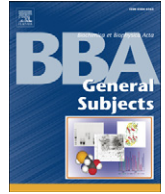
Biochimica et Biophysica Acta xxx (2016) xxx–xxx

Contents lists available at ScienceDirect

Biochimica et Biophysica Acta

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



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

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

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

ARTICLE INFO

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

Keywords:
Adhesion
Crystallography
Lectin
Phylogenesis
Proliferation

ABSTRACT

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

© 2016 Elsevier B.V. All rights reserved.

1. Introduction

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

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

http://dx.doi.org/10.1016/j.bbagen.2016.06.001
0304-4165/© 2016 Elsevier B.V. All rights reserved.

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 intrafamily (C-GRP vs CGs) and inter-species (C-GRP vs GRP of other species) comparisons were processed manually to spot positions of highly conserved amino acids relevant for binding the canonical sugar ligand lactose (Lac) (sequence signature of galectins).

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

2.2. Cloning, expression and purification

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

Please cite this article as: G. García Caballero, et al., Galectin-related protein: An integral member of the network of chicken galectins 1. From strong sequence conservation of the gene ..., Biochim. Biophys. Acta (2016), http://dx.doi.org/10.1016/j.bbagen.2016.06.001

Download English Version:

https://daneshyari.com/en/article/10799791

Download Persian Version:

https://daneshyari.com/article/10799791

Daneshyari.com