Phylogenetic and specificity studies of two-domain GNA-related lectins: generation of multispecificity through domain duplication and divergent evolution

Els J. M. VAN DAMME^{*1}, Sachiko NAKAMURA-TSURUTA[†], David F. SMITH[‡], Maté ONGENAERT[§], Harry C. WINTER^{||}, Pierre ROUGÉ[¶], Irwin J. GOLDSTEIN^{||}, Hanqing MO^{||}, Junko KOMINAMI^{†**}, Raphaël CULERRIER[¶], Annick BARRE[¶], Jun HIRABAYASHI[†] and Willy J. PEUMANS^{*}

^{*}Department of Molecular Biotechnology, Laboratory of Biochemistry and Glycobiology, Ghent University, Coupure Links 653, B-9000 Gent, Belgium, [†]Glycostructure Analysis Team, Research Center for Glycoscience, National Institute of Advanced Industrial Science and Technology, AIST Tsukuba Central 2, 1-1-1, Umezono, Tsukuba, Ibaraki 305-8568, Japan, [‡]Department of Biochemistry, Emory University School of Medicine, Atlanta, GA 30322, U.S.A., [§]Department of Molecular Biotechnology, Laboratory for Bioinformatics and Computational Genomics, Ghent University, Coupure Links 653, B-9000 Gent, Belgium, ^{||}Department of Biological Chemistry, University of Michigan Medical School, Ann Arbor, MI 48109-0606, U.S.A., [¶]Surfaces Cellulaires et Signalisation chez les Végétaux, UMR-CNRS 5546, 24 Chemin de Borde Rouge, 31326 Castanet Tolosan, France, and ^{**}Fine Chemical and Foods Laboratories, J-OIL MILLS, INC., 11, Kagetoricho, Totsuka-ku, Yokohama, Kanagawa 245-0064, Japan

A re-investigation of the occurrence and taxonomic distribution of proteins built up of protomers consisting of two tandem arrayed domains equivalent to the GNA [*Galanthus nivalis* (snowdrop) agglutinin] revealed that these are widespread among monokotyledonous plants. Phylogenetic analysis of the available sequences indicated that these proteins do not represent a monophylogenetic group but most probably result from multiple independent domain duplication/in tandem insertion events. To corroborate the relationship between inter-domain sequence divergence and the widening of specificity range, a detailed comparative analysis was made of the sequences and specificity of a set of two-domain GNA-related lectins. Glycan microarray analyses, frontal affinity chromatography and surface plasmon resonance measurements demonstrated that the two-domain GNA-related lectins acquired a marked diversity in carbohydrate-binding specificity that strikingly contrasts the canonical exclusive specificity of their single domain counterparts towards mannose. Moreover, it appears that most two-domain GNA-related lectins

interact with both high mannose and complex N-glycans and that this dual specificity relies on the simultaneous presence of at least two different independently acting binding sites. The combined phylogenetic, specificity and structural data strongly suggest that plants used domain duplication followed by divergent evolution as a mechanism to generate multispecific lectins from a single mannose-binding domain. Taking into account that the shift in specificity of some binding sites from high mannose to complex type N-glycans implies that the two-domain GNA-related lectins are primarily directed against typical animal glycans, it is tempting to speculate that plants developed two-domain GNA-related lectins for defence purposes.

Key words: carbohydrate-binding specificity, *Galanthus nivalis* (snowdrop) agglutinin (GNA), lectin, mannose-binding, molecular evolution.

INTRODUCTION

Carbohydrate-binding proteins comprising one or two domains equivalent to the GNA [*Galanthus nivalis* (snowdrop) agglutinin] form one of the major plant lectin families [1]. GNA itself was originally isolated from snowdrop (*G. nivalis*) bulbs and described as a lectin with an exclusive specificity towards mannose. Cloning of the corresponding gene [2] and X-ray diffraction analysis [3] revealed that GNA represents a novel lectin family with a unique β -prism fold. After the identification of GNA, similar lectins were isolated and characterized from many other plant species belonging to different families of the Liliopsida (monocots), including Amariyllidaceae, Alliaceae, Orchidaceae, Araceae, Liliaceae and Bromeliaceae [1]. Accordingly, GNA and related lectins were classified into the so-called ‘monocot mannose-binding lectins’. In the meantime, closely related lectins were identified in the liverwort *Marchantia polymorpha* [4], the gymnosperm *Taxus media* [5] and the fish *Fugu rubripes* [6]. Since this implies

that the original term ‘monocot mannose-binding lectins’ is no longer appropriate the lectin family will further be referred to as ‘GNA-related lectins’.

The majority of all characterized plant GNA-related lectins consist of subunits derived from primary translation products comprising a single GNA domain of approx. 110 AA (amino acid) residues flanked by an N-terminal signal peptide and a C-terminal propeptide. Apart from a few monomeric mannose-binding orchid proteins, these lectin subunits associate into homodimers (e.g. dimeric orchid lectins) or homotetramers (e.g. GNA) [1]. Besides these single-domain GNA-related lectins, several proteins have been isolated that are built up of protomers derived from primary translation products comprising two homologous GNA domains arranged in tandem. The eventual molecular structure of the two-domain GNA-related lectins is determined by the degree of oligomerization and the post-translational processing of the precursors [1]. Unlike the single-domain GNA-related lectins, which all have an exclusive specificity towards mannose and

Abbreviations used: AA, amino acid; AMA, *Arum maculatum* agglutinin; ASA-I, *Allium sativum* (garlic) bulb agglutinin I; ASAL, *Allium sativum* (garlic) leaf agglutinin; CAA, *Colchicum autumnale* agglutinin; CVA, *Crocus vernus* agglutinin; FAC, frontal affinity chromatography; GNA, *Galanthus nivalis* (snowdrop) agglutinin; FAC, frontal affinity chromatography; NHS, N-hydroxysuccinimide; PA, pyridylaminated; PAUP, phylogenetic analysis using parsimony; RFU, relative fluorescence units; SPR, surface plasmon resonance; TBS, Tris-buffered saline; TxLC-I, *Tulipa* hybrid lectin I with complex specificity.

¹ To whom correspondence should be addressed (email ElsJM.VanDamme@UGent.be).

oligomannosides, most two-domain GNA-related lectins exhibit a 'complex' specificity. Preliminary specificity studies provided circumstantial evidence that the apparently complex specificity of a two-domain lectin from tulip, TxLC-I (*Tulipa* hybrid lectin I with complex specificity), relies on the simultaneous occurrence of two distinct binding sites [7]. Hapten inhibition assays yielded similar conclusions for the two-domain lectins from *Arum maculatum* [AMA (*A. maculatum* agglutinin)] [8] and *Hyacinthoides hispanica* [9]. Detailed quantitative precipitation and hapten inhibition assays confirmed that another two-domain lectin from *Xanthosoma sagittifolium*, which closely resembles AMA has two different types of carbohydrate-combining sites recognizing oligomannosides and complex N-linked carbohydrates respectively [10].

For all these lectins it was suggested that the presumed occurrence of two distinct carbohydrate-binding sites could be explained in terms of a pronounced sequence divergence between the two GNA domains. The role of inter-domain sequence divergence to generate binding domains with a different specificity is further supported by the observation that the two-domain GNA-related lectin ASA-I [*Allium sativum* (garlic) bulb agglutinin I], which consists of two nearly identical domains [11], exhibits an exclusive specificity towards oligomannosides [12]. To corroborate the relationship between inter-domain sequence divergence and the widening of specificity range, we made a detailed comparative analysis of the sequences and specificity of several two-domain GNA-related lectins using both previously published results and new data obtained by high performance analytical methods. In addition, we searched the publicly available databases for possible yet unidentified GNA orthologues. Two-domain GNA-related lectins are fairly widespread among flowering plants. Phylogenetic analyses indicated that multiple independent domain duplication/in tandem insertion events gave rise to distinct subgroups with a different inter-domain sequence identity and residual sequence identity to single-domain GNA-related lectins. All evidence suggests that the two-domain lectins evolved more rapidly than their single-domain counterparts and that there was a strong tendency to generate an inter-domain sequence divergence that eventually resulted in the formation of binding sites with a totally different specificity. The physiological relevance of this particular evolutionary mechanism is discussed.

EXPERIMENTAL

Lectins

Samples of AMA [8], TxLC-I [7], ASA-I [11], CVA (*Crocus vernus* agglutinin) [13] and CAA (*Colchicum autumnale* agglutinin) [14] were purified as described previously.

SPR (surface plasmon resonance) measurements

The specific interaction of AMA and TxLC-I with immobilized N-glycosylated proteins was analysed by SPR using a biosensor BIAcore X (BIAcore AB, Uppsala, Sweden) according to standard techniques. Details of the preparation of the sensorchips and the measurements are given in the Supplemental Materials (see <http://www.BiochemJ.org/bj/404/bj4040051add.htm>).

FAC (frontal affinity chromatography) analysis

Chemicals

The PA (pyridylaminated) N-linked glycans used in the present study are listed in Supplementary Figure 3 (see <http://www.BiochemJ.org/bj/404/bj4040051add.htm>).

N-linked glycans were purchased from Takara Bio (Kyoto, Japan) and Seikagaku Co. (Tokyo, Japan). Methotrexate derivatized Man9GlcNAc2 was a generous gift from Dr Y. Ito and Dr K. Totani (RIKEN) [15].

Preparation of AMA, and TxLC-I columns

Purified AMA and TxLC-I were dissolved in coupling buffer (50 mM NaHCO₃, pH 8.3, containing 0.5 M NaCl) and coupled to NHS (*N*-hydroxysuccinimide)-activated Sepharose 4FF (Amersham Pharmacia Biotech, Bucks, U.K.) according to the manufacturer's instructions. After deactivation of excess NHS groups by 1 M monoethanolamine, followed by extensive washing with coupling buffer and acetate buffer (0.1 M sodium acetate, pH 4.0, containing 0.5 M NaCl), the AMA and TxLC-I-Sepharose was suspended in 10 mM Tris/HCl (pH 7.4) containing 0.8 % NaCl. The amount of immobilized protein was determined by measuring the amount of uncoupled protein in the above wash fraction by the method of Bradford [16]. The protein concentrations of the lectin columns used were as follows: AMA, 2.0 and 4.8 mg/ml; and TxLC-I, 1.0 and 2.5 mg/ml. The slurry was then packed into a capsule-type miniature column (inner diameter, 2 mm; length, 10 mm; bed volume, 31.4 μ l) (Shimadzu Co., Kyoto, Japan).

FAC

FAC was performed using an automated FAC system (FAC-1, Shimadzu Co.) as described previously [17]. Briefly, the lectin columns were slotted into stainless steel holders, and were connected to the FAC-1. The columns were equilibrated with 10 mM Tris/HCl (pH 7.4) containing 0.8 % NaCl at a flow rate of 0.125 ml/min at 25 °C. After equilibration of the columns, PA-oligosaccharides (2.5 nM) dissolved in 10 mM Tris/HCl (pH 7.4) containing 0.8 % NaCl were successively injected onto the columns. Elution of PA-oligosaccharides was monitored by fluorescence (excitation and emission wavelengths of 310 and 380 nm respectively), and the elution front (*V*) was calculated according to the method originally described by Arata et al. [18]. Retardation of the elution front relative to that of an appropriate standard (LNnT-PA), $V - V_0$, is inversely proportional to the dissociation constant, K_d , (i.e., proportional to K_a) under the employed conditions [19].

The effective ligand concentration (B_i) of the AMA column was determined by concentration-dependence analysis, using methotrexate-derivatized Man9GlcNAc2 glycan. The B_i value of an AMA column was calculated to be 2.1 nmol. According to the simplified basic-equation of FAC, i.e., $K_d = B_i/V - V_0$, K_d values were calculated.

Glycan microarray

The glycan microarrays were printed as described previously [20]. Lyophilized lectin preparations were dissolved in PBS at 1 mg/ml and labelled with TFP (tetrafluorophenyl)-Alexa Fluor[®] 488 using the Invitrogen protein labelling kit, following the manufacturer's instructions. Assuming a molar absorption coefficient of 1.5 M⁻¹ · cm⁻¹ for a 1.0 mg/ml solution for all of the proteins, the molar ratios of Alexa Fluor[®] 488 to protein are 0.34 for ASA-I (25 kDa), 1.2 for CAA (100 kDa), 4.1 for TxLC-I (100 kDa) and 0.34 for CVA (48 kDa). The labelled lectins were diluted to 0.2 mg/ml in TBS (Tris-buffered saline; 20 mM Tris/HCl, pH 7.4, 150 mM NaCl, 2 mM CaCl₂ and 2 mM MgCl₂) containing 1 % (w/v) BSA and 0.05 % Tween 20. An aliquot (70 μ l) of each labelled lectin solution was applied to a separate microarray slide and incubated under a cover slip for 1 h in a dark, humidified chamber

Table 1 List of purified and characterized two-domain GNA-related lectins

The superscript number outside the parentheses in the Molecular structure column refers to the number of subunits. ??, No sequence information available or the molecular structure is uncertain.

Species/family (agglutinin)	Lectin	Molecular structure	Specificity	Reference
ALLIACEAE				
<i>A. sativum</i> (ASA-I)	Bulb Natural mixture of isoforms	[(12 kDa + 12kDa)]	High mannose oligosaccharide chains	[11,25]
ARACEAE				
<i>A. cucullata</i>	Tuber lectin	[13.5 kDa] ⁴ , ?? [(12 kDa + 12 kDa)] ^{2*}	Not determined	[30]
<i>A. helleborifolium</i>	Tuber lectin	[13.4 kDa] ⁴ , ?? [(12 kDa + 12 kDa)] ^{2*}	Not determined	[31]
<i>A. tortuosum</i>	Tuber lectin	[13.5 kDa] ⁴ , ?? [(12 kDa + 12 kDa)] ^{2*}	Not determined	[32]
<i>A. heterophyllum</i>	Recombinant	Not characterized	Not determined	[33]
<i>A. maculatum</i> (AMA)	Tuber lectin Natural mixture of isoforms	[(12 kDa + 12 kDa)] ²	Thyrogobulin Asialofetuin	[8]
<i>C. esculenta</i> (CEA)	Tuber lectin Natural mixture of isoform	[(12 kDa + 12 kDa)] ²	Thyrogobulin Asialofetuin	[8]
<i>X. sagittifolium</i> (XSA)	Tuber lectin Natural mixture of isoforms	[(12 kDa + 12 kDa)] ²	High mannose and complex N-glycans	[8,10]
<i>D. sequina</i> (DSA)	Tuber lectin Natural mixture of isoforms	[(12 kDa + 14kDa)] ²	Thyrogobulin Asialofetuin	[8]
COLCHICACEAE				
<i>C. autumnale</i> (CAA)	Tuber Natural mixture of isoforms	[(10 kDa + 15 kDa)] ⁴	GalNAc	[14]
HYACINTHACEAE				
<i>H. hispanica</i> (SCAFet)	Bulb Natural mixture of isoforms	[28 kDa] ⁴	Complex N-linked oligosaccharides	[9]
LILIACEAE				
<i>Tulipa</i> hybrid (TxLC-I)	Bulb Natural mixture of isoforms	[(14 kDa + 14 kDa)] ⁴ & [28 kDa] ⁴ (partly)	Galactosylated triantennary N-glycans	[7]
IRIDACEAE				
<i>C. sativus</i>	Bulb Natural mixture of isoforms	[(12 kDa + 12 kDa)] ²	Yeast mannan	[34]
<i>C. vernus</i> (CVA)	Bulb Natural mixture of isoforms	[(12 kDa + 12 kDa)] ²	Man α 1,3Man high mannose-type glycans Man3GlcNAc	[13] [27] [28]

* Probable molecular structure (based on the analogy of purified/cloned orthologues from closely related species).

at room temperature (25°C). After incubation, the cover slips were gently removed in a solution of TBS containing 0.05 % Tween 20 and washed by gently dipping the slides four times in successive washes of TBS containing 0.05 % Tween 20, TBS and then deionized water. After the last wash, the slides were spun in a slide centrifuge for approx. 15 s to dry and immediately scanned in a PerkinElmer ProScanArray MicroArray scanner using an excitation wavelength of 488 nm and ImaGene software (BioDiscovery, El Segundo, CA, U.S.A.) to quantify the fluorescence emissions. The data are reported as average RFUs (relative fluorescence units) for six replicates for each glycan represented on the array.

Phylogenetic analysis

Protein sequences were aligned using the ClustalW program [21]. Parsimony analyses on the alignments were conducted with PAUP* (phylogenetic analysis using parsimony) version 4.0b10 [22]. Non-parametric bootstrap support was obtained by resampling the data 1000 times using parsimony. Heuristic searching used ten random taxon addition replicates, holding 100 trees at each step, tree bisection-reconnection branch swapping, MulTrees, Collapse and Steep Descent options, and no upper limit for trees held in memory. The phylogenetic tree is visualized using Treeillustrator [23].

RESULTS AND DISCUSSION

Overview of purified and characterized two-domain GNA-related lectins

Previous to the present study, nine proteins had been isolated that, based on the sequence of the corresponding genes, have been unambiguously identified as two-domain GNA-related lectins (Table 1, see Supplementary Table 1 at <http://www.BiochemJ.org/bj/404/bj4040051add.htm>). These lectins were purified from various tissues of *A. sativum* (garlic), *A. maculatum* L. (lords and ladies), *Colocasia esculenta* (taro), *X. sagittifolium* (yautia blanco), *Dieffenbachia sequina* (mother-in-law plant), *Crocus sativus* (saffron crocus), *Crocus vernus* (dutch crocus), *Tulipa* hybrid (tulip) and *H. hispanica* (Spanish bluebell). In addition, lectins have been isolated from *Alocasia cucullata*, *Arisaema helleborifolium* and *Arisaema tortuosum* that most probably are closely related orthologues of the two-domain Araceae lectins. Finally, N-terminal sequencing of the 10 kDa (SAANNLMFSGEALRSE-SQLV) and 15 kDa (EENNVLLTGVDLETGRSLLS) subunits strongly indicates that the CAA described previously [14] is also a typical two-domain GNA-related lectin. An overview of the molecular structure and sugar-binding specificity of these characterized two-domain lectins is presented in Table 1.

It should also be mentioned here that a two-domain GNA-related protein called putidacin L1 has been isolated from the

Table 2 List of putative two-domain GNA-related lectins retrieved in protein, genomic and transcriptome databases

Proteins listed in Table 1 are not included. (E), EST; U, unpublished; sequences retrieved from the NCBI database.

Species/family	Code	Nucleic acid	Reference
ACORACEAE			
<i>A. americanus</i>	AcoamD1	mRNA (E)	U
ARACEAE			
<i>A. macrorrhizos</i>	AlomaD1	mRNA	U
<i>A. lobatum</i>	AriLoD1	Genomic DNA	U
<i>A. heterophyllum</i>	AriHeD1	Genomic DNA	[33]
	AriHeD2	mRNA	U
<i>P. pedatisecta</i>	PinpeD1	Genomic DNA	U
<i>P. ternata</i>	PinteD1	Genomic DNA	[35]
	PinteD2	Genomic DNA	U
	PinteD3	mRNA	[36]
AGAVACEAE			
<i>Y. filamentosa</i>	YucfiD1	mRNA (E)	U
	YucfiD2	mRNA (E)	U
ALLIACEAE			
<i>A. cepa</i>	AllceD1	mRNA (E)	U
	AllceD2	mRNA (E)	U
<i>A. sativum</i>	AllsaD11	mRNA	[11]
RUSCACEAE			
<i>P. multiflorum</i>	PolmuD1	mRNA	[37]
BROMELIACEAE			
<i>A. comosus</i>	AnacoD	mRNA (E)	U
ZINGIBERACEAE			
<i>Z. officinale</i>	ZinofD1	mRNA (E)	U
	ZinofD2	mRNA (E)	U
	ZinofD3	mRNA (E)	U

bacterium *Pseudomonas* sp. BW11M1. However, no lectin or carbohydrate-binding activity could be detected in putidacin L1 preparations [24].

Other putative/expressed two-domain GNA-related proteins

Besides the above-mentioned genuine lectins, cDNAs encoding two-domain GNA-related proteins were identified in libraries prepared from garlic bulbs and *Polygonatum multiflorum* (common Solomon's seal) rhizomes (Table 2). Although the corresponding proteins were not isolated, the identification of these cDNAs indicated that two-domain GNA-related lectins or proteins might be more widespread than can be inferred from protein data. To corroborate this issue, the publicly accessible databases were screened for (i) proteins comprising two tandem arrayed GNA domains and (ii) (expressed) genes encoding such proteins. As shown in Table 2, several novel (putative) proteins were retrieved. The sequences identified in *Alocasia macrorrhizos*, *Arisaema lobatum*, *Arisaema heterophyllum*, *Pinellia pedatisecta* and *Pinellia ternata* only confirm the occurrence of two-domain GNA-related lectins within the family Araceae. However, the sequences identified in *Acorus americanus* (sweetflag), *Yucca filamentosa* (Adam's-needle, bear-grass), *Allium cepa* (onion), *Ananas comosus* (pineapple) and *Zingiber officinale* (ginger) represent novel types of putative two-domain GNA-related proteins occurring in different taxonomic groups. Moreover, a closer examination of the sequences indicates that some species like *A. cepa* and *Z. officinale* express multiple but distantly related two-domain GNA-related proteins.

Two-domain GNA-related lectins/proteins are widespread in Liliopsida but are not a monophylogenetic group

All of the collected data showed that proteins comprising two GNA domains are fairly widespread among Liliopsida

(monocots) and occur in families of at least six different orders (Acorales, Alismatales, Asparagales, Liliales, Poales and Zingiberales). Phylogenetic analysis using MacClade/PAUP yielded a fairly complex dendrogram with several obvious phylogenetic incongruities (Figure 1). Three main branches can be distinguished. Branch 1 groups YucfiD1 and YucfiD2 from *Y. filamentosa*, AllceD2 and AllceD1 from *A. cepa*, PolmuD1 from *P. multiflorum* and AnacoD from *Ananas comosus*. Branch 2 comprises, besides a fairly homogenous cluster of Araceae lectins, the agglutinins from *Tulipa*, *Crocus* sp. and one of the *Z. officinale* proteins. The rest of the sequences cluster in a heterogeneous branch with several distinct side branches. Evidently, the dendrogram does not coincide with the phylogeny of the species in which the sequences are found. This obvious phylogenetic anomaly, combined with the scattering of different sequences from some single species (e.g. *Z. officinale*) or closely related species (e.g. *A. cepa* and *A. sativum*) over different branches, strongly argues against a monophylogenetic origin of all two-domain GNA-related proteins. Accordingly, it is tempting to speculate that multiple independent domain duplications/in tandem insertions gave rise to the group of two-domain GNA-related proteins. If so one can reasonably expect that the subgroups descending from each distinct duplication/in tandem insertion differ from each other with respect to: (i) the degree of interdomain sequence divergence, and; (ii) the degree of residual sequence identity/similarity with the original single-domain lectin. To test this hypothesis the sequence identity between the N- and C-terminal domains of two-domain GNA-related proteins was determined (for a summary of the results see Supplementary Table 2 at <http://www.BiochemJ.org/bj/404/bj4040051add.htm>). Since no sequence information is available about the original duplicated GNA domains, the degree of sequence divergence of the modern two-domain proteins from their respective ancestors cannot be determined. Fortunately, in eight species (*Y. filamentosa*, *A. cepa*, *A. sativum*, *H. hispanica*, *P. multiflorum*, *Tulipa* hybrid, *A. comosus* and *Z. officinale*), belonging to seven different families classified into four different orders (Asparagales, Liliales, Poales and Zingiberales), two-domain GNA-related proteins coexist with one or more single domain lectins. This simultaneous occurrence allows an indirect estimation of the similarity/divergence between two-domain and single-domain GNA-related proteins in the very same species. No single-domain lectin could be identified in any of the Araceae species in which two-domain lectins occur. However, the identification of single-domain lectins in three other Araceae species (*Amorphophallus konjac*, *Typhonium divaricatum* and *Zantedeschia aethiopica*) allows a similar comparison to be made for single- and two-domain GNA-related lectins from the family Araceae. Since no conspecific single-domain lectin could be identified in *A. americanus* and *Crocus* species, the two-domain lectins of these species were compared with single-domain lectins that shared the highest sequence identity (see Supplementary Table 2). This sequence analysis revealed that there are marked differences in the degree of interdomain sequence conservation between the two-domain lectins, with the number of identical residues varying between 98 and < 30. Alignments of the individual domains with a single-domain lectin from the same species (or if not available with such a lectin from a related species), further indicated a striking correlation between the degree of interdomain sequence conservation and the residual sequence identity with the respective single-domain lectins.

Superposing the sequence data on the dendrogram (Figure 1), revealed a marked intra-branch homogeneity and inter-branch heterogeneity for the interdomain sequence conservation and the residual sequence identity with the respective single-domain

Figure 1 Phylogenetic tree of two-domain GNA-related proteins

Figures next to the different entries indicate the number of identical residues shared between the N- and C-terminal domains of the respective proteins. For lectin abbreviations and sequence data, see Supplementary Table 1 and Supplementary Figure 4 respectively (which can be found at: <http://www.BiochemJ.org/bj/404/bj4040051add.htm>).

lectins. All proteins clustering in branch 1 exhibit a low (< 40 AA) internal sequence conservation and share a very limited sequence identity with the single domain lectins from the same species. The proteins from branch 2 share a considerably higher interdomain sequence identity (30–43 AA and 41–50 AA for the two sub-branches respectively), as well as a higher sequence identity with the respective single-domain lectins. In the third branch, the internal sequence conservation varies between 48 and 98 AA, which confirms the heterogeneity between the different side branches. Although the dendrogram, as well as the results summarized in Supplementary Table 2, should be interpreted with care because of the limited number of sequences, all evidence points towards the occurrence of multiple duplication/in tandem insertion events. Neither the exact number nor the timing of the distinct evolutionary events can be accurately estimated. However, it seems likely that the duplication/in tandem insertion that gave rise to the proteins in branch 1 has taken place in the rather distant past, whereas the duplication/in tandem insertion leading to the garlic lectin ASA-I is of the recent past in evolutionary terms.

Comparative study of the specificity of AMA, TxCL-I, ASA-I, CVA and CAA

Irrespective of the details of the molecular evolution, it is evident that the two-domain GNA-related proteins exhibit a marked sequence divergence that is far more pronounced than that observed between single-domain GNA-related lectins from the same taxonomic group. A similar conclusion can be drawn with respect to the carbohydrate-binding specificity. Unlike

the single-domain GNA-related lectins, which all exhibit an exclusive specificity towards mannose/oligomannosides [1], the two-domain lectins cover a wide range of specificities. For example, ASA-I binds only mannose/oligomannosides [12], whereas some Araceae lectins react with an unusually broad range of glycans [10]. These observations suggest that evolution of the two-domain lectins is less conservative with respect to specificity retention than that of their single domain counterparts, which, in turn, might indicate that domain duplication/in tandem insertion followed by a pronounced divergence between the individual domains represents an evolutionary mechanism to generate multi-specific lectins. To test this hypothesis, a comparative analysis was made of the specificity of two-domain lectins with a different degree of interdomain sequence divergence using AMA, TxLC-I, ASA-I, CVA and CAA as models.

ASA-I

Previous studies using an ELISA technique and SPR demonstrated that ASA-I binds mannose and oligomannosides with a preference for α 1,2-linked mannose residues [12,25]. Man9GlcNAc2Asn, which carries several α 1,2-linked mannose residues, was the best manno-oligosaccharide ligand (binding affinity $K_a = 1.2 \times 10^6 \text{ M}^{-1}$ at 25 °C). Analyses of the interaction with glycoproteins further indicated that ASA-I binds to the core pentasaccharide of N-linked glycans.

The analysis of ASA-I using glycan array screening confirmed the strong interaction of ASA-I with a terminal α 1,2-linked mannose as well as with high mannose N-glycans comprised

Table 3 Comparison of the specificity of CAA, TxLC-I, AMA, ASA and CVA as determined by the glycan array analysis

Glycan no.	Glycan name	Binding as a percentage of maximum RFU bound				
		CAA	TxLC-I	AMA	ASA	CVA
High-mannose N-glycans						
50	Man α 1-3(Man α 1-6)Man β 1-4GlcNAc β 1-4GlcNAc β -Gly	77	83	37	99	85
189	Man α 1-2Man α 1-2Man α 1-3Man α -Sp9	<10	<10	<10	100	<10
190	Man α 1-2Man α 1-3(Man α 1-2Man α 1-6)Man α -Sp9	<10	<10	<10	35	<10
192	Man α 1-6(Man α 1-2Man α 1-3)Man α 1-6(Man α 2Man α 1-3)Man β 1-4GlcNAc β 1-4GlcNAc β -N	26	<10	49	47	26
193	Man α 1-2Man α 1-6(Man α 1-3)Man α 1-6(Man α 2Man α 2Man α 1-3)Man β 1-4GlcNAc β 1-4GlcNAc β -N	69	<10	61	40	<10
194	Man α 1-2Man α 1-2Man α 1-3(Man α 1-2Man α 1-3(Man α 1-2Man α 1-6)Man α 1-6)Man β 1-4GlcNAc β 1-4GlcNAc β -N	56	<10	45	25	<10
195	Man α 1-3(Man α 1-6)Man α -Sp9	63	<10	<10	<10	100
196	Man α 1-3(Man α 1-2Man α 1-2Man α 1-6)Man α -Sp9	0	<10	<10	57	<10
197	Man α 1-6(Man α 1-3)Man α 1-6(Man α 2Man α 1-3)Man β 1-4GlcNAc β 1-4GlcNAc β -N	49	<10	54	91	49
198	Man α 1-6(Man α 1-3)Man α 1-6(Man α 1-3)Man β 1-4GlcNAc β 1-4GlcNAc β -N	86	<10	41	82	90
199	Man5.9mix N	83	<10	60	93	98
Complex N-glycans						
51	GlcNAc β 1-2Man α 1-3(GlcNAc β 1-2Man α 1-6)Man β 1-4GlcNAc β 1-4GlcNAc β -Gly	53	74	100	<10	15
52	Gal β 1-4GlcNAc β 1-2Man α 1-3(Gal β 1-4GlcNAc β 1-2Man α 1-6)Man β 1-4GlcNAc β 1-4GlcNAc β -Gly	81	89	63	<10	<10
53	Neu5Ac α 2-6Gal β 1-4GlcNAc β 1-2Man α 1-3(Neu5Ac α 2-6Gal β 1-4GlcNAc β 1-2Man α 1-6)Man β 1-4GlcNAc β 1-4GlcNAc β -Gly	80	87	<10	<10	<10
54	Neu5Ac α 2-6Gal β 1-4GlcNAc β 1-2Man α 1-3(Neu5Ac α 2-6Gal β 1-4GlcNAc β 1-2Man α 1-6)Man β 1-4GlcNAc β 1-4GlcNAc β -Sp8	69	88	<10	<10	<10
186	GlcA β 1-6Gal β -Sp8	28	<10	<10	<10	<10
201	Neu5Ac α 2-3(Gal β 1-3GalNAc β 1-4)Gal β 1-4Glc β -Sp0	15	<10	<10	<10	<10
203	NeuAc α 2-8NeuAc α 2-8NeuAc α 2-3(GalNAc β 1-4)Gal β 1-4Glc β -Sp0	32	<10	<10	<10	<10
232	Neu5Ac α 2-3Gal β 1-4(Fuc α 1-3)GlcNAc β 1-3Gal β -Sp8	25	<10	<10	<10	<10
264	Neu5Gc α -Sp8	18	<10	<10	<10	<10
234	Neu5Ac α 2-3Gal β 1-4GlcNAc β 1-3Gal β 1-4(Fuc α 1-3)GlcNAc-Sp0	17	<10	<10	<10	<10
219	Neu5Ac α 2-3Gal β 1-3(Neu5Ac α 2-3Gal β 1-4)GlcNAc β -Sp8	16	<10	<10	<10	<10
156	GlcNAc α 1-3Gal β 1-4GlcNAc β -Sp8	16	<10	<10	<10	<10

of tri-, penta- and hexa-mannosyl structures containing the core GlcNAc β 1-4GlcNAc β 1-Asn. The glycan array is a screening method and should be used cautiously in attempting to quantify interpretations. For example, Man9GlcNAc2Asn was not the best ligand by the glycan array analysis, and the order of binding was not related to the number of terminal α 1,2-linked mannose residues (Table 3, see Supplementary Table 3 at <http://www.BiochemJ.org/bj/404/bj4040051add.htm>). This may be related to the imprecision of the printing methods or could represent the complexity of the lectin-carbohydrate interaction with the array, which is suggested by a large difference in the binding of the tri-mannosyl core with (structure 50) and without (structure 195) the chitobiose, and the similar disparity between the mannose α 1,2-terminated mannans (structures 189 and 191) that differ by a single mannose residue. It is important to note that the glycan array screening detected no binding of ASA-I to complex N-linked glycans.

AMA

In contrast with ASA-I, the specificity of AMA differs markedly from GNA and related single-domain lectins. Preliminary studies based on inhibition assays have already indicated that AMA does not bind to mannose [8], but has a strong affinity for immobilized thyroglobulin and is specifically inhibited by N-acetyllactosamine (Gal β 1,4GlcNAc) [26]. Since the three-dimensional structure of the GNA domain had not yet been resolved, no explanation could be given for the aberrant and apparently complex specificity of AMA. To check whether the complex specificity of AMA possibly relies on the simultaneous presence of two distinct carbohydrate-binding sites that recognize structurally unrelated sugars/glycans, the carbohydrate binding properties of AMA were reinvestigated using a combination of classical sugar hapten inhibition assays, SPR analysis, FAC and glycan array screening.

Affinity chromatography experiments demonstrated that AMA bound to immobilized asialofetuin as well as to an invertase–

Table 4 Conservation/lack of conservation of the key amino acid residues of the three mannose-binding sites of the GNA domain in three different two-domain GNA-related lectins

Conserved residues are indicated in bold in the subdomains I, II and III of GNA and other monocot lectin protomers. Charged residues that replace key residues are indicated by an asterisk (*). Neutral residues that replace key residues are underlined.

Lectin/domain	Sub-domain III	Sub-domain II	Sub-domain I
GNA†	Q D N V Y	Q D N V Y	Q D N V Y
ASAL‡	Q D N V Y	Q D N V Y	Q D N V Y
ASAI-N	Q D N V Y	Q D N V Y	Q D N V Y
ASAI-C	Q D N V Y	Q D N V Y	Q D N V Y
AmokoS§	Q D N V Y	Q D N V Y	Q D N V Y
AMA-N	Q D N V Y	<u>T F E</u> *V K*	H E*R*V Y
AMA-C	Q D N V Y	<u>I K</u> *E*V K*	Q D L I Y
TxLM-II‡	Q D N V Y	Q D N V Y	Q D N V Y
TxLC-I-N	Q D N V Y	<u>N N H I N</u>	R* <u>A D</u> *A Y
TxLC-I-C	<u>L K</u> *Q S S	D*R*H S L	E*G A Y Y

† Residues in sub-domain III, II and I correspond to Gln²⁶ Asp²⁸ Asn³⁰ Val³² Tyr³⁴, Gln⁵⁷ Asp⁵⁹ Asn⁶¹ Val⁶³ Tyr⁶⁵ and Gln⁸⁹ Asp⁹¹ Asn⁹³ Val⁹⁵ Tyr⁹⁷ respectively, of mature GNA.

‡ ASAL and TxLM-II are conspecific single-domain homologues of the two-domain lectins ASA-I and TxLC-I respectively.

§ AmokoS is the only single-domain homologue of AMA found in the family Araceae.

Sepharose 4B matrix. In addition, AMA gave precipitation curves with both yeast mannan and asialo-orosomucoid (results not shown). The inhibition of precipitation of the yeast mannan–AMA systems revealed that mannose oligosaccharides were the best inhibitors, in the order Man α 1,3[Man α 1,6Man]Man > Man α 1,3Man. Man α 1,2Man was a non-inhibitor at the highest level (25 mM) at which the former two disaccharides inhibited the precipitation reaction (see Supplementary Table 4 at <http://www.BiochemJ.org/bj/404/bj4040051add.htm>). Even at a level of 250 mM, methyl α -mannoside was a non-inhibitor. The best

Figure 2 FAC analysis of the carbohydrate-binding specificity of AMA (A) and TxLC-I (B)

Results are expressed as affinity constants and $V - V_0$ values for AMA and TxLC-I respectively. Numbers at the bottom of the graphs correspond to sugar numbers indicated in Supplementary Figure 3 (which can be found at: <http://www.BiochemJ.org/bj/404/bj4040051add.htm>). Asterisks (*) refer to glycans for which no data are available. The symbols used to represent pyranose rings of monosaccharides and the bars used to indicate linkage are shown in the box at the bottom of the figure.

carbohydrate ligand inhibitors were non-sialylated, triantennary oligosaccharides with *N*-acetylglucosamine (i.e., Gal β 1,4GlcNAc-) or lacto-*N*-biose (i.e. Gal β 1,3GlcNAc-) groups at the three non-reducing termini.

These results strongly suggest that AMA recognizes and binds to two distinctly different classes of carbohydrate structures: high mannose and asialo-*N*-linked oligosaccharide chains. Moreover, since in precipitin experiments each of the precipitin systems is inhibited only by 'homologous' carbohydrate structures (i.e. the yeast mannan system by mannose oligosaccharides, and the asialo-orosomucoid system by complex asialo-*N*-linked carbohydrate chains), it can be concluded that different sites are involved in the binding to high mannose and asialo-*N*-linked oligosaccharide chains respectively.

SPR analysis confirmed the results of the affinity chromatography and glycan-glycoprotein precipitation experiments. AMA strongly interacted with immobilized RNase B and immobilized

arcelin-1 (both containing exclusively high mannose *N*-glycans). Mannopentaose and mannotriose were strong inhibitors of the interaction between AMA and RNase, whereas neither mannose nor the dimannosides Man α 1,2Man, Man α 1,3Man, Man α 1,4Man and Man α 1,6Man had any effect (see Supplementary Figure 2 at <http://www.BiochemJ.org/bj/404/bj4040051add.htm>). AMA also strongly interacted with asialofetuin and fetuin (containing predominantly complex *N*-glycans). Virtually no inhibition was observed with Gal and GalNAc (or any other monosaccharide) even when applied at very high concentrations (200 $\mu\text{g/ml}$).

Glycan array screening (Table 3, see Supplementary Table 3 and Supplementary Figure 1A at <http://www.BiochemJ.org/bj/404/bj4040051add.htm>) and FAC analysis (Figure 2A) confirmed that AMA binds to both high mannose and complex *N*-linked oligosaccharide chains. In addition, FAC analysis allowed the quantification of the affinity of the lectin for the different types of glycans. Considering the substantial affinity of AMA for glycans

003 and 015 ($K_d = 322$ and $119 \mu\text{M}$ respectively), Man3GlcNAc2 is probably the minimal structure required for binding of high mannose N-glycans. However, since the lectin exhibited a much higher affinity for glycans 005 (M5A; $K_d = 31 \mu\text{M}$), 006 (M6A; $23 \mu\text{M}$), 008 (M7A; $26 \mu\text{M}$), 009 (M7B; $21 \mu\text{M}$) and 012 (M8A; $24 \mu\text{M}$), it is evident that extension of the Man α 1-3Man β branch with α 1-2Man residue substantially increases the affinity. AMA did not interact with position isomers 007 (M6C), 010 (M7D), 011 (M8B), 013 (M8C) and 014 (M9A), all of which have an α 1-2Man residue in common at the non-reducing end of Man α 1-3Man α 1-6Man β . This indicates that the addition of an α 1-2Man residue to the Man α 1-3Man α 1-6Man β branch completely abolishes the interaction with AMA.

Besides N-glycans, AMA strongly interacts with diverse complex type N-glycans (101–506). AMA exhibited a relatively high affinity for mono-antennary glycans, i.e., 102 ($K_d = 63 \mu\text{M}$), 302 ($105 \mu\text{M}$) and 402 ($30 \mu\text{M}$), but had only low affinity for, or was not reactive towards, position isomers 101 (no detectable binding), 301 ($489 \mu\text{M}$) and 401 ($152 \mu\text{M}$). This implies that AMA prefers the α 1-3 branch in mono-antennary complex-type glycans. In addition, AMA recognized with a higher affinity bi-antennary glycans like 103 ($75 \mu\text{M}$), 202 ($21 \mu\text{M}$), 307 ($100 \mu\text{M}$) and 405 ($33 \mu\text{M}$), as well as tri-antennary glycans like 105 ($20 \mu\text{M}$), 313 ($28 \mu\text{M}$) and 410 ($10 \mu\text{M}$). Interestingly, AMA was, however, unreactive towards all tetra-antennary glycans (107, 323, 413, 414, 418 and 420). Core (α 1-6) fucosylation enhanced the affinity (e.g., 003 compared with 015, 103 compared with 202, and 313 compared with 410) regardless of the types of glycans, whereas bisecting GlcNAc completely abolished the affinity (e.g., 405 compared with 406). Accordingly, α 1-6 fucosylated, tri-antennary N-linked glycan (410, $K_d = 10 \mu\text{M}$) was the best ligand for AMA, which is in agreement with the results of the inhibition assay.

In summary, the combined results from the different analytical techniques unambiguously demonstrate that AMA interacts with both high mannose and complex N-glycans, and that this unusual specificity relies on the presence of two distinct binding sites with markedly different specificities.

TxLC-I

Preliminary specificity studies revealed that TxLC-I behaves as a lectin with a complex specificity. Hapten inhibition assays of the agglutination activity of the lectin towards rabbit and human erythrocytes provided indirect evidence that the apparent complex specificity might rely on the occurrence of a mannose-binding site and an N-acetylgalactosamine-binding site, which act independently of each other [7]. SPR experiments with immobilized RNase B/arcelin and fetuin/asialofetuin confirmed the predicted dual specificity of TxLC-I, but yielded no clear inhibition data (see Supplementary Figure 2). However, FAC provided additional and more quantitative data about the specificity of TxLC-I. No B_i value could be determined for the TxLC-I column, because no suitable sugar derivative for concentration-dependence analysis was available. Therefore the specificity of TxLC-I can only be interpreted in terms of $V - V_0$ values, which are proportional to K_d values (which are inversely proportional to K_d values). FAC analysis indicated that TxLC-I recognized a broad range of complex-type glycans (101–506) including both agalactosylated and galactosylated glycans (Figure 2). This feature is consistent with the result obtained by SPR experiments using glycoproteins. None of the high-mannose type glycans was recognized by TxLC-I, except glycan 015 (α 1-6 fucosylated Man3GlcNAc2; $V - V_0 = 5.1 \mu\text{l}$). A comparison of the figures obtained with

glycans 301–323 and 401–418 indicates that α 1-6 fucosylation apparently enhances the affinity. This might explain why glycan 015 has a low but measurable affinity.

A comparison of the binding to different mono-antennary, complex-type glycans (101, 102, 301, 302, 401 and 402) revealed that TxLC-I preferred α 1-3 branched to α 1-6 branched glycans. For example, the lectin showed no detectable binding to glycans 101 and 301, but reacted well with glycans 102 ($4.9 \mu\text{l}$) and 302 ($7.3 \mu\text{l}$). TxLC-I also bound bi- and tri-antennary glycans. The affinity for these glycans apparently increased with branching number, as is illustrated by a comparison of the figures obtained for glycans 410 (tri-antennary; $96.9 \mu\text{l}$), 405 (bi-antennary, $68.4 \mu\text{l}$) and 402 (mono-antennary; $29.7 \mu\text{l}$). However, the binding was completely abolished by the addition of β 1-6GlcNAc transferred by GnT-V. In contrast with di- and tri-antennary glycans, TxLC-I did not interact with any of tetra-antennary glycans (205, 323, 413 and 418).

Glycan array screening assays largely confirmed the conclusions drawn from the SPR and FAC experiments. TxLC-I strongly reacted with sialylated and unsialylated complex N-glycans and with Man α 1-3(Man α 1-6)Man1-4GlcNAc β 1-4GlcNAc (Table 3, see Supplementary Table 3 and Supplementary Figure 1B and <http://www.BiochemJ.org/bj/404/bj4040051add.htm>). Although the latter observation leaves no doubt that TxLC-I possesses a binding site for high mannose N-glycans, the apparent inability to react with all other high mannose N-glycans present on the array indicates that this binding site exhibits a very narrow specificity range, especially when compared with that of the homologous sites from e.g. ASA-I and CVA.

CVA

Unlike AMA and TxCL-I, the two-domain lectins found in *Crocus* species are considered mannose-binding lectins. Initially the lectin from *C. vernus* was described as an α -1,3-mannosyl-mannose-recognizing protein [27]. Additional specificity studies using a solution phase method (fluorescence polarization) and three solid phase methods (flow injection, SPR and microtitre plate binding) revealed that the lectin specifically recognizes Man3GlcNAc in the N-glycan core structure [28]. To check whether the *C. vernus* lectin possibly possesses a second type of binding site, recognizing glycans other than those containing Man3GlcNAc, the specificity was further corroborated by glycan array screening experiments (Table 3, see Supplementary Table 3). These assays confirmed that CVA interacts very strongly with several types of high mannose N-glycans, but is virtually unreactive towards complex type N-glycans. Accordingly, it seems likely that the *Crocus* lectin possesses a binding site with a high affinity for high mannose N-glycans and no or only a weakly active complex N-glycan-binding site.

CAA

Preliminary studies indicated that CAA exhibits an unusual carbohydrate-binding specificity, because its agglutination activity was readily inhibited by lactose, galactose, N-acetylgalactosamine and related sugars when assayed with human red blood cells, but not in assays with rabbit erythrocytes [14]. Since a similar observation was made for TxLC-I the specificity of CAA was studied in more detail by glycan array screening (Table 3, see Supplementary Table 3). The results of this assay clearly indicated that CAA strongly interacts with both high-mannose and complex type N-glycans, and in this respect it closely resembles some of the two-domain GNA-related lectins discussed above.

Figure 3 Ribbon diagram of the GNA protomer showing the conserved residues (orange sticks) involved in the binding of methyl-mannopyranoside (MeMan, blue sticks) to the carbohydrate-binding sites of sub-domains I, II and III respectively

The network of hydrogen-bonds connecting the amino acid residues of the binding site to MeMan is shown by red dashes. The conserved valine residues occurring in the carbohydrate-binding sites are not represented.

A major conclusion to be drawn from the specificity studies described above is that most two-domain GNA-related lectins strongly interact with both high-mannose and complex N-glycans. It should be emphasized, however, that the latter lectins markedly differ from each other with respect to their fine specificity towards both high-mannose and complex N-glycans. These differences in specificity most probably account for the marked differences in biological activity between the different two-domain GNA-related lectins.

Structural basis for the complex dual specificity of two-domain GNA-related lectins: the amino acid sequences of the putative binding sites are markedly less conserved than in the single-domain GNA-related lectins

To explain the unusual and dual specificity, the sequence of the putative sugar-binding sites of AMA and TxLC-I were compared with that of the mannose-binding sites of GNA. Structural analyses of GNA and several orthologues demonstrated that the exclusive specificity towards mannose depends on the presence of three functional sites located at each of the three faces of the β -prism structure of the protomer [3,29]. Each binding site

comprises four conserved residues (Gln, Asp, Asn and Tyr) clustered in a short linear consensus sequence (QxDxNxVxY), and three such sequences are located in subdomains I, II and III of GNA (Figure 3 and Table 4). The exclusive specificity of GNA and related lectins towards the monosaccharide mannose is due to the fact that the four hydrogen-bonds that anchor the sugar to the binding site involve the axial O₂, because of the particular orientation of the side chains of the aspartic acid and asparagine residues. This particular hydrogen-bonding is reinforced by a hydrophobic interaction between the pyranose ring of mannose and a hydrophobic residue (usually valine) in the mannose-binding site.

Virtually all single-domain GNA-related lectins share the consensus sequences with GNA and accordingly contain three functional mannose-binding sites. The same applies to the N- and C-terminal domain of ASA-I, which explains why the specificity of ASA-I closely resembles that of GNA. In contrast, the sequences of all other known two-domain lectins are markedly less conserved in the amino acid stretches that build up the binding sites. For example, a closer examination of the sequence of AMA revealed that the canonical residues, Gln, Asp, Asn, Val and Tyr, required for the specific recognition of mannose

are only conserved in subdomain III of both the N-terminal and C-terminal domains (Table 4). This implies that both domains contain a fully functional mannose-binding site that presumably accounts for the strong interaction of AMA with the high-mannose N-glycans. Taking into account the results of the glycan array and FAC analyses, it seems likely that the mannose-binding sites of AMA are sufficiently extended to accommodate glycan chains of the high-mannose type. All other potential carbohydrate-binding sites of both the N-terminal and C-terminal domain of AMA definitely lack most of the residues required for the specific recognition of mannose. The replacement of some of these residues by negatively (Asp, Glu) and positively (Lys, Arg) charged residues might create binding sites capable of interacting with the charged sugars (e.g. GalNAc, Neu5Ac). Other putative binding sites of AMA contain predominantly uncharged residues at key positions and accordingly might be completely devoid of any carbohydrate-binding activity. Although only predictive, the results of the sequence alignments are in perfect agreement with the simultaneous occurrence of two distinct binding sites with a completely different specificity.

A similar conclusion can be drawn for the two-domain lectin from tulip. In TxLC-I, only subdomain III of the N-terminal domain contains all residues (Gln, Asp, Asn, Val and Tyr) required for the specific recognition of mannose, indicating that the N-terminal domain possesses a functional mannose/high-mannose N-glycan-binding site. All other potential binding sites of both the N-terminal and C-terminal domains lack most of the residues required for the specific recognition of mannose (Table 4). For this two-domain lectin, the results of sequence alignments also support the simultaneous occurrence of two distinct binding sites with completely different specificity.

Conclusions

Two-domain GNA-related lectins and/or corresponding genes are fairly widespread among monotyledonous plants. Phylogenetic analysis indicates that these proteins/genes do not represent a monophyletic group, but are the eventual result of multiple independent domain duplication/in tandem insertion events. Comparative studies revealed that the two-domain GNA-related lectins acquired a marked diversity in carbohydrate-binding specificity, which strongly contrasts with the fairly strict conservation of specificity among their single-domain counterparts. Moreover, it seems that there has been a pronounced tendency to generate binding sites with a totally different specificity within the very same lectin molecule. As a result, most of the modern two-domain GNA-related lectins interact with both high mannose and complex N-glycans and accordingly are capable of interacting with a wide variety of foreign glycoproteins. Evidently this dramatic change in specificity profoundly extends the range of target glycans of the two-domain GNA-related lectins. Furthermore the shift in specificity of some binding sites from high mannose to complex type N-glycans implies that the two-domain GNA-related lectins are primarily directed against typical animal glycans. Although circumstantial, these considerations suggest that plants developed, and are still developing, two-domain GNA-related lectins for defence purposes.

The financial support of Fund for Scientific Research-Flanders (E.J.M.V.D., grant G.0201.04) and CNRS is gratefully acknowledged (R.C., A.B., P.R.). This work was supported in part by the Consortium for Functional Glycomics under NIGMS (National Institute of General Medical Sciences), NIH Grant GM62116. Furthermore we would like to acknowledge grant GM29470 to I.J.G. We thank Dr Y. Ito and Dr K. Totani for giving us methotrexate-derived M8A-glycan. This work was supported in part by NEDO (New Energy and Industrial Technology Organization) under the METI (The Ministry of Economy, Trade, and Industry, Japan).

REFERENCES

- 1 Van Damme, E. J. M., Peumans, W. J., Barre, A. and Rougé, P. (1998) Plant lectins: a composite of several distinct families of structurally and evolutionarily related proteins with diverse biological roles. *Crit. Rev. Plant Sci.* **17**, 575–692
- 2 Van Damme, E. J. M., Kaku, H., Perini, F., Goldstein, I. J., Peeters, B., Yagi, F., Decock, B. and Peumans, W. J. (1991) Biosynthesis, primary structure and molecular cloning of snowdrop (*Galanthus nivalis* L.) lectin. *Eur. J. Biochem.* **202**, 23–30
- 3 Hester, G., Kaku, H., Goldstein, I. J. and Wright, C. S. (1995) Structure of mannose-specific snowdrop (*Galanthus nivalis*) lectin is representative of a new plant lectin family. *Nat. Struct. Biol.* **2**, 472–479
- 4 Peumans, W. J., Barre, A., Bras, J., Rougé, P., Proost, P. and Van Damme, E. J. M. (2002) The liverwort contains a lectin that is structurally and evolutionarily related to the monocot mannose-binding lectins. *Plant Physiol.* **129**, 1054–1065
- 5 Kai, G., Zhao, L., Zheng, J., Zhang, L., Miao, Z., Sun, X. and Tang, K. (2004) Isolation and characterization of a new mannose-binding lectin gene from *Taxus media*. *J. Biosci.* **29**, 399–407
- 6 Tsutsui, S., Tasumi, S., Suetake, H. and Suzuki, Y. (2003) Lectins homologous to those of monocotyledonous plants in the skin mucus and intestine of pufferfish, *Fugu rubripes*. *J. Biol. Chem.* **278**, 20882–20889
- 7 Van Damme, E. J. M., Briké, F., Winter, H. C., Van Leuven, F., Goldstein, I. J. and Peumans, W. J. (1996) Molecular cloning of two different mannose-binding lectins from tulip bulbs. *Eur. J. Biochem.* **236**, 419–427
- 8 Van Damme, E. J. M., Goossens, K., Smeets, K., Van Leuven, F., Verhaert, P. and Peumans, W. J. (1995) The major tuber storage protein of Araceae species is a lectin: characterization and molecular cloning of the lectin from *Arum maculatum* L. *Plant Physiol.* **107**, 1147–1158
- 9 Wright, L. M., Van Damme, E. J. M., Barre, A., Allen, A. K., Van Leuven, F., Reynolds, C. D., Rougé, P. and Peumans, W. J. (1999) Isolation, characterization, molecular cloning and molecular modelling of two lectins of different specificities from bluebell (*Scilla campanulata*) bulbs. *Biochem. J.* **340**, 299–308
- 10 Mo, H., Rice, K. G., Evers, D. L., Winter, H. C., Peumans, W. J., Van Damme, E. J. M. and Goldstein, I. J. (1999) *Xanthosoma sagittifolium* tubers contain a lectin with two different types of carbohydrate-binding sites. *J. Biol. Chem.* **274**, 33300–33305
- 11 Van Damme, E. J. M., Smeets, K., Torrekens, S., Van Leuven, F., Goldstein, I. J. and Peumans, W. J. (1992) The closely related homomeric and heterodimeric mannose-binding lectins from garlic are encoded by one-domain and two-domain lectin genes, respectively. *Eur. J. Biochem.* **206**, 413–420
- 12 Dam, T. K., Bachawat, K., Rani, P. G. and Surolia, A. (1998) Garlic (*Allium sativum*) lectins bind to high mannose oligosaccharide chains. *J. Biol. Chem.* **273**, 5528–5535
- 13 Van Damme, E. J. M., Houllès-Astoul, C., Barre, A., Rougé, P. and Peumans, W. J. (2000) Cloning and characterization of a monocot mannose-binding lectin from *Crocus vernus* (family Iridaceae). *Eur. J. Biochem.* **267**, 5067–5077
- 14 Peumans, W. J., Allen, A. K. and Cammue, B. P. (1986) A new lectin from meadow saffron (*Colchicum autumnale*). *Plant Physiol.* **82**, 1036–1039
- 15 Totani, K., Ihara, Y., Matsuo, I., Koshino, H. and Ito, Y. (2005) Synthetic substrates for an endoplasmic reticulum protein-folding sensor, UDP-glucose: glycoprotein glucosyltransferase. *Angew. Chem. Int. Ed. Engl.* **44**, 7950–7954
- 16 Bradford, M. M. (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein–dye binding. *Anal. Biochem.* **72**, 248–254
- 17 Hirabayashi, J. (2004) Lectin-based structural glycomics: glycoproteomics and glycan profiling. *Glycoconjugate J.* **21**, 35–40
- 18 Arata, Y., Hirabayashi, J. and Kasai, K. (2001) Application of reinforced frontal affinity chromatography and advanced processing procedure to the study of the binding property of a *Caenorhabditis elegans* galectin. *J. Chromatogr. A* **905**, 337–343
- 19 Nakamura, S., Yagi, F., Totani, K., Ito, Y. and Hirabayashi, J. (2005) Comparative analysis of carbohydrate-binding properties of two tandem repeat-type jacalin-related lectins, *Castanea crenata* agglutinin and *Cycas revoluta* leaf lectin. *FEBS J.* **272**, 2784–2799
- 20 Blixt, O., Head, S., Mondala, T., Scanlan, C., Hufelt, M. E., Alvarez, R., Bryan, M. C., Fazio, F., Calarese, D., Stevens, J. et al. (2004) Printed covalent glycan array for ligand profiling of diverse glycan binding proteins. *Proc. Natl. Acad. Sci. U.S.A.* **101**, 17033–17038
- 21 Chenna, R., Sugawara, H., Koike, T., Lopez, R., Gibson, T. J., Higgins, D. G. and Thompson, J. D. (2003) Multiple sequence alignment with the Clustal series of programs. *Nucleic Acids Res.* **31**, 3497–3500
- 22 Swofford, D. L. (2002) PAUP*: phylogenetic analysis using parsimony (* and other methods) (Sinauer Associates, Sunderland, MA), version 4
- 23 Trooskens, G., De Beule, D., Decouttere, F. and Van Criekinge, W. (2005) Phylogenetic trees: visualizing, customizing and detecting incongruence. *Bioinformatics* **21**, 3801–3802

- 24 Parret, A. H. A., Schoofs, G., Proost, P. and De Mot, R. (2003) Plant lectin-like bacteriocin from a rhizosphere-colonizing pseudomonas isolate. *J. Bacteriol.* **185**, 897–908
- 25 Bachhawat, K., Thomas, C. J., Amutha, B., Krishnasastri, M. V., Khan, M. I. and Surolia, A. (2001) On the stringent requirement of mannosyl substitution in mannooligosaccharides for the recognition by garlic (*Allium sativum*) lectin: a surface plasmon resonance study. *J. Biol. Chem.* **276**, 5541–5546
- 26 Allen, A. K. (1995) Purification and characterization of an N-acetyllactosamine-specific lectin from tubers of *Arum maculatum*. *Biochim. Biophys. Acta* **1244**, 129–132
- 27 Misaki, A., Kakuta, M., Meah, Y. and Goldstein, I. J. (1997) Purification and characterization of the α -1,3-mannosylmannose-recognizing lectin of *Crocus vernus* bulbs. *J. Biol. Chem.* **272**, 25455–25461
- 28 Oda, Y., Nakayama, K., Abdul-Rahman, B., Kinoshita, M., Hashimoto, O., Kawasaki, N., Hayakawa, T., Kakehi, K., Tomiya, N. and Lee, Y. C. (2000) *Crocus sativus* lectin recognizes Man3GlcNAc in the N-glycan core structure. *J. Biol. Chem.* **275**, 26772–26779
- 29 Chandra, N. R., Ramachandiraiah, G., Bachhawat, K., Dam, T. K., Surolia, A. and Vijayan, M. (1999) Crystal structure of a dimeric mannose-specific agglutinin from garlic: quaternary association and carbohydrate specificity. *J. Mol. Biol.* **285**, 1157–1168
- 30 Kaur, A., Kamboj, S. S., Singh, J., Saxena, A. K. and Dhuna, V. (2005) Isolation of a novel N-acetyl-D-lactosamine specific lectin from *Alocasia cucullata* (Schott.). *Biotechnol. Lett.* **27**, 1815–1820
- 31 Kaur, M., Singh, K., Rup, P. J., Saxena, A. K., Khan, R. H., Ashraf, M. T., Kamboj, S. S. and Singh, J. (2006) A tuber lectin from *Arisaema helleborifolium* Schott with anti-insect activity against melon fruit fly, *Bactrocera cucurbitae* (Coquillett) and anti-cancer effect on human cancer cell lines. *Arch. Biochem. Biophys.* **445**, 156–165
- 32 Dhuna, V., Bains, J. S., Kamboj, S. S., Singh, J., Kamboj, S. and Saxena, A. K. (2005) Purification and characterization of a lectin from *Arisaema tortuosum* Schott having in-vitro anticancer activity against human cancer cell lines. *J. Biochem. Mol. Biol.* **38**, 526–532
- 33 Zhao, X., Chen, Z., Lin, J., Kong, W., Sun, X. and Tang, K. (2006) Expression and purification of *Arisaema heterophyllum* agglutinin in *Escherichia coli*. *J. Plant Physiol.* **163**, 206–212
- 34 Escribano, J., Rubio, A., Alvarez-Orti, M., Molina, A. and Fernandez, J. A. (2000) Purification and characterization of a mannan-binding lectin specifically expressed in corms of saffron plant (*Crocus sativus* L.). *J. Agric. Food Chem.* **48**, 457–463
- 35 Lin, J., Yao, J., Zhou, X., Sun, X. and Tang, K. (2003) Expression and purification of a novel mannose-binding lectin from *Pinellia ternata*. *Mol. Biotechnol.* **25**, 215–222
- 36 Yao, J. H., Zhao, X. Y., Liao, Z. H., Lin, J., Chen, Z. H., Chen, F., Song, J., Sun, X. F. and Tang, K. X. (2003) Cloning and molecular characterization of a novel lectin gene from *Pinellia ternata*. *Cell Res.* **13**, 301–308
- 37 Van Damme, E. J. M., Barre, A., Rougé, P., Van Leuven, F., Balzarini, J. and Peumans, W. J. (1996) Molecular cloning of the lectin and a lectin-related protein from common Solomon's seal (*Polygonatum multiflorum*). *Plant Mol. Biol.* **31**, 657–672

Received 6 December 2006/30 January 2007; accepted 8 February 2007

Published as BJ Immediate Publication 8 February 2007, doi:10.1042/BJ20061819