFE (CG-1B, CG-3) und des FAE (CG-3) sowie in B-Zellen der Bursa (CG-8). Die Bindung von fluoreszenz-markiertem C-GRP an B-Zellen konnte durch Heparin blockiert werden, was nicht notwendigerweise bedeutet, dass die Bindung über die vermutliche CRD erfolgen muss. Zukünftige Untersuchungen bestehen darin, die Rolle der einzelnen kanonischen/nicht-kanonischen Galektine als Bestandteil eines Netzwerks in der embryonalen Entwicklung zu definieren. Die Isolierung von Liganden für GRIFIN und GRP (sowohl bei den menschlichen als auch den aviären Proteinen) ist zwingend notwendig, um deren Funktionen und ihre Beteiligungen an physiologischen und pathologischen Prozessen zu bestimmen.

REFERENCES

1. Sharon, N. (1975) *Complex Carbohydrates. Their Chemistry, Biosynthesis, and Functions*, Addison-Wesley, Reading, MA, USA
2. Kaltner, H., and Gabius, H.-J. (2012) A toolbox of lectins for translating the sugar code: the galectin network in phylogenesis and tumors. *Histol and Histopathol* **27**, 397-416
3. Roseman, S. (2001) Reflections on glycobiology. *J. Biol. Chem.* **276**, 41527-41542
4. Hart, G. W. (2013) Glycan functions pervade biology at all levels. *J Biol Chem* **288**, 6903
5. Gabius, H.-J. (2015) The magic of the sugar code. *Trends Biochem Sci* **40**, 341
6. Winterburn, P. J., and Phelps, C. F. (1972) The significance of glycosylated proteins. *Nature* **236**, 147-151
7. Kasai, K.-i., and Hirabayashi, J. (1996) Galectins: a family of animal lectins that decipher glycocodes. *J Biochem* **119**, 1-8
8. Kaltner, H., Kübler, D., López-Merino, L., Lohr, M., Manning, J. C., Lensch, M., Seidler, J., Lehmann, W. D., André, S., Solís, D., and Gabius, H.-J. (2011) 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 (Hoboken)* **294**, 427-444
9. Manning, J. C., Garcia Caballero, G., Knospe, C., Kaltner, H., and Gabius, H.-J. (2017) Network analysis of adhesion/growth-regulatory galectins and their binding sites in adult chicken retina and choroid. *J Anat* **231**, 23-37
10. Kaltner, H., Toegel, S., Caballero, G. G., Manning, J. C., Ledeen, R. W., and Gabius, H.-J. (2017) Galectins: their network and roles in immunity/tumor growth control. *Histochem Cell Biol* **147**, 239-256
11. Rüdiger, H. (2009) The History of Lectinology. in *The Sugar Code. Fundamentals of Glycosciences* (Gabius, H.-J. ed.), Wiley-VCH, Weinheim,pp 261-268
12. Boyd, W. C. (1954) *The Proteins*, Academic Press, New York
13. Murphy, P. V., André, S., and Gabius, H.-J. (2013) The third dimension of reading the sugar code by lectins: design of glycoclusters with cyclic scaffolds as tools with the aim to define correlations between spatial presentation and activity. *Molecules* **18**, 4026-4053
14. Solís, D., Bovin, N. V., Davis, A. P., Jiménez-Barbero, J., Romero, A., Roy, R., Smetana, K., Jr., and Gabius, H.-J. (2015) A guide into glycosciences: How chemistry, biochemistry and biology cooperate to crack the sugar code. *Biochim Biophys Acta* **1850**, 186-235
15. André, S., Kaltner, H., Manning, J. C., Murphy, P. V., and Gabius, H.-J. (2015) Lectins: getting familiar with translators of the sugar code. *Molecules* **20**, 1788-1823
16. Gabius, H.-J., André, S., Jiménez-Barbero, J., Romero, A., and Solís, D. (2011) From lectin structure to functional glycomics: principles of the sugar code. *Trends Biochem Sci* **36**, 298-313
17. Gabius, H.-J. (2011) The how and why of Ca^{2+} involvement in lectin activity. *Trends Glycosci Glycotechnol* **23**, 168-177
18. Barondes, S. H. (1997) Galectins: A personal overview. *Trends Glycosci Glycotechnol* **9**, 1-7
19. Cooper, D. N. W. (2002) Galectinomics: finding themes in complexity. *Biochim Biophys Acta* **1572**, 209-231
20. Ruiz, F. M., Fernandez, I. S., López-Merino, L., Lagartera, L., Kaltner, H., Menéndez, M., André, S., Solís, D., Gabius, H.-J., and Romero, A. (2013) Fine-tuning of prototype

chicken galectins: structure of CG-2 and structure-activity correlations. *Acta Crystallogr D Biol Crystallogr* **69**, 1665-1676
21. Lobsanov, Y. D., and Rini, J. M. (1997) Galectin structure. *Trends Glycosci and Glycotechnol* **9**, 145-154
22. Taroni, C., Jones, S., and Thornton, J. M. (2000) Analysis and prediction of carbohydrate binding sites. *Protein Eng* **13**, 89-98
23. Gabius, H.-J. (1997) Animal lectins. *Eur J Biochem* **243**, 543-576
24. Kasai, K.-i., and Hirabayashi, J. (1996) Galectins: a family of animal lectins that decipher glycocodes. *J Biochem* **119**, 1-8
25. Tsai, C. M., Guan, C. H., Hsieh, H. W., Hsu, T. L., Tu, Z., Wu, K. J., Lin, C. H., and Lin, K. I. (2011) Galectin-1 and galectin-8 have redundant roles in promoting plasma cell formation. *J Immunol* **187**, 1643-1652
26. Toegel, S., Bieder, D., André, S., Kayser, K., Walzer, S. M., Hobusch, G., Windhager, R., and Gabius, H.-J. (2014) 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**, 373-388
27. Teichberg, V. I., Silman, I., Beitsch, D. D., and Resheff, G. (1975) A β-D-galactoside binding protein from electric organ tissue of *Electrophorus electricus*. *Proc Natl Acad Sci U S A* **72**, 1383-1387
28. Marchant, J., Cowper, B., Liu, Y., Lai, L., Pinzan, C., Marq, J. B., Friedrich, N., Sawmynaden, K., Liew, L., Chai, W., Childs, R. A., Saouros, S., Simpson, P., Roque Barreira, M. C., Feizi, T., Soldati-Favre, D., and Matthews, S. (2012) Galactose recognition by the apicomplexan parasite *Toxoplasma gondii*. *J Biol Chem* **287**, 16720-16733
29. Walser, P. J., Kües, U., Aebi, M., and Künzler, M. (2005) Ligand interactions of the *Coprinopsis cinerea* galectins. *Fungal Genet Biol* **42**, 293-305
30. Takeuchi, T., Arata, Y., and Kasai, K.-i. (2016) Galactoseβ1-4fucose: A unique disaccharide unit found in N-glycans of invertebrates including nematodes. *Proteomics* **16**, 3137-3147
31. Pfeifer, K., Haasemann, M., Gamulin, V., Bretting, H., Fahrenholz, F., and Müller, W. E. G. (1993) S-type lectins occur also in invertebrates: high conservation of the carbohydrate recognition domain in the lectin genes from the marine sponge *Geodia cydonium*. *Glycobiology* **3**, 179-184
32. Wagner-Hülsmann, C., Bachinski, N., Diehl-Seifert, B., Blumbach, B., Steffen, R., Pancer, Z., and Müller, W. E. G. (1996) A galectin links the aggregation factor to cells in the sponge (*Geodia cydonium*) system. *Glycobiology* **6**, 785-793
33. Goletz, S., Hanisch, F. G., and Karsten, U. (1997) Novel αGalNAc containing glycans on cytokeratins are recognized in vitro by galectins with type II carbohydrate recognition domains. *J Cell Sci* **110**, 1585-1596
34. Stalz, H., Roth, U., Schleuder, D., Macht, M., Haebel, S., Strupat, K., Peter-Katalinič, J., and Hanisch, F. G. (2006) The *Geodia cydonium* galectin exhibits prototype and chimera-type characteristics and a unique sequence polymorphism within its carbohydrate recognition domain. *Glycobiology* **16**, 402-414
35. Nemoto-Sasaki, Y., Hayama, K., Ohya, H., Arata, Y., Kaneko, M. K., Saitou, N., Hirabayashi, J., and Kasai, K.-i. (2008) *Caenorhabditis elegans* galectins LEC-1-LEC-11: structural features and sugar-binding properties. *Biochim Biophys Acta* **1780**, 1131-1142
36. Nemoto-Sasaki, Y., Takai, S., Takeuchi, T., Arata, Y., Nishiyama, K., Yamada, A., Takahashi, H., Natsugari, H., and Kasai, K.-i. (2011) The *DC2.3* gene in *Caenorhabditis elegans* encodes a galectin that recognizes the galactoseβ1→4fucose disaccharide unit. *Biol Pharm Bull* **34**, 1635-1639

37. Takeuchi, T., Nemoto-Sasaki, Y., Arata, Y., and Kasai, K.-i. (2011) Galectin LEC-6 interacts with glycoprotein F57F4.4 to cooperatively regulate the growth of *Caenorhabditis elegans*. *Biol Pharm Bull* **34**, 1139-1142
38. Greenhalgh, C. J., Loukas, A., Donald, D., Nikolaou, S., and Newton, S. E. (2000) A family of galectins from *Haemonchus contortus*. *Mol Biochem Parasitol* **107**, 117-121
39. Li, C., Wei, X., Xu, L., and Li, X. (2007) Recombinant galectins of male and female *Haemonchus contortus* do not hemagglutinate erythrocytes of their natural host. *Vet Parasitol* **144**, 299-303
40. Lu, M., Tian, X., Yang, X., Yuan, C., Ehsan, M., Liu, X., Yan, R., Xu, L., Song, X., and Li, X. (2017) The N- and C-terminal carbohydrate recognition domains of *Haemonchus contortus* galectin bind to distinct receptors of goat PBMC and contribute differently to its immunomodulatory functions in host-parasite interactions. *Parasit Vectors* **10**, 409
41. Tasumi, S., and Vasta, G. R. (2007) A galectin of unique domain organization from hemocytes of the Eastern oyster (*Crassostrea virginica*) is a receptor for the protistan parasite Perkinsus marinus. *J Immunol* **179**, 3086-3098
42. Pace, K. E., and Baum, L. G. (2002) Insect galectins: roles in immunity and development. *Glycoconj J* **19**, 607-614
43. Pace, K. E., Lebestky, T., Hummel, T., Arnoux, P., Kwan, K., and Baum, L. G. (2002) Characterization of a novel *Drosophila melanogaster* galectin. Expression in developing immune, neural, and muscle tissues. *J Biol Chem* **277**, 13091-13098
44. Ogawa, T., Ishii, C., Kagawa, D., Muramoto, K., and Kamiya, H. (1999) Accelerated evolution in the protein-coding region of galectin cDNAs, congerin I and congerin II, from skin mucus of conger eel (*Conger myriaster*). *Biosci Biotechnol Biochem* **63**, 1203-1208
45. Ahmed, H., Du, S. J., O'Leary, N., and Vasta, G. R. (2004) Biochemical and molecular characterization of galectins from zebrafish (*Danio rerio*): notochord-specific expression of a prototype galectin during early embryogenesis. *Glycobiology* **14**, 219-232
46. Vasta, G. R., Ahmed, H., Du, S., and Henrikson, D. (2004) Galectins in teleost fish: Zebrafish (*Danio rerio*) as a model species to address their biological roles in development and innate immunity. *Glycoconj J* **21**, 503-521
47. Shoji, H., Nishi, N., Hirashima, M., and Nakamura, T. (2003) Characterization of the *Xenopus* galectin family. Three structurally different types as in mammals and regulated expression during embryogenesis. *J Biol Chem* **278**, 12285-12293
48. Gitt, M. A., Colnot, C., Poirier, F., Nani, K. J., Barondes, S. H., and Leffler, H. (1998) Galectin-4 and galectin-6 are two closely related lectins expressed in mouse gastrointestinal tract. *J Biol Chem* **273**, 2954-2960
49. Gitt, M. A., Xia, Y. R., Atchison, R. E., Lusis, A. J., Barondes, S. H., and Leffler, H. (1998) Sequence, structure, and chromosomal mapping of the mouse Lgals6 gene, encoding galectin-6. *J Biol Chem* **273**, 2961-2970
50. Houzelstein, D., Goncalves, I. R., Orth, A., Bonhomme, F., and Netter, P. (2008) Lgals6, a 2-million-year-old gene in mice: a case of positive Darwinian selection and presence/absence polymorphism. *Genetics* **178**, 1533-1545
51. Gitt, M. A., Wiser, M. F., Leffler, H., Herrmann, J., Xia, Y. R., Massa, S. M., Cooper, D. N.W., Lusis, A. J., and Barondes, S. H. (1995) Sequence and mapping of galectin-5, a β-galactoside-binding lectin, found in rat erythrocytes. *J Biol Chem* **270**, 5032-5038
52. Kaltner, H., and Gabius, H.-J. (2012) A toolbox of lectins for translating the sugar code: the galectin network in phylogenesis and tumors. *Histol Histopathol* **27**, 397-416
53. Den, H., and Malinzak, D. A. (1977) Isolation and properties of β-D-galactoside-specific lectin from chick embryo thigh muscle. *J Biol Chem* **252**, 5444-5448

54. Nowak, T. P., Kobiler, D., Roel, L. E., and Barondes, S. H. (1977) Developmentally regulated lectin from embryonic chick pectoral muscle. Purification by affinity chromatography. *J Biol Chem* **252**, 6026-6030
55. Beyer, E. C., Zweig, S. E., and Barondes, S. H. (1980) Two lactose binding lectins from chicken tissues. Purified lectin from intestine is different from those in liver and muscle. *J Biol Chem* **255**, 4236-4239
56. Oda, Y., and Kasai, K.-i. (1983) Purification and characterization of β-galactoside-binding lectin from chick embryonic skin. *Biochim Biophys Acta* **761**, 237-245
57. Sakakura, Y., Hirabayashi, J., Oda, Y., Ohyama, Y., and Kasai, K.-i. (1990) Structure of chicken 16-kDa β-galactoside-binding lectin. Complete amino acid sequence, cloning of cDNA, and production of recombinant lectin. *J Biol Chem* **265**, 21573-21579
58. Varela, P. F., Solís, D., Diaz-Maurino, T., Kaltner, H., Gabius, H.-J., and Romero, A. (1999) The 2.15 Å crystal structure of CG-16, the developmentally regulated homodimeric chicken galectin. *J Mol Biol* **294**, 537-549
59. Kaltner, H., Solís, D., Kopitz, J., Lensch, M., Lohr, M., Manning, J. C., Mürnseer, M., Schnölzer, M., André, S., Saíz, J. L., and Gabius, H.-J. (2008) Prototype chicken galectins revisited: characterization of a third protein with distinctive hydrodynamic behaviour and expression pattern in organs of adult animals. *Biochem J* **409**, 591-599
60. López-Lucendo, M. F., Solís, D., Saíz, J. L., Kaltner, H., Russwurm, R., André, S., Gabius, H.-J., and Romero, A. (2009) 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**, 366-378
61. Crittenden, S. L., Roff, C. F., and Wang, J. L. (1984) Carbohydrate-binding protein 35: identification of the galactose-specific lectin in various tissues of mice. *Mol Cell Biol* **4**, 1252-1259
62. Nurminskaya, M., and Linsenmayer, T. F. (1996) Identification and characterization of up-regulated genes during chondrocyte hypertrophy. *Dev Dyn* **206**, 260-271
63. Gorski, J. P., Liu, F. T., Artigues, A., Castagna, L. F., and Osdoby, P. (2002) New alternatively spliced form of galectin-3, a member of the β-galactoside-binding animal lectin family, contains a predicted transmembrane-spanning domain and a leucine zipper motif. *J Biol Chem* **277**, 18840-18848
64. Geatrell, J. C., Gan, P. M., Mansergh, F. C., Kisiswa, L., Jarrin, M., Williams, L. A., Evans, M. J., Boulton, M. E., and Wride, M. A. (2009) Apoptosis gene profiling reveals spatio-temporal regulated expression of the p53/Mdm2 pathway during lens development. *Exp Eye Res* **88**, 1137-1151
65. Kaltner, H., Solís, D., André, S., Lensch, M., Manning, J. C., Mürnseer, M., Saíz, J. L., and Gabius, H.-J. (2009) 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**, 4403-4416
66. Levi, G., and Teichberg, V. I. (1985) Isolation and characterization of chicken thymic electrolectin. *Biochem J* **226**, 379-384
67. Cerra, R. F., Gitt, M. A., and Barondes, S. H. (1985) Three soluble rat β-galactoside-binding lectins. *J Biol Chem* **260**, 10474-10477
68. Gabius, H.-J., and Wu, A. M. (2006) The emerging functionality of endogenous lectins: A primer to the concept and a case study on galectins including medical implications. *Chang Gung Med J* **29**, 37-62
69. Kaltner, H., Seyrek, K., Heck, A., Sinowatz, F., and Gabius, H.-J. (2002) Galectin-1 and galectin-3 in fetal development of bovine respiratory and digestive tracts. Comparison of cell type-specific expression profiles and subcellular localization. *Cell Tissue Res* **307**, 35-46

70. Wasano, K., and Yamamoto, T. (1989) Rat lung 29 kD β-galactoside-binding lectin is secreted by bronchiolar Clara cells into airways. *Histochemistry* **90**, 447-451
71. Cornillot, J. D., Pontet, M., Dupuy, C., Chadli, A., Caron, M., Joubert-Caron, R., Bourin, P., and Bladier, D. (1998) Production and characterization of a monoclonal antibody able to discriminate galectin-1 from galectin-2 and galectin-3. *Glycobiology* **8**, 425-432
72. Kopitz, J., von Reitzenstein, C., Burchert, M., Cantz, M., and Gabius, H.-J. (1998) Galectin-1 is a major receptor for ganglioside GM1, a product of the growth-controlling activity of a cell surface ganglioside sialidase, on human neuroblastoma cells in culture. *J Biol Chem* **273**, 11205-11211
73. Schwarz, G., Jr., Remmelink, M., Decaestecker, C., Gielen, I., Budel, V., Burchert, M., Darro, F., Danguy, A., Gabius, H.-J., Salmon, I., and Kiss, R. (1999) Galectin fingerprinting in tumor diagnosis. Differential expression of galectin-3 and galectin-3 binding sites, but not galectin-1, in benign vs malignant uterine smooth muscle tumors. *Am J Clin Pathol* **111**, 623-631
74. Harrison, F. L., and Chesterton, C. J. (1980) Erythroid developmental agglutinin is a protein lectin mediating specific cell-cell adhesion between differentiating rabbit erythroblasts. *Nature* **286**, 502-504
75. Rabinovich, G. A., and Croci, D. O. (2012) Regulatory circuits mediated by lectin-glycan interactions in autoimmunity and cancer. *Immunity* **36**, 322-335
76. Allen, H. J., Sucato, D., Gottstine, S., Kisailus, E., Nava, H., Petrelli, N., Castillo, N., and Wilson, D. (1991) Localization of endogenous β-galactoside-binding lectin in human cells and tissues. *Tumour Biol* **12**, 52-60
77. Nio, J., Kon, Y., and Iwanaga, T. (2005) Differential cellular expression of galectin family mRNAs in the epithelial cells of the mouse digestive tract. *J Histochem Cytochem* **53**, 1323-1334
78. Nio-Kobayashi, J., Takahashi-Iwanaga, H., and Iwanaga, T. (2009) Immunohistochemical localization of six galectin subtypes in the mouse digestive tract. *J Histochem Cytochem* **57**, 41-50
79. Gabius, H.-J., and Kayser, K. (2014) Introduction to glycopathology: the concept, the tools and the perspectives. *Diagn Pathol* **9**, 4
80. Kaltner, H., Singh, T., Manning, J. C., Raschta, A. S., André, S., Sinowatz, F., and Gabius, H.-J. (2015) Network monitoring of adhesion/growth-regulatory galectins: localization of the five canonical chicken proteins in embryonic and maturing bone and cartilage and their introduction as histochemical tools. *Anat Rec (Hoboken)* **298**, 2051-2070
81. Smetana, K., Jr., André, S., Kaltner, H., Kopitz, J., and Gabius, H.-J. (2013) Context-dependent multifunctionality of galectin-1: a challenge for defining the lectin as therapeutic target. *Expert Opin Ther Targets* **17**, 379-392
82. Ledeen, R. W., Wu, G., André, S., Bleich, D., Huet, G., Kaltner, H., Kopitz, J., and Gabius, H.-J. (2012) Beyond glycoproteins as galectin counterreceptors: tumor-effector T cell growth control via ganglioside GM1 [corrected]. *Ann N Y Acad Sci* **1253**, 206-221
83. Sturm, A., Lensch, M., André, S., Kaltner, H., Wiedenmann, B., Rosewicz, S., Dignass, A. U., and Gabius, H.-J. (2004) Human galectin-2: novel inducer of T cell apoptosis with distinct profile of caspase activation. *J Immunol* **173**, 3825-3837
84. 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., and Gabius, H.-J. (2007) Tumor suppressor $p16^{INK4a}$ - modulator of glycomic profile and galectin-1 expression to

increase susceptibility to carbohydrate-dependent induction of anoikis in pancreatic carcinoma cells. *FEBS J* **274**, 3233-3256

85. Valenzuela, H. F., Pace, K. E., Cabrera, P. V., White, R., Porvari, K., Kaija, H., Vihko, P., and Baum, L. G. (2007) *O*-glycosylation regulates LNCaP prostate cancer cell susceptibility to apoptosis induced by galectin-1. *Cancer Res* **67**, 6155-6162
86. Sanchez-Ruderisch, H., Detjen, K. M., Welzel, M., André, S., Fischer, C., Gabius, H.-J., and Rosewicz, S. (2011) Galectin-1 sensitizes carcinoma cells to anoikis via the fibronectin receptor α5β1-integrin. *Cell Death Differ* **18**, 806-816
87. Jung, T. Y., Jung, S., Ryu, H. H., Jeong, Y. I., Jin, Y. H., Jin, S. G., Kim, I. Y., Kang, S. S., and Kim, H. S. (2008) Role of galectin-1 in migration and invasion of human glioblastoma multiforme cell lines. *J Neurosurg* **109**, 273-284
88. Camby, I., Belot, N., Lefranc, F., Sadeghi, N., de Launoit, Y., Kaltner, H., Musette, S., Darro, F., Danguy, A., Salmon, I., Gabius, H.-J., and Kiss, R. (2002) Galectin-1 modulates human glioblastoma cell migration into the brain through modifications to the actin cytoskeleton and levels of expression of small GTPases. *J Neuropathol Exp Neurol* **61**, 585-596
89. Camby, I., Decaestecker, C., Lefranc, F., Kaltner, H., Gabius, H.-J., and Kiss, R. (2005) Galectin-1 knocking down in human U87 glioblastoma cells alters their gene expression pattern. *Biochem Biophys Res Commun* **335**, 27-35
90. Kopitz, J., von Reitzenstein, C., André, S., Kaltner, H., Uhl, J., Ehemann, V., Cantz, M., and Gabius, H.-J. (2001) Negative regulation of neuroblastoma cell growth by carbohydrate-dependent surface binding of galectin-1 and functional divergence from galectin-3. *J Biol Chem* **276**, 35917-35923
91. André, S., Kaltner, H., Lensch, M., Russwurm, R., Siebert, H. C., Fallsehr, C., Tajkhorshid, E., Heck, A. J. R., von Knebel Doeberitz, M., Gabius, H.-J., and Kopitz, J. (2005) Determination of structural and functional overlap/divergence of five prototype galectins by analysis of the growth-regulatory interaction with ganglioside GM1 *in silico* and *in vitro* on human neuroblastoma cells. *Int J Cancer* **114**, 46-57
92. Fischer, C., Sanchez-Ruderisch, H., Welzel, M., Wiedenmann, B., Sakai, T., André, S., Gabius, H.-J., Khachigian, L., Detjen, K. M., and Rosewicz, S. (2005) Galectin-1 interacts with the α5β1 fibronectin receptor to restrict carcinoma cell growth via induction of p21 and p27. *J Biol Chem* **280**, 37266-37277
93. Sanchez-Ruderisch, H., Fischer, C., Detjen, K. M., Welzel, M., Wimmel, A., Manning, J. C., André, S., and Gabius, H.-J. (2010) 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**, 3552-3563
94. Yang, R. Y., Hsu, D. K., and Liu, F.-T. (1996) Expression of galectin-3 modulates T-cell growth and apoptosis. *Proc Natl Acad Sci U S A* **93**, 6737-6742
95. Akahani, S., Nangia-Makker, P., Inohara, H., Kim, H. R., and Raz, A. (1997) Galectin-3: a novel antiapoptotic molecule with a functional BH1 (NWGR) domain of Bcl-2 family. *Cancer Res* **57**, 5272-5276
96. Harazono, Y., Kho, D. H., Balan, V., Nakajima, K., Zhang, T., Hogan, V., and Raz, A. (2014) Galectin-3 leads to attenuation of apoptosis through Bax heterodimerization in human thyroid carcinoma cells. *Oncotarget* **5**, 9992-10001
97. Gabius, H.-J., Manning, J. C., Kopitz, J., André, S., and Kaltner, H. (2016) Sweet complementarity: the functional pairing of glycans with lectins. *Cell Mol Life Sci* **73**, 1989-2016
98. Percec, V., Leowanawat, P., Sun, H. J., Kulikov, O., Nusbaum, C. D., Tran, T. M., Bertin, A., Wilson, D. A., Peterca, M., Zhang, S., Kamat, N. P., Vargo, K., Moock, D., Johnston, E. D., Hammer, D. A., Pochan, D. J., Chen, Y., Chabre, Y. M., Shiao, T. C., Bergeron-Brlek, M., André, S., Roy, R., Gabius, H.-J., and Heiney, P. A. (2013)

Modular synthesis of amphiphilic Janus glycodendrimers and their self-assembly into glycodendrimersomes and other complex architectures with bioactivity to biomedically relevant lectins. *J Am Chem Soc* **135**, 9055-9077

99. Zhang, S., Moussodia, R. O., Murzeau, C., Sun, H. J., Klein, M. L., Vértesy, S., André, S., Roy, R., Gabius, H.-J., and Percec, V. (2015) Dissecting molecular aspects of cell interactions using glycodendrimersomes with programmable glycan presentation and engineered human lectins. *Angew Chem Int Ed Engl* **54**, 4036-4040
100. Vértesy, S., Michalak, M., Miller, M. C., Schnölzer, M., André, S., Kopitz, J., Mayo, K. H., and Gabius, H.-J. (2015) Structural significance of galectin design: impairment of homodimer stability by linker insertion and partial reversion by ligand presence. *Protein Eng Des Sel* **28**, 199-210
101. Zhang, S., Moussodia, R. O., Vértesy, S., André, S., Klein, M. L., Gabius, H.-J., and Percec, V. (2015) Unraveling functional significance of natural variations of a human galectin by glycodendrimersomes with programmable glycan surface. *Proc Natl Acad Sci U S A* **112**, 5585-5590
102. Zhang, S., Xiao, Q., Sherman, S. E., Muncan, A., Ramos Vicente, A. D., Wang, Z., Hammer, D. A., Williams, D., Chen, Y., Pochan, D. J., Vértesy, S., André, S., Klein, M. L., Gabius, H.-J., and Percec, V. (2015) Glycodendrimersomes from sequence-defined Janus glycodendrimers reveal high activity and sensor capacity for the agglutination by natural variants of human lectins. *J Am Chem Soc* **137**, 13334-13344
103. Jessell, T. M., Hynes, M. A., and Dodd, J. (1990) Carbohydrates and carbohydrate-binding proteins in the nervous system. *Annu Rev Neurosci* **13**, 227-255
104. Spicer, S. S., and Schulte, B. A. (1992) Diversity of cell glycoconjugates shown histochemically: a perspective. *J Histochem Cytochem* **40**, 1-38
105. Kleene, R., and Schachner, M. (2004) Glycans and neural cell interactions. *Nat Rev Neurosci* **5**, 195-208
106. Roth, J. (2011) Lectins for histochemical demonstration of glycans. *Histochem Cell Biol* **136**, 117-130
107. Gabius, H.-J., Kaltner, H., Kopitz, J., and André, S. (2015) The glycobiology of the CD system: a dictionary for translating marker designations into glycan/lectin structure and function. *Trends Biochem Sci* **40**, 360-376
108. Manning, J. C., Romero, A., Habermann, F. A., Garcia Caballero, G., Kaltner, H., and Gabius, H.-J. (2017) Lectins: a primer for histochemists and cell biologists. *Histochem Cell Biol* **147**, 199-222
109. Park, J. W., Voss, P. G., Grabski, S., Wang, J. L., and Patterson, R. J. (2001) Association of galectin-1 and galectin-3 with Gemin4 in complexes containing the SMN protein. *Nucleic Acids Res* **29**, 3595-3602
110. Purkrábková, T., Smetana, K., Jr., Dvoránková, B., Holíková, Z., Böck, C., Lensch, M., André, S., Pytlik, R., Liu, F.-T., Klima, J., Smetana, K., Motlík, J., and Gabius, H.-J. (2003) New aspects of galectin functionality in nuclei of cultured bone marrow stromal and epidermal cells: biotinylated galectins as tool to detect specific binding sites. *Biol Cell* **95**, 535-545
111. Haudek, K. C., Patterson, R. J., and Wang, J. L. (2010) SR proteins and galectins: what's in a name? *Glycobiology* **20**, 1199-1207
112. Schlötzer-Schrehardt, U., André, S., Janko, C., Kaltner, H., Kopitz, J., Gabius, H.-J., and Herrmann, M. (2012) Adhesion/growth-regulatory galectins in the human eye: localization profiles and tissue reactivities as a standard to detect disease-associated alterations. *Graefes Arch Clin Exp Ophthalmol* **250**, 1169-1180
113. Gabius, H.-J., Wosgien, B., Hendrys, M., and Bardosi, A. (1991) Lectin localization in human nerve by biochemically defined lectin-binding glycoproteins, neoglycoprotein and lectin-specific antibody. *Histochemistry* **95**, 269-277

114. Schwarz, A., and Futerman, A. H. (1997) Determination of the localization of gangliosides using anti-ganglioside antibodies: comparison of fixation methods. *J Histochem Cytochem* **45**, 611-618
115. Celie, J. W., Beelen, R. H., and van den Born, J. (2005) Effect of fixation protocols on in situ detection of L-selectin ligands. *J Immunol Methods* **298**, 155-159
116. Lohr, M., Kaltner, H., Schwartz-Albiez, R., Sinowatz, F., and Gabius, H.-J. (2010) Towards functional glycomics by lectin histochemistry: strategic probe selection to monitor core and branch-end substitutions and detection of cell-type and regional selectivity in adult mouse testis and epididymis. *Anat Histol Embryol* **39**, 481-493
117. Gabius, H.-J. (2009) Animal and Human Lectins. in *The Sugar Code. Fundamentals of Glycosciences* (Gabius, H.-J. ed.), Wiley-VCH, Weinheim, pp 317-328
118. Cooper, D. N. W., and Barondes, S. H. (1999) God must love galectins; he made so many of them. *Glycobiology* **9**, 979-984
119. Weller, P. F., Bach, D. S., and Austen, K. F. (1984) Biochemical characterization of human eosinophil Charcot-Leyden crystal protein (lysophospholipase). *J Biol Chem* **259**, 15100-15105
120. Dvorak, A. M. (1996) Human basophil recovery from secretion. A review emphasizing the distribution of Charcot-Leyden crystal protein in cells stained with the postfixation electron-dense tracer, cationized ferritin. *Histol Histopathol* **11**, 711-728
121. Swaminathan, G. J., Leonidas, D. D., Savage, M. P., Ackerman, S. J., and Acharya, K. R. (1999) Selective recognition of mannose by the human eosinophil Charcot-Leyden crystal protein (galectin-10): A crystallographic study at 1.8 Å resolution. *Biochemistry* **38**, 15406
122. Sammar, M., Nisamblatt, S., Gonen, R., Huppertz, B., Gizurarson, S., Osol, G., and Meiri, H. (2014) The role of the carbohydrate recognition domain of placental protein 13 (PP13) in pregnancy evaluated with recombinant PP13 and the DelT221 PP13 variant. *PLoS One* **9**, e102832
123. Ogden, A. T., Nunes, I., Ko, K., Wu, S., Hines, C. S., Wang, A. F., Hegde, R. S., and Lang, R. A. (1998) GRIFIN, a novel lens-specific protein related to the galectin family. *J Biol Chem* **273**, 28889-28896
124. Barton, K. A., Hsu, C. D., and Petrash, J. M. (2009) Interactions between small heat shock protein α-crystallin and galectin-related interfiber protein (GRIFIN) in the ocular lens. *Biochemistry* **48**, 3956-3966
125. Ahmed, H., and Vasta, G. R. (2008) Unlike mammalian GRIFIN, the zebrafish homologue (DrGRIFIN) represents a functional carbohydrate-binding galectin. *Biochem Biophys Res Commun* **371**, 350-355
126. Zhang, Q. H., Ye, M., Wu, X. Y., Ren, S. X., Zhao, M., Zhao, C. J., Fu, G., Shen, Y., Fan, H. Y., Lu, G., Zhong, M., Xu, X. R., Han, Z. G., Zhang, J. W., Tao, J., Huang, Q. H., Zhou, J., Hu, G. X., Gu, J., Chen, S. J., and Chen, Z. (2000) 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**, 1546-1560
127. Zhou, D., Ge, H., Sun, J., Gao, Y., Teng, M., and Niu, L. (2008) 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**, 1582-1588
128. Winding, P., and Berchtold, M. W. (2001) The chicken B cell line DT40: a novel tool for gene disruption experiments. *J Immunol Methods* **249**, 1-16
129. Buddecke, H. (2009) Proteoglycans. in *The Sugar Code. Fundamentals of Glycosciences* (Gabius, H.-J. ed.), Wiley-VCH, Weinheim, pp 199-216
130. Moise, A., André, S., Eggers, F., Krzeminski, M., Przybylski, M., and Gabius, H.-J. (2011) Toward bioinspired galectin mimetics: identification of ligand-contacting peptides by proteolytic-excision mass spectrometry. *J Am Chem Soc* **133**, 14844-14847

131. Kohnke-Godt, B., and Gabius, H.-J. (1989) Heparin-binding lectin from human placenta: purification and partial molecular characterization and its relationship to basic fibroblast growth factors. *Biochemistry* **28**, 6531-6538
132. Iwaki, J., Minamisawa, T., Tateno, H., Kominami, J., Suzuki, K., Nishi, N., Nakamura, T., and Hirabayashi, J. (2008) Desulfated galactosaminoglycans are potential ligands for galectins: evidence from frontal affinity chromatography. *Biochem Biophys Res Commun* **373**, 206-212
133. Wälti, M. A., Walser, P. J., Thore, S., Grunler, A., Bednar, M., Künzler, M., and Aebi, M. (2008) Structural basis for chitotetraose coordination by CGL3, a novel galectin-related protein from *Coprinopsis cinerea*. *J Mol Biol* **379**, 146-159
134. Barres, C., Blanc, L., Bette-Bobillo, P., André, S., Mamoun, R., Gabius, H.-J., and Vidal, M. (2010) Galectin-5 is bound onto the surface of rat reticulocyte exosomes and modulates vesicle uptake by macrophages. *Blood* **115**, 696-705
135. Elad-Sfadia, G., Haklai, R., Balan, E., and Kloog, Y. (2004) Galectin-3 augments K-Ras activation and triggers a Ras signal that attenuates ERK but not phosphoinositide 3-kinase activity. *J Biol Chem* **279**, 34922-34930
136. Belanis, L., Plowman, S. J., Rotblat, B., Hancock, J. F., and Kloog, Y. (2008) Galectin-1 is a novel structural component and a major regulator of h-ras nanoclusters. *Mol Biol Cell* **19**, 1404-1414
137. Thurston, T. L., Wandel, M. P., von Muhlinen, N., Foeglein, A., and Randow, F. (2012) Galectin 8 targets damaged vesicles for autophagy to defend cells against bacterial invasion. *Nature* **482**, 414-418
138. Kim, B. W., Hong, S. B., Kim, J. H., Kwon, D. H., and Song, H. K. (2013) Structural basis for recognition of autophagic receptor NDP52 by the sugar receptor galectin-8. *Nat Commun* **4**, 1613
139. Chen, H. Y., Fermin, A., Vardhana, S., Weng, I. C., Lo, K. F., Chang, E. Y., Maverakis, E., Yang, R. Y., Hsu, D. K., Dustin, M. L., and Liu, F. T. (2009) Galectin-3 negatively regulates TCR-mediated $CD4^+$ T-cell activation at the immunological synapse. *Proc Natl Acad Sci U S A* **106**, 14496-14501
140. Ogino, H., and Yasuda, K. (1998) Induction of lens differentiation by activation of a bZIP transcription factor, L-Maf. *Science* **280**, 115-118
141. Reza, H. M., and Yasuda, K. (2004) Roles of Maf family proteins in lens development. *Dev Dyn* **229**, 440-448
142. Than, N. G., Romero, R., Goodman, M., Weckle, A., Xing, J., Dong, Z., Xu, Y., Tarquini, F., Szilagyi, A., Gal, P., Hou, Z., Tarca, A. L., Kim, C. J., Kim, J. S., Haidarian, S., Uddin, M., Bohn, H., Benirschke, K., Santolaya-Forgas, J., Grossman, L. I., Erez, O., Hassan, S. S., Zavodszky, P., Papp, Z., and Wildman, D. E. (2009) A primate subfamily of galectins expressed at the maternal-fetal interface that promote immune cell death. *Proc Natl Acad Sci U S A* **106**, 9731-9736
143. Bhat, R., Lerea, K. M., Peng, H., Kaltner, H., Gabius, H.-J., and Newman, S. A. (2011) A regulatory network of two galectins mediates the earliest steps of avian limb skeletal morphogenesis. *BMC Dev Biol* **11**, 6
144. Astorgues-Xerri, L., Riveiro, M. E., Tijeras-Raballand, A., Serova, M., Neuzillet, C., Albert, S., Raymond, E., and Faivre, S. (2014) Unraveling galectin-1 as a novel therapeutic target for cancer. *Cancer Treat Rev* **40**, 307-319
145. Thijssen, V. L., and Griffioen, A. W. (2014) Galectin-1 and -9 in angiogenesis: a sweet couple. *Glycobiology* **24**, 915-920
146. Ebrahim, A. H., Alalawi, Z., Mirandola, L., Rakhshanda, R., Dahlbeck, S., Nguyen, D., Jenkins, M., Grizzi, F., Cobos, E., Figueroa, J. A., and Chiriva-Internati, M. (2014) Galectins in cancer: carcinogenesis, diagnosis and therapy. *Ann Transl Med* **2**, 88

147. Gentilini, L. D., Jaworski, F. M., Tiraboschi, C., Perez, I. G., Kotler, M. L., Chauchereau, A., Laderach, D. J., and Compagno, D. (2017) Stable and high expression of Galectin-8 tightly controls metastatic progression of prostate cancer. *Oncotarget* **8**, 44654-44668
148. Wu, M. H., Hong, T. M., Cheng, H. W., Pan, S. H., Liang, Y. R., Hong, H. C., Chiang, W. F., Wong, T. Y., Shieh, D. B., Shiau, A. L., Jin, Y. T., and Chen, Y. L. (2009) Galectin-1-mediated tumor invasion and metastasis, up-regulated matrix metalloproteinase expression, and reorganized actin cytoskeletons. *Mol Cancer Res* **7**, 311-318
149. Matsuda, Y., Yamagiwa, Y., Fukushima, K., Ueno, Y., and Shimosegawa, T. (2008) Expression of galectin-3 involved in prognosis of patients with hepatocellular carcinoma. *Hepatol Res* **38**, 1098-1111
150. Zhang, S., Moussodia, R. O., Sun, H. J., Leowanawat, P., Muncan, A., Nusbaum, C. D., Chelling, K. M., Heiney, P. A., Klein, M. L., André, S., Roy, R., Gabius, H.-J., and Percec, V. (2014) Mimicking biological membranes with programmable glycan ligands self-assembled from amphiphilic Janus glycodendrimers. *Angew Chem Int Ed Engl* **53**, 10899-10903
151. Wang, J., Lu, Z. H., Gabius, H.-J., Rohowsky-Kochan, C., Ledeen, R. W., and Wu, G. (2009) Cross-linking of GM1 ganglioside by galectin-1 mediates regulatory T cell activity involving TRPC5 channel activation: possible role in suppressing experimental autoimmune encephalomyelitis. *J Immunol* **182**, 4036-4045
152. Wang, G. N., André, S., Gabius, H.-J., and Murphy, P. V. (2012) Bi- to tetravalent glycoclusters: synthesis, structure-activity profiles as lectin inhibitors and impact of combining both valency and headgroup tailoring on selectivity. *Org Biomol Chem* **10**, 6893-6907
153. Roy, R., Cao, Y., Kaltner, H., Kottari, N., Shiao, T. C., Belkhadem, K., André, S., Manning, J. C., Murphy, P. V., and Gabius, H.-J. (2017) Teaming up synthetic chemistry and histochemistry for activity screening in galectin-directed inhibitor design. *Histochem Cell Biol* **147**, 285-301
154. Toegel, S., Weinmann, D., André, S., Walzer, S. M., Bilban, M., Schmidt, S., Chiari, C., Windhager, R., Krall, C., Bennani-Baiti, I. M., and Gabius, H.-J. (2016) Galectin-1 couples glycobiology to inflammation in osteoarthritis through the activation of an NF-κB-regulated gene network. *J Immunol* **196**, 1910-1921

SUPPLEMENTARY MATERIAL

Publication 1 - Supplementary Material

All Supplementary Material to the respective publication

Chicken GRIFIN: A homodimeric member of the galectin network with canonical properties and a unique expression profile

comprised of:

Fig. S1
Fig. S2
Fig. S3
Fig. S4
Fig. S5
Fig. S6
Fig. S7
Fig. S8
Table S1
Table S2A
Table S2B
Table S3
Table S4

can be found and downloaded via the following link:

http://www.sciencedirect.com/science/article/pii/S0300908416301109#appsec1

In the cases in which the large size of the figures/tables make an implementation into this text problematic, we kindly ask to visit the online version of the Supplementary Material using the link provided above.
The following figures are included here (but can also be viewed in larger size and resolution online):

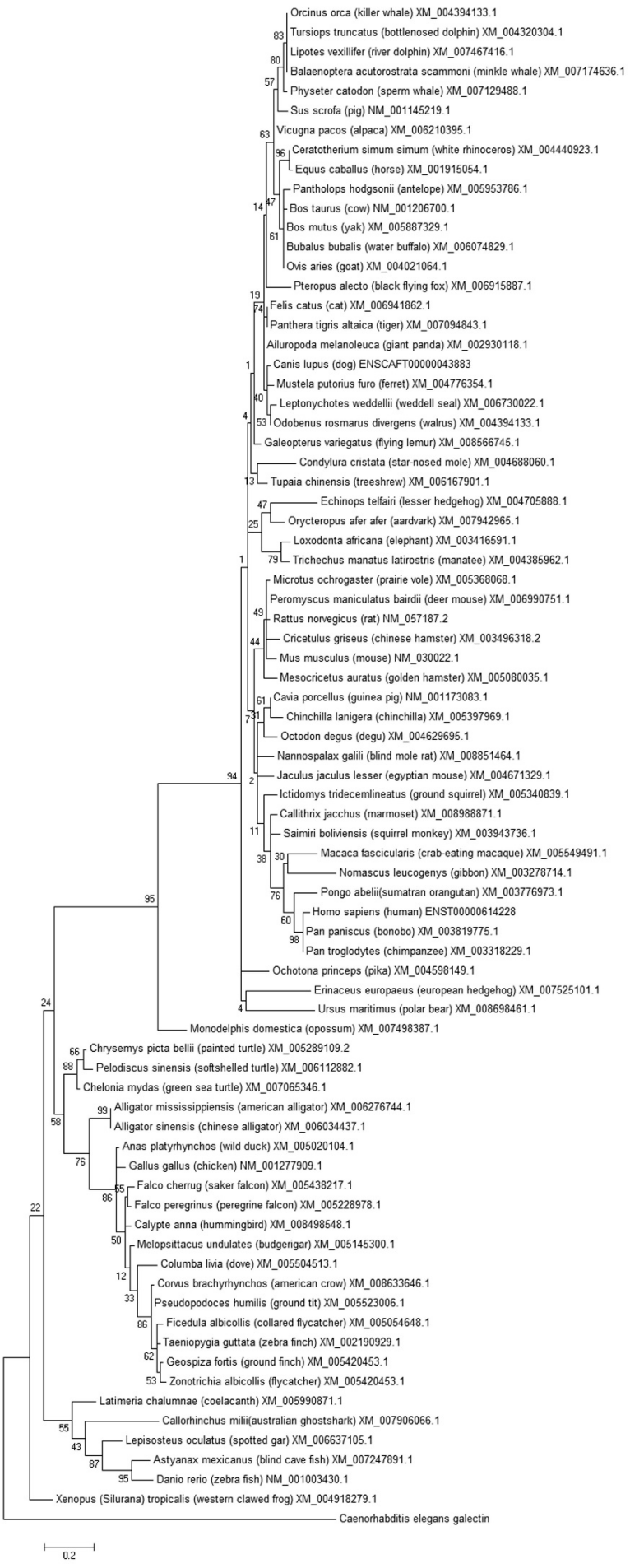


Fig. S1: Molecular phylogenetic analysis by the Maximum Likelihood method using the complete set of GRIFIN sequences currently available from the NCBI (last search performed May 2016) and ENSEMBL (release 84 – March 2016) databases (77 species). Amino acid sequences were deduced from coding sequences annotated on the genomic level, except for human GRP, C-GRIFIN, rat GRIFIN and zebrafish GRIFIN, where the information on the level of the protein is available. Deduced amino acid sequence for a *Caenorhabditis elegans* galectin Lec-6 (BAA09794) is used as an outgroup to display a rooted tree.

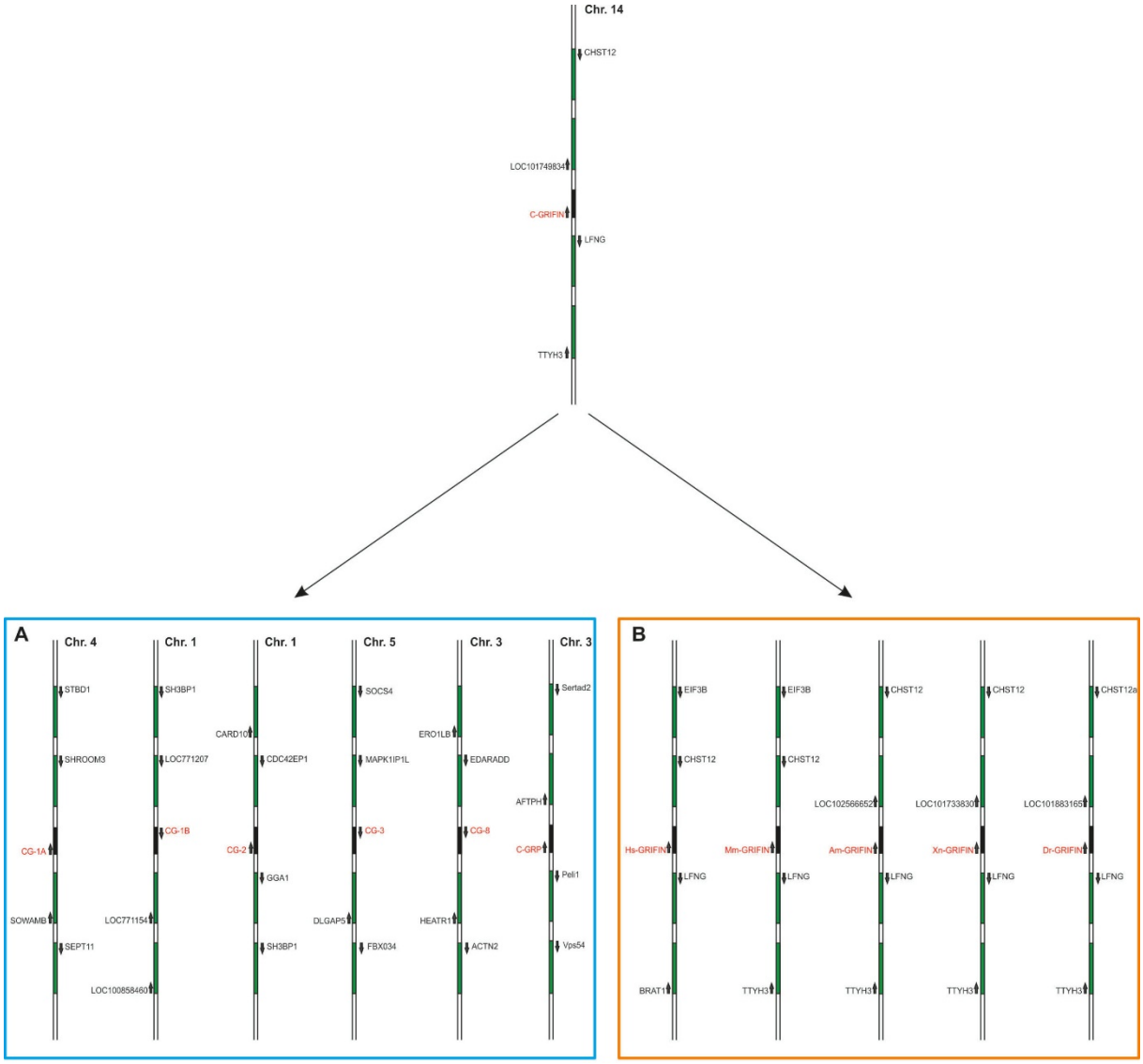


Fig. S2: Comparative documentation of the chromosomal environment of the C-GRIFIN gene with those of the five genes for canonical CGs and C-GRP (blue box, A, chromosome localization is indicated) and with GRIFIN genes of five selected species (orange box, B; Hs, *Homo sapiens*, chromosome 7; Mm, *Mus musculus*, chromosome 5; Am, *Alligator mississippiensis,* unplaced scaffold; Xn, *Xenopus tropicalis*, unplaced scaffold; Dr, *Danio rerio,* chromosome 3). Gene orientation is indicated by arrows; box lengths and spacers are not drawn to scale. Gene abbreviations: ACTN2, actin α2; AFTPH, aftifilin; BRAT1, BRCA1-associated ATM activator 1; CARD10, caspase recruitment domain family member 10; CDC42EP1, CDC42 effector protein Rho GTPase binding 1; CHST12, carbohydrate (chondroitin 4) sulfotransferase 2; CHST12a, carbohydrate (chondroitin 4) sulfotransferase 2a; DLGAP5, discs, large (*Drosophila*) homolog-associated protein; EDARADD, EDAR-associated death domain; EIF3B, eukaryotic translation initiation factor 3, subunit B; ERO1LB, ERO1-like β (*S. cerevisiae*); FBX034, F-box protein 34; GGA1, Golgi-associated, γ-adaptin ear containing, ARF binding protein 1; HEATR1, HEAT repeat containing 1; LFNG, LFNG O-fucosylpeptide 3-β-N-acetylglucosaminyltransferase; LOC101733830, probable G-protein coupled receptor No18; LOC101883165, neuropeptide FF receptor 2-like; LOC102566652, galanin receptor type 1-like; LOC771154, urotensin-2 receptor-like; MAPK1IP1L, chromosome 14 open reading frame 32;; Peli1, pellino E3 ubiquitin protein ligase; SEPT11, septin 11; Sertad2, SERTA domain containing 2; SH3BP1, SH3-domain binding protein 1; SHROOM3, shroom family member 3; SOCS4, suppressor of cytokine signaling 4; SOWAMB, sosondowah ankyrin repeat domain family member B; STBD1, startch binding domain 1; TTYH3, tweety family member 3; Vps54, vacuolar protein sorting 54 homolog (*S. cerevisiae*).

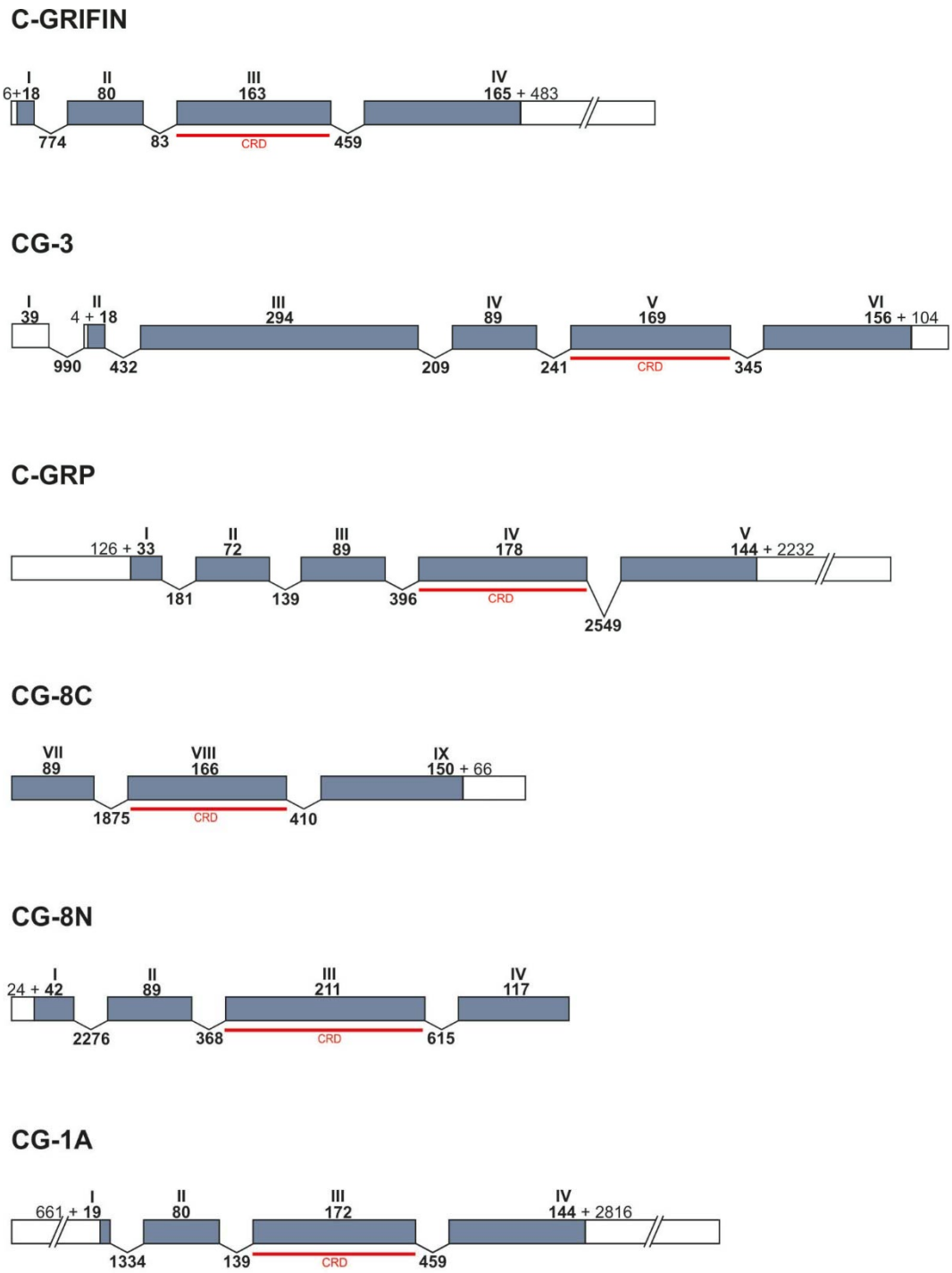


Fig. S3: Organization of the genes for C-GRIFIN, CG-3, C-GRP, the respective parts of the gene for CG-8 separated into its C- and N-terminal CRDs and finally the gene for CG-1A, given in the degree of vicinity based on sequence aligments, as shown in Fig. 2. Exons are graphically displayed as boxes (5′/3′-untranslated exon parts in white, coding exons in gray, the exon encoding the CRD underlined in red, size proportional to exon lengths), and introns are shown as lines (size not proportional to intron lengths). Roman numbers above exon boxes define the spatial order of exon occurrence. The lengths of exons and introns are given in base pairs (bp).

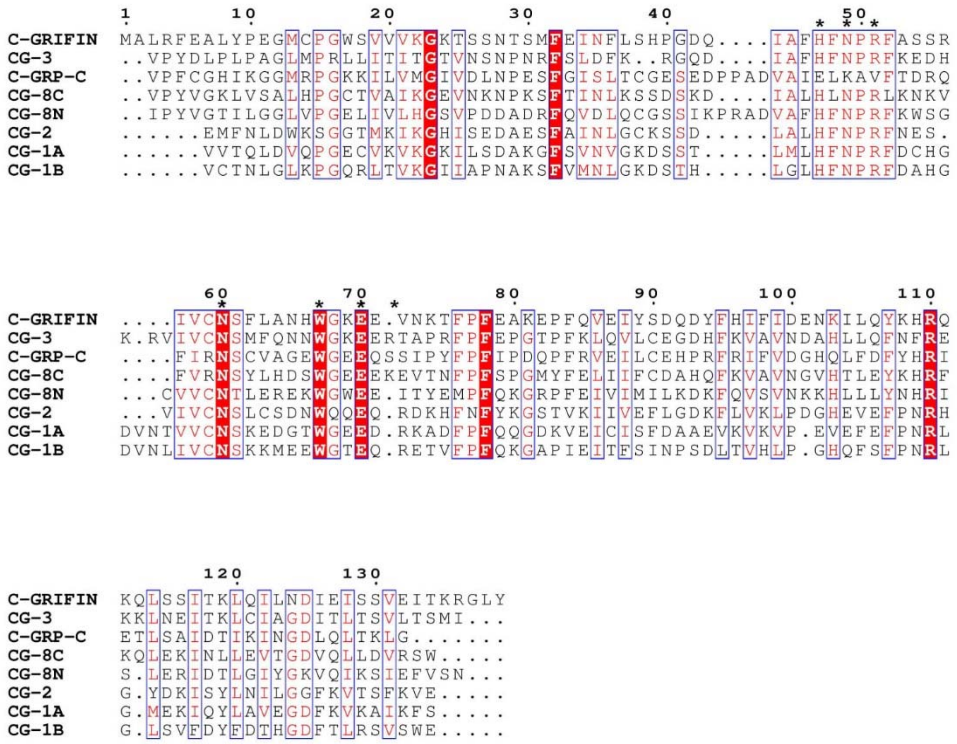


Fig. S4: Comparison of amino acid sequences of C-GRIFIN with those of the CRDs for the five canonical CGs and the shortened version of C-GRP, i.e. C-GRP-C (amino acid 37 is set to position 3 for alignment).

Fig. S5: Illustrations of multiple sequence alignments for predicted GRIFIN proteins, edited in Jalview (vs. 2.7): Please view online and/or download using the link given above.

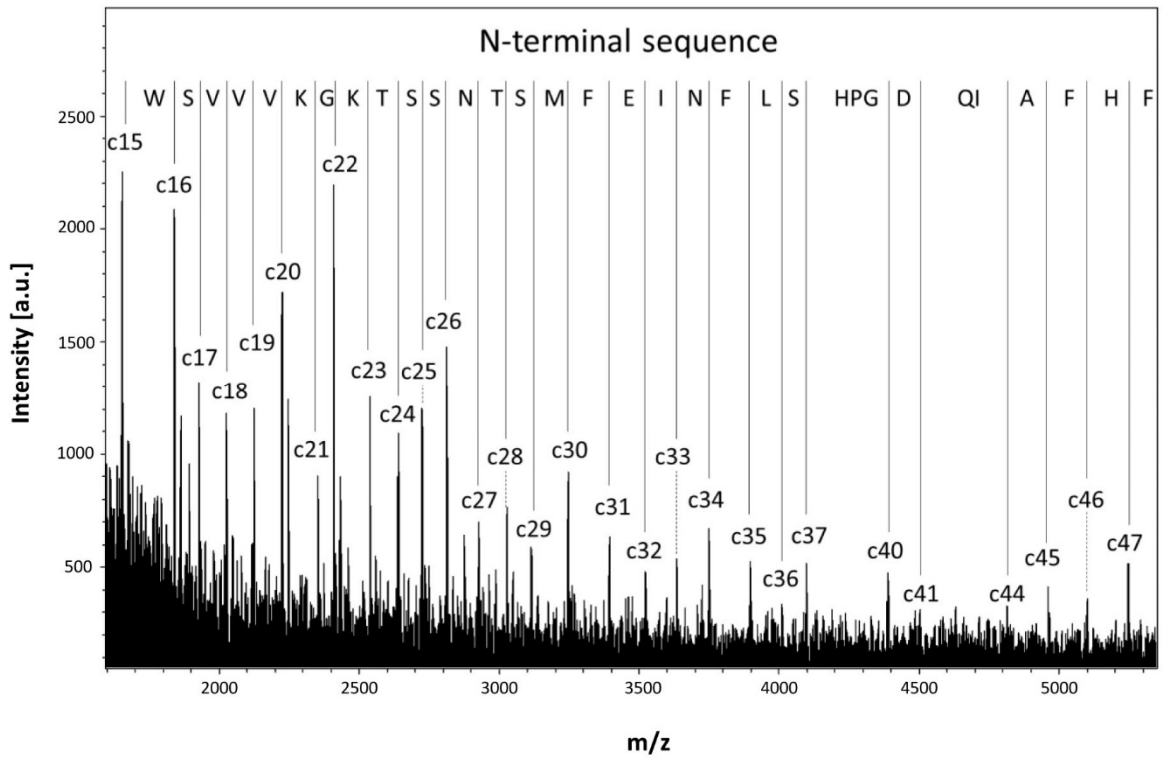


Fig. S6: Sequence analysis of the N-terminus of C-GRIFIN by relISD. Mass spectrum of C-GRIFIN using sinapinic acid (SA) as matrix presents the smallest detectable peptide at the c15/c16 position and the ensuing peptide ladder.

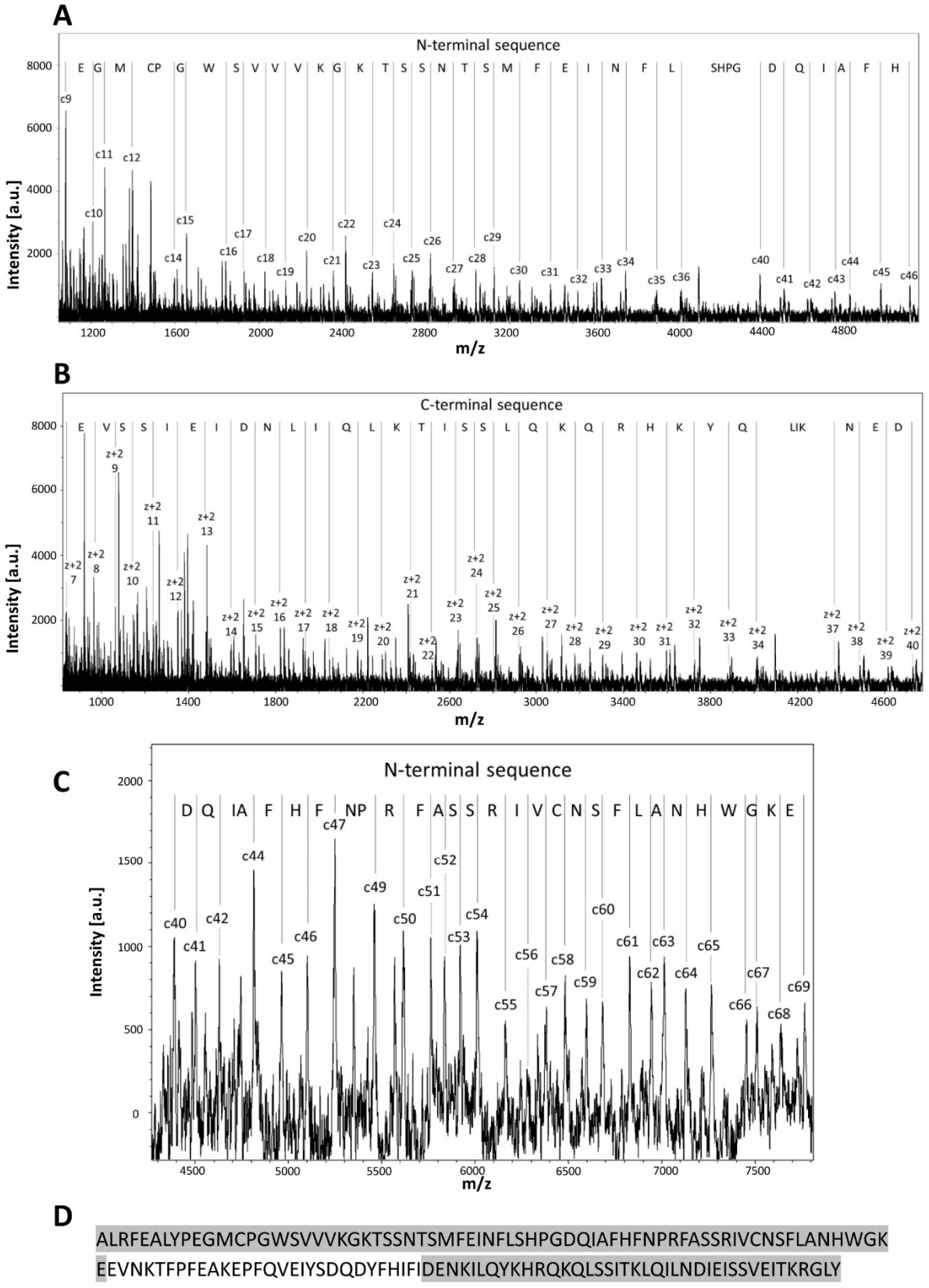


Fig. S7: Sequence analysis of the N- and C-termini of C-GRIFIN by reISD and linISD. Mass spectra of C-GRIFIN using super 2,5-dihydroxybenzoic acid (sDHB) as matrix show the smallest detectable peptide at the c9/c10 (A) and 7/8 (z+2) (B) positions and the ensuing stepwise peptide ladder. (C) LinISD spectrum using sinapinic acid (SA) as matrix shows the smallest detectable peptide at c40/c41 and the ensuing stepwise peptide ladder up to c68/69. (D) Sequence coverage by ISD shown by shaded sections. Experimental and calculated mass values are given in Table S2A and S2B.

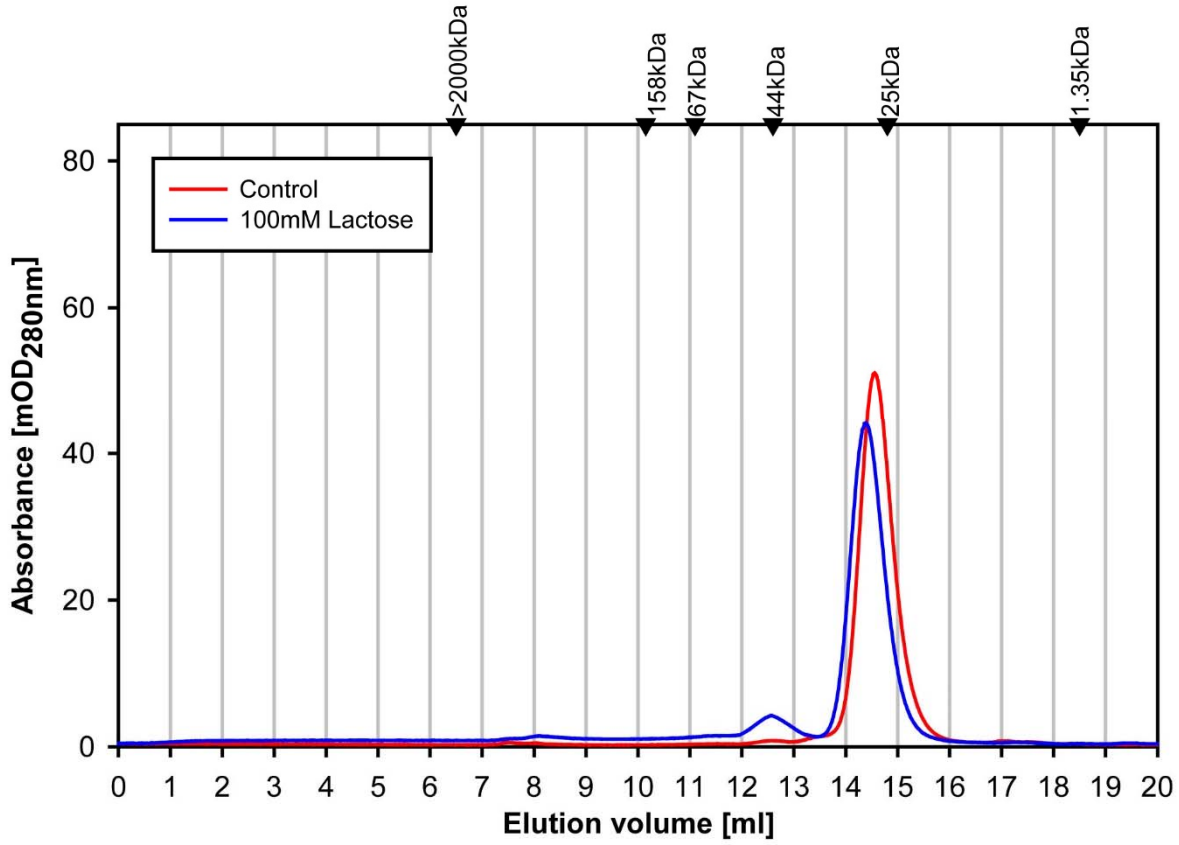


Fig. S8: Gel filtration analysis of C-GRIFIN. Elution profiles of C-GRIFIN in the absence and in the presence of 100 mM lactose. Black arrowheads indicate elution volume of molecular weight markers used for calibration.

Table S1: Compilation of genomic orthologs/paralogs of the GRIFIN gene:
Please view online and/or download using the link given above.

Table S4: Glycan microarray data for C-GRIFIN:
Please view online and/or download using the link given above.

Publication 2 – Supplementary Material

All Supplementary Material to the respective publication

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

comprised of:

Fig. S1
Fig. S2
Fig. S3
Fig. S4
Fig. S5
Fig. S6
Fig. S7
Fig. S8
Fig. S9
Fig. S10
Fig. S11
Table S1A
Table S1B
Table S2

can be found and downloaded via the following link:

http://www.sciencedirect.com/science/article/pii/S0304416516302057?via%3Dihub#ec0005

In the cases in which the large size of the figures/tables make an implementation into this text problematic, we kindly ask to visit the online version of the Supplementary Material using the link provided above.

The following figures are included here (but can also be viewed in larger size and resolution online):

Fig. S1: Molecular phylogenetic analisis by the Maximum Likelihood method using the complete set of GRP sequences currently available (last search performed at 13.01.2016) from the NCBI and ENSEMBL databases (119 species).

Please view online and/or download using the link given above.

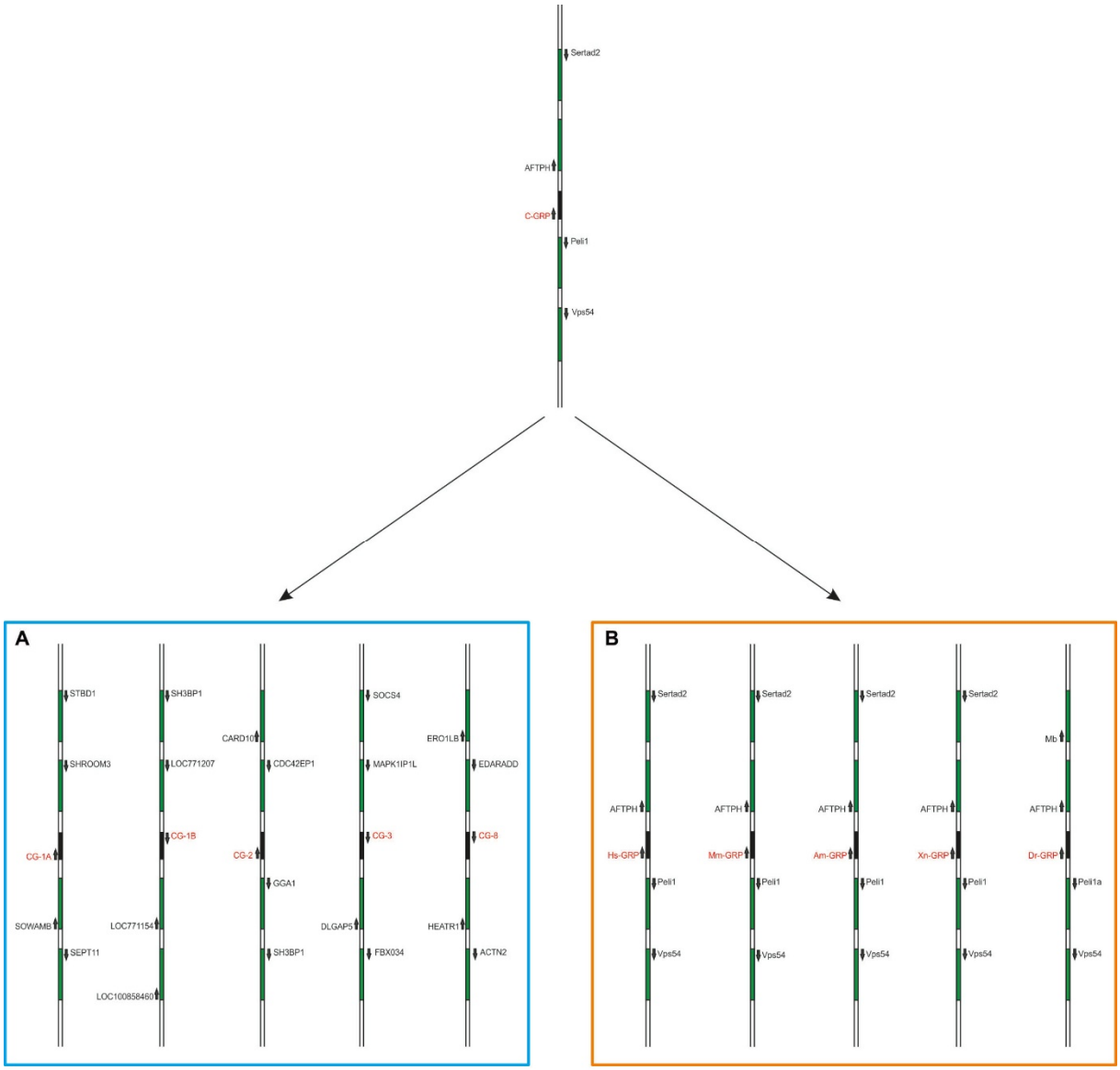


Fig. S2: Comparative documentation of the chromosomal environment of the C-GRP gene with those of the five genes for canonical CGs (blue box, A) and with GRP genes of five selected species (orange box, B; Hs, *Homo sapiens*, chromosome 2; Mm, *Mus musculus*, chromosome 11; Am, *Alligator mississippiensis,* unplaced scaffold; Xn, *Xenopus laeviensis*, unplaced scaffold; Dr, *Danio rerio,* chromosome 1). Gene orientation is indicated by arrows; box lengths and spacers are not drawn to scale. Gene abbreviations: ACTN2, actin α2; AFTH, aftifilin; CARD10, caspase recruitment domain family member 10; CDC42EP1, CDC42 effector protein Rho GTPase binding 1; DLGAP5, discs, large (*Drosophila*) homolog-associated protein; EDARADD, EDAR-associated death domain; ERO1LB, ERO1-like β (*S. cerevisiae*); FBX034, F-box protein 34; GGA1, Golgi-associated, γ-adaptin ear containing, ARF binding protein 1; HEATR1, HEAT repeat containing 1; LOC100858460, arf-GAP with dual PH domain-containing protein 1-like; LOC771154, urotensin-2 receptor-like; LOC771207, pyrodoxal phosphate phosphatase-like; MAPK1IP1L, chromosome 14 open reading frame 32; Mb, myoglobin; PELI, pellino E3 ubiquitin protein ligase; SEPT11, septin 11; Sertad2, SERTA domain containing 2; SH3BP1, SH3-domain binding protein 1; SHROOM3, shroom family member 3; SOCS4, supressor of cytokine signaling 4; SOWAMB, sosondowah ankyrin repeat domain family member B; STBD1, startch binding domain 1; Vps54, vacuolar protein sorting 54 homolog (*S. cerevisiae*).

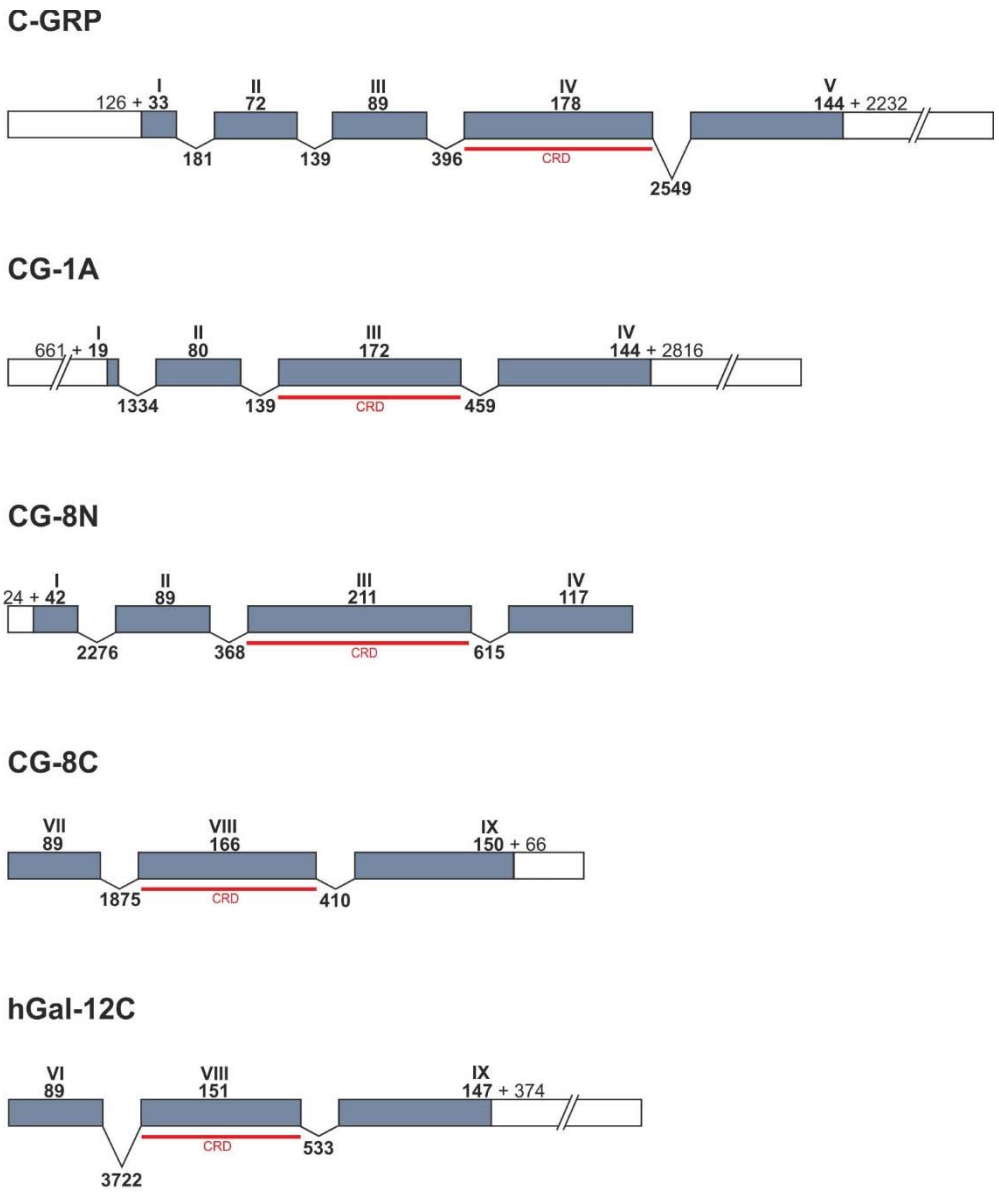
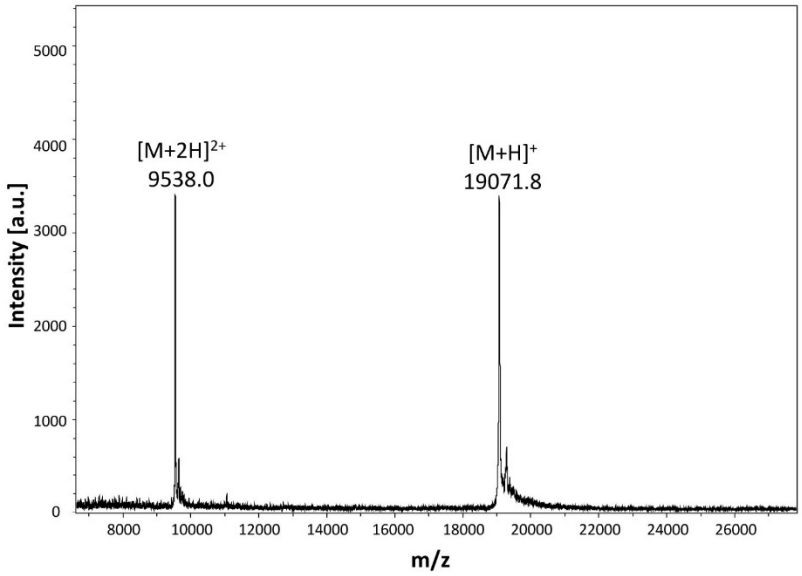


Fig. S3: Organization of the genes for C-GRP, CG-1A, the N- and C-terminal domains of CG-8, and the C-terminal domain of human galectin-12. Exons are graphically displayed as boxes (5′/3′-untranslated exon parts in white, coding exons in gray, the exon encoding the CRD underlined in red, size proportional to exon lengths), and introns are shown as lines (size not proportional to intron lengths). Roman numbers above exon boxes define the spatial order of exon occurrence. The lengths of exons and introns are given in base pairs (bp).

Fig. S4: Illustrations of multiple sequence alignments for predicted GRP proteins, edited in Jalview (vs. 2.7; Waterhouse *et al*. 2009 [25]).
Please view online and/or download using the link given above.

A



B

Sequence	[MH]+ (mono) calc	[MH]+ (mono) exp
QFIR	563.3	563.3
GGMRPGK	702.4	702.4
AVFTDR	708.4	708.3
GPAGTVAER	857.4	857.5
VEILCEHPR	1095.6	1095.7
VEILCEHPR*	1152.6	1152.7
IETLSAIDTIK	1203.7	1203.7
IFVDGHQLFDFYHR	1793.9	1794.0
LIVPFCGHIKGGMRPGK	1810.0	1809.9
IEDGHLNNSLGSPVQADVYFPR	2428.2	2428.3
NSCVAGEWGEEQSSIPYFPFIPDQPFR*	3157.4	3157.6
VEILCEHPRFRIFVDGHQLFDFYHR	3173.6	3173.6

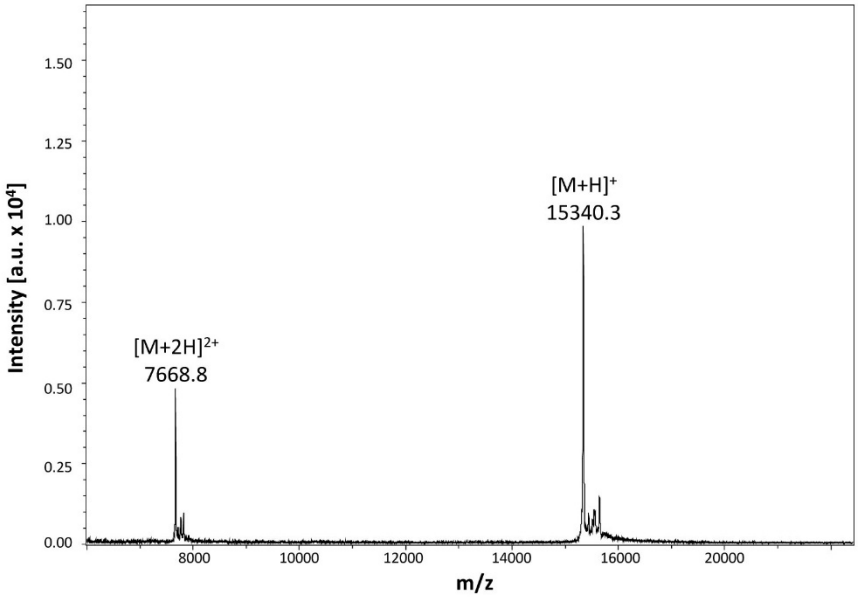
*Carbamidomethyl at C

C

GPAGTVAERDAPKIEDGHLNNSLGSPVQADVYFPRLIVPFCGHIKGGMRPG
KKILVMGIVDLNPESFGISLTCGESEDPPADVAIELKAVFTDRQFIRNSCV
AGEWGEEQSSIPYFPFIPDQPFRVEILCEHPRFRIFVDGHQLFDFYHRIET
LSAIDTIKINGDLQLTKLG

Fig. S5: MALDI TOF MS of C-GRP. (A) Molecular mass determination. The two peaks represent its singly and doubly charged molecular ions (calculated mass: 19068.9 Da). (B, C) Peptide mass fingerprinting by MALDI TOF MS. The list of detected peptides with their calculated (calc) and experimentally measured (exp) mass values (B) and the sequence coverage (C) are shown.

A



B

Sequence	[MH]⁺ (mono) calc	[MH]⁺ (mono) exp
QFIR	563.3	563.3
GGMRPGK	702.4	702.4
AVFTDR	708.4	708.4
VEILCEHPR	1095.6	1095.7
GPVPFCGHIK*	1111.6	1111.7
VEILCEHPR*	1152.6	1152.7
IETLSAIDTIK	1203.7	1203.7
IFVDGHQLFDFYHR	1793.9	1794.0
NSCVAGEWGEEQSSIPYFPFIPDQPFR*	3157.4	3157.7

*Carbamidomethyl at C

C

GPVPFCGHIKGGMRPGKKILVMGIVDLNPESFGISLTCGESEDPPADVAIE
LKAVFTDRQFIRNSCVAGEWGEEQSSIPYFPFIPDQPFRVEILCEHPRFRI
FVDGHQLFDFYHRIETLSAIDTIKINGDLQLTKLG

Fig. S6: MALDI TOF MS of C-GRP-C. (A) Molecular mass determination. The two peaks represent its singly and doubly charged molecular ions (calculated mass: 15338.8 Da). (B, C) Peptide mass fingerprinting by MALDI TOF MS. The list of detected peptides with their calculated (calc) and experimentally measured (exp) mass values (B) and the sequence coverage (C) are shown.

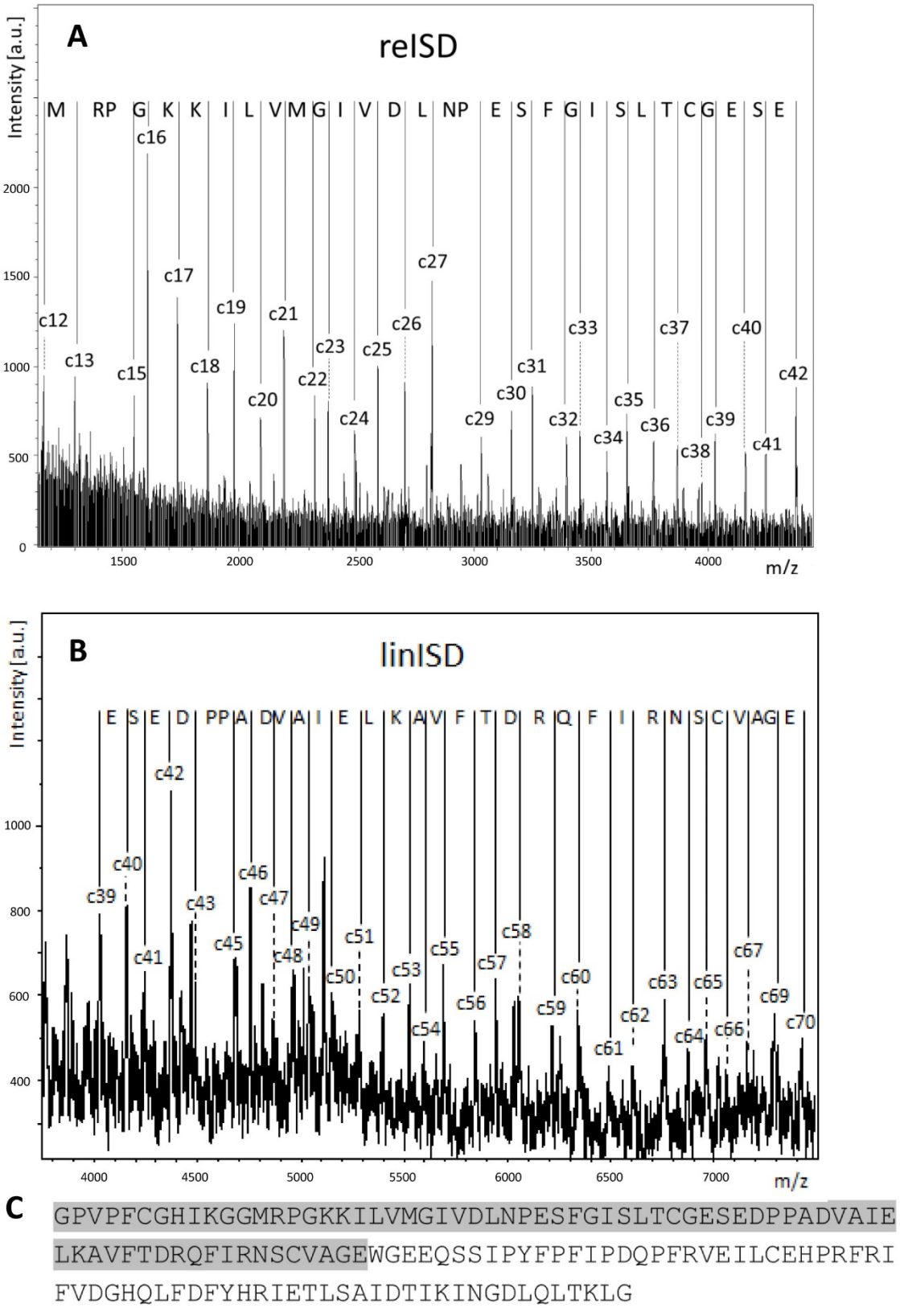


Fig. S7: Sequencing of the N-terminus of C-GRP-C by reISD and linISD. (A) reISD mass spectrum of C-GRP-C shows the smallest detectable peptide at the c12/c13 and the ensuing stepwise peptide ladder. (B) linISD spectrum shows the smallest detectable peptide at c39/c40 and the ensuing stepwise peptide ladder up to c69/c70. (C) Sequence coverage by ISD. Experimental and calculated mass values are given in Table S1B.

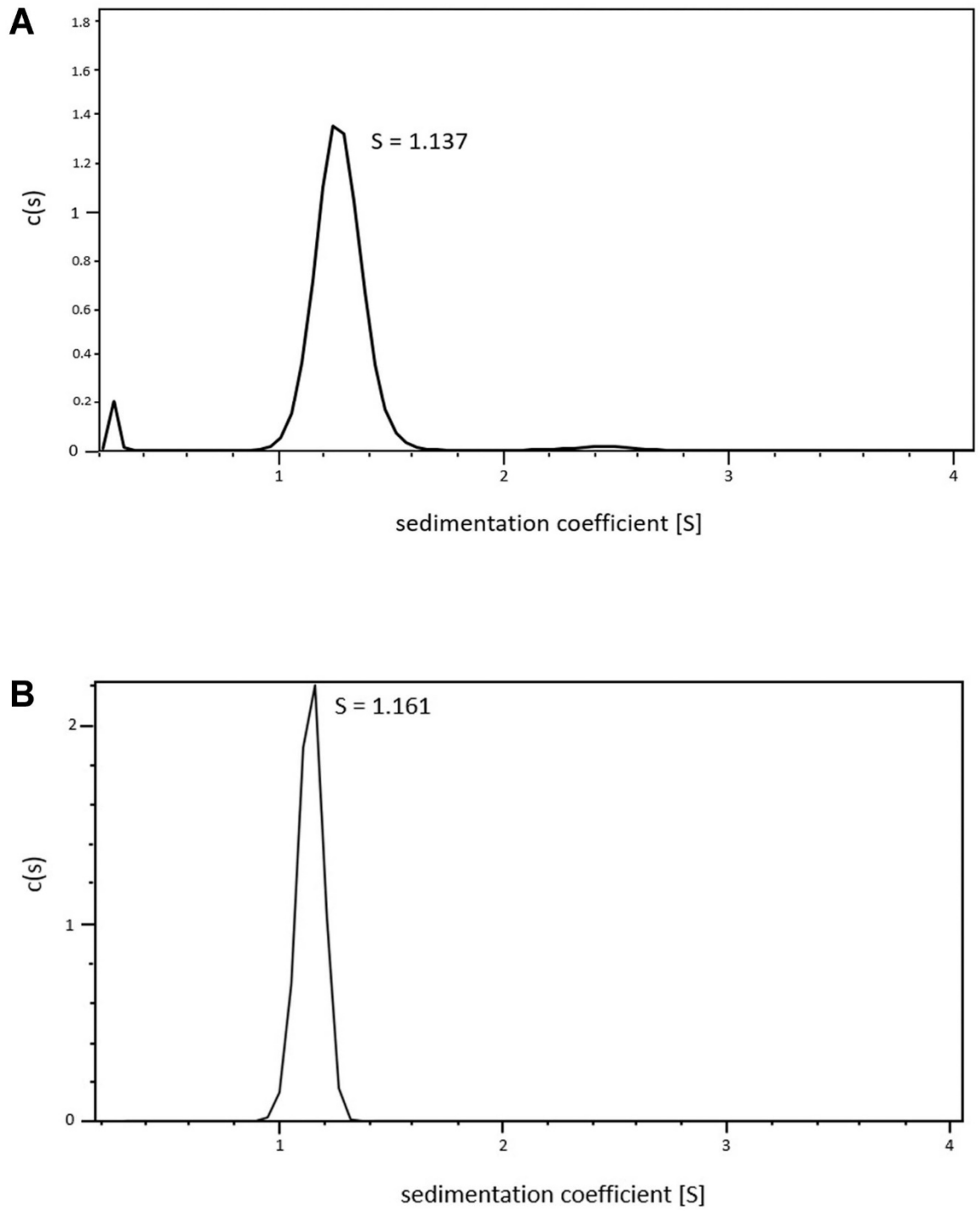


Fig. S8: Analytical ultracentrifugation experiments of C-GRP and C-GRP-C in solution. Sedimentation velocity at 2.0 mg/mL monitored by absorbance at 280 nm. The differential coefficient distribution c(s) curves suggest that both C-GRP (A) and C-GRP-C (B) are predominantly monomeric in solution with sedimentation coefficients of 1.137 ± 0.008 S and 1.161 ± 0.006 S for C-GRP and C-GRP-C, respectively. The apparent molecular weights of 19045 Da (C-GRP) and 16551 Da (C-GRP-C) were calculated from the c(s) distribution.

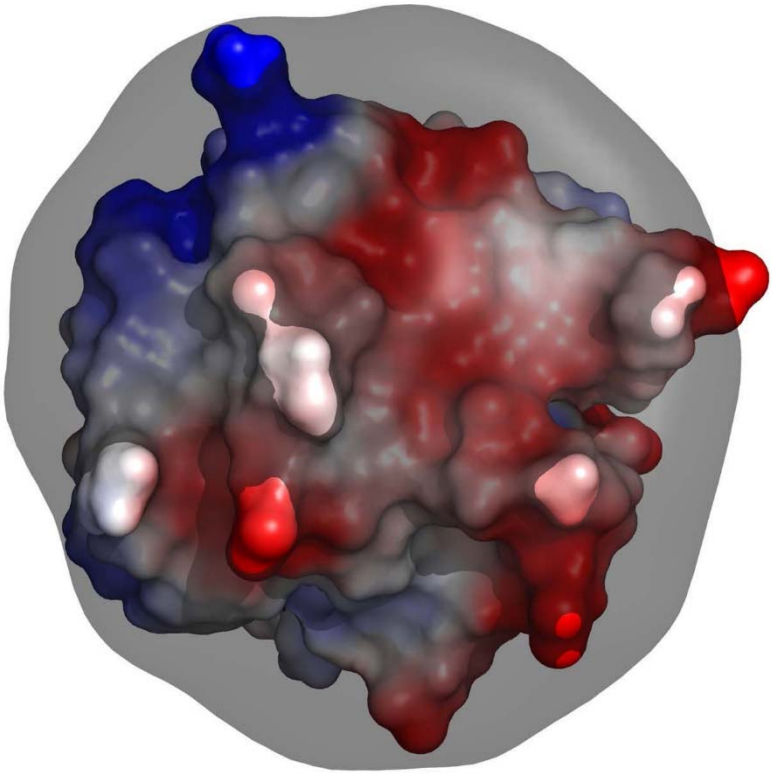


Fig. S9: Fitting of data of a SAXS experiment with C-GRP-C (represented as a surface with the electrostatic potential) using the ab initio SAXS model (grey) as platform.

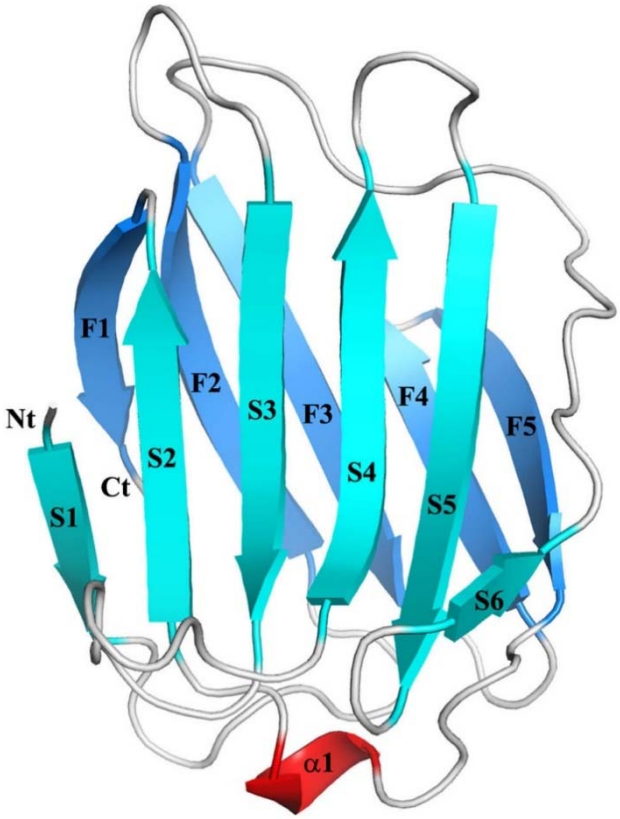


Fig. S10: Overall structure of C-GRP-C. The C-GRP-C monomer presents a jelly roll topology typical of the galectin family. The strands of the two β-sheets are labeled, and a short 3_{10} helix placed between strands F5 and S2.

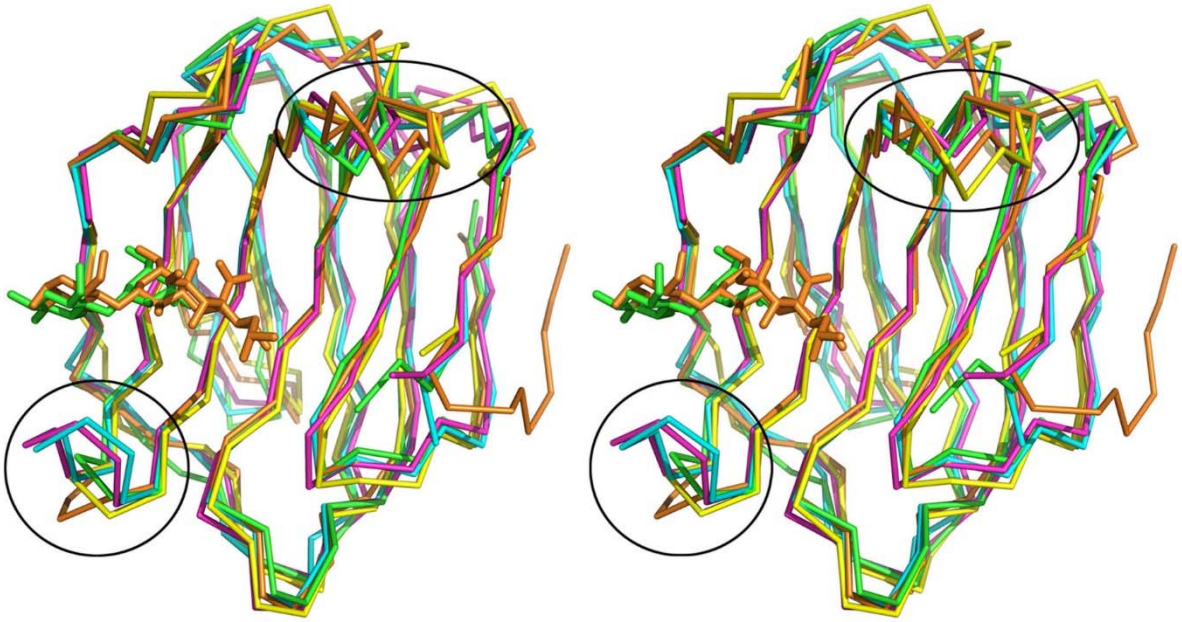


Fig. S11: Superposition of the fold C-GRP-C (yellow) and of four CRDs of canonical CGs, i.e., CG-1A (cyan; 1QMJ), CG-1B (purple; 3DUI), CG-2 (green; 2YMZ) and CG-8N (orange; 4WVW), highlighting the major structural differences in two regions involving connecting loops at the concave face of the carbohydrate-binding pocket.

Table S1A: Calculated and experimental masses of c-ions observed in the reISD spectra for C-GRP.
Please view online and/or download using the link given above.

Table S1B: Calculated and experimental masses of c-ions observed in the reISD spectra for C-GRP-C.

Table S2: Glycan microarray data for C-GRP

Publication 3 – Supplementary Material

All Supplementary Material to the respective publication

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.
comprised of:

Table S1: MatInspector-based search for putative transcription-factor-binding sites in the promoter region (- 2500 bp upstream of translation start) of the C-GRP gene (using the following setting: core similarity = 1.0), listed alphabetically.

can be found and downloaded via the following link:

http://www.sciencedirect.com/science/article/pii/S0304416516302069?via%3Dihub#ec0005

Because of the large size of the table, an implementation into this text is problematic, we kindly ask to visit the online version of the Supplementary Material using the link provided above.

ACKNOWLEDGEMENTS

First of all, I want to say “thank you” to Univ. Prof. Dr. Dr. Hans-Joachim Gabius for giving me the opportunity of doing this thesis and participating in the Marie Curie Initial Training Network GLYCOPHARM. But my thankfulness goes much further than that. Working by your side made me discover the thrill of an unanswered question, the hard work and patience of its study, the required integrity of its answer and, of course, the satisfaction of a well-done job. I shall never forget these last three years and my gratitude will remain.

Special thanks to Prof. Dr. Herbert Kaltner for his guidance, support, patience and invaluable time spent in the conception and realization of this work. Long evenings in the laboratory were easier in your company.

Further thanks go to the European Marie Curie Initial Training Network GLYCOPHARM for the financial support of this project.

I also would like to thank Prof. Dr. Clemens Knospe for his inestimable and friendly advice in the field of anatomy.

Of course, I want to express my gratitude to all my lab co-workers for their prompt help, comfort and advice. I am thankful to Dr. Sabine André, Dr. Carmen Bednorz, Dr. Tanja Kutzner, Dr. Anna-Kristin Ludwig, Dr. Joachim Manning, Dr. Sebastian Schmidt and Dr. Tanuja Singh. It was a pleasure to share this experience with all of you. I would also like to thank the technical assistants Annelise Helfrich, Cornelia Oehme and Beate Hofer for excellent technical support.

I also would like to thank my friends, who in these three years made me feel at home. And to the ones I left in Spain, because distance only strengthened our friendship.

I dedicate this work to my parents, for their wise advice, support and, above all, their endless love. Os quiero, papás.

Finally, I would like to thank E. Theuerkauff for making these years the best ones of my life, as well as for diligent food delivery service.