



**HAL**  
open science

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

Els Jm van Damme, Sachiko Nakamura-Tsurata, David F Smith, Maté Ongenaert, Harry C Winter, Pierre Rougé, Irwin J Goldstein, Hanqing Mo, Junko Kominami, Raphael Culerrier, et al.

### ► To cite this version:

Els Jm van Damme, Sachiko Nakamura-Tsurata, David F Smith, Maté Ongenaert, Harry C Winter, et al.. Phylogenetic and specificity studies of two-domain GNA-related lectins: generation of multispecificity through domain duplication and divergent evolution. *Biochemical Journal*, 2007, 404 (1), pp.51-61. 10.1042/BJ20061819 . hal-00478712

**HAL Id: hal-00478712**

**<https://hal.science/hal-00478712>**

Submitted on 30 Apr 2010

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

## PHYLOGENETIC AND SPECIFICITY STUDIES OF TWO-DOMAIN GNA-RELATED LECTINS: GENERATION OF MULTISPECIFICITY THROUGH DOMAIN DUPLICATION AND DIVERGENT EVOLUTION

Els J.M. Van Damme<sup>\*</sup>, Sachiko Nakamura-Tsuruta<sup>†</sup>, David F. Smith<sup>‡</sup>, Maté Ongenaert<sup>§</sup>, Harry C. Winter<sup>||</sup>, Pierre Rougé<sup>¶</sup>, Irwin J. Goldstein<sup>||</sup>, Hanqing Mo<sup>||</sup>, Junko Kominami<sup>†,\*\*</sup>, Raphaël Culerrier<sup>¶</sup>, Annick Barre<sup>¶</sup>, Jun Hirabayashi<sup>†</sup>, Willy J. Peumans<sup>\*</sup>

<sup>\*</sup>Department of Molecular Biotechnology, Laboratory of Biochemistry and Glycobiology, Ghent University, Coupure Links 653, B-9000 Gent, Belgium

<sup>†</sup>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

<sup>‡</sup>Department of Biochemistry, Emory University School of Medicine, Atlanta, GA 30322, USA

<sup>§</sup>Department of Molecular Biotechnology, Laboratory for Bioinformatics and Computational Genomics, Ghent University, Coupure Links 653, B-9000 Gent, Belgium

<sup>||</sup>Department of Biological Chemistry, University of Michigan Medical School, Ann Arbor, Michigan 48109-0606, USA

<sup>¶</sup>Surfaces Cellulaires et Signalisation chez les Végétaux, UMR-CNRS 5546, 24 Chemin de Borde Rouge, 31326 Castanet Tolosan, France

<sup>\*\*</sup>Fine Chemical & Foods Laboratories, J-OIL MILLS, INC., 11, Kagetoricho, Totsuka-ku, Yokohama, Kanagawa 245-0064, Japan

\*Address correspondence to: Els J.M. Van Damme, Department of Molecular Biotechnology, Lab. Biochemistry and Glycobiology, Ghent University, Coupure Links 653, B-9000 Gent, Belgium, Tel. +32 92646086; Fax : +32 92646219; E-mail : [ElsJM.VanDamme@UGent.be](mailto:ElsJM.VanDamme@UGent.be)

**Short title:** Molecular evolution of the GNA-related lectins

### **Keywords**

carbohydrate-binding specificity, GNA, lectin, mannose-binding, molecular evolution

### **Abbreviations**

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, GNA: *Galanthus nivalis* (snowdrop) agglutinin, FAC: frontal affinity chromatography, PA: pyridylaminated, pNP: *p*-nitrophenyl, RFU: relative fluorescence units, TxLC-I: *Tulipa* hybrid lectin I with complex specificity

## Synopsis

A reinvestigation of the occurrence and taxonomic distribution revealed that proteins built up of protomers consisting of two in tandem arrayed domains equivalent to the *Galanthus nivalis* agglutinin (GNA) are widespread among monotyledonous 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 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 plant 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 defense purposes.

## Introduction

Carbohydrate-binding proteins comprising one or two-domains equivalent to the *Galanthus nivalis* agglutinin (GNA) form one of the major plant lectin families [1]. GNA itself was originally isolated from snowdrop (*Galanthus 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  $\beta$ -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 Amaryllidaceae, 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 consists of subunits derived from primary translation products comprising a single GNA domain of approximately 110 AA-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 in tandem arrayed GNA domains. 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 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) 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) [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 oligomannoses 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 from garlic (*Allium sativum*) bulbs, 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 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 performing analytical methods. In addition, we searched the publicly available databases for possible yet unidentified GNA orthologs. 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 [13] and CAA [14] were purified as described previously.

### **Surface plasmon resonance measurements**

The specific interaction of AMA and TxLC-I with immobilized N-glycosylated proteins was analyzed by surface plasmon resonance (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 supplemental data.

### **Frontal affinity chromatography (FAC) analysis**

#### ***Chemicals***

Pyridylaminated (PA) N-linked glycans used are listed in Supplementary Fig. 3. N-linked glycans were purchased from Takara Bio Inc. (Kyoto, Japan) and Seikagaku Co. (Tokyo, Japan). Methotrexate derivatized Man9AGlcNAc2 was generous gift from Drs. Y. Ito and K. Totani (RIKEN) [15].

#### ***Preparation of AMA, and TxLC-I columns***

Purified AMA and TxLC-I were dissolved in 50 mM NaHCO<sub>3</sub> buffer, pH 8.3, containing 0.5M NaCl (coupling buffer) and coupled to NHS-activated Sepharose 4FF (Amersham Pharmacia Biotech, Bucks, UK) 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, pH 4.0, containing 0.5 M NaCl, the AMA and TxLC-I-Sepharose was suspended in 10 mM Tris-HCl buffer, pH 7.4, containing 0.8% NaCl (TBS). The amount of immobilized protein was determined by measuring the amount of uncoupled protein in the above wash fraction by the method of Bradford. The protein concentrations of lectin columns used were as follows; AMA, 2.0 and 4.8 mg ml<sup>-1</sup>; TxLC-I, 1.0 and 2.5 mg ml<sup>-1</sup>. The slurry was then packed into a capsule-type miniature column (inner diameter, 2 mm; length, 10 mm; bed volume, 31.4 µl) (Shimadzu Co., Kyoto, Japan).

#### ***Procedure of FAC***

FAC was performed using an automated FAC system (FAC-1, Shimadzu Co.) as described previously [16]. Briefly, the lectin columns were slotted into stainless holders, and were connected to the FAC-1. The columns were equilibrated with TBS at a flow rate of 0.125 ml min<sup>-1</sup> at 25 °C. After equilibrium of the columns, PA-oligosaccharides (2.5 nM) dissolved in TBS were successively injected into the columns in an automatic manner. Elution of PA-oligosaccharides was monitored by fluorescence (excitation and emission wavelengths of 310 and 380 nm, respectively), and elution front (V) was calculated according to the method originally described by Arata et al. [17]. Retardation of elution front relative to that of an appropriate standard (LNnT-PA),  $V-V_0$ , is

inversely proportional to dissociation constant,  $K_d$ , (i.e., proportional to  $K_a$ ) under the employed condition [18].

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

### Glycan microarray

The glycan microarrays were printed as previously described [19]. Lyophilized lectin preparations were dissolved in PBS at 1 mg ml<sup>-1</sup> and labeled with tetrafluorophenyl (TFP)-Alexa Fluor 488 using the Invitrogen protein labeling kit following the manufacturers instructions. Assuming an extinction coefficient of 1.5 for a 1.0 mg ml<sup>-1</sup> solution for all of the proteins, the molar ratios of Alexa488 to protein are 0.34 for *Allium sativium* agglutinin (25 kDa); 1.2 for *Colchicum autumnale* agglutinin (100 kDa); 4.1 for *Tulipa* lectin C1 (100 kDa); and 0.34 for *Crocus vernus* agglutinin (48 kDa). The labeled lectins were diluted to 0.2 mg ml<sup>-1</sup> in Tris-buffered saline (20 mM Tris, 150 mM NaCl, 2 mM CaCl<sub>2</sub>, 2 mM MgCl<sub>2</sub>, pH 7.4) containing 1% BSA and 0.05 % Tween-20. An aliquot (70 µl) of each labeled lectin solution was applied to separate microarray slides and incubated under a cover slip for 60 min in a dark, humidified chamber at room temperature. After the incubation, the cover slips were gently removed in a solution of Tris buffered saline containing 0.05% Tween-20 and washed by gently dipping the slides 4 times in successive washes of Tris-buffered saline containing 0.05% Tween-20, Tris-buffered saline, and deionized water. After the last wash, the slides were spun in a slide centrifuge for approximately 15 sec to dry and immediately scanned in a PerkinElmer ProScanArray MicroArray Scanner using an excitation wavelength of 488 nm and ImaGene software (BioDiscovery, Inc., El Segundo, CA) to quantify fluorescence. The data are reported as average relative fluorescence units (RFU) of 6 replicates for each glycan represented on the array.

### Phylogenetic analysis

Protein sequences were aligned using the ClustalW program [20]. Parsimony analyses on the alignments were conducted with PAUP\* version 4.0b10 [21]. Nonparametric bootstrap support was obtained by resampling the data 1,000 times using parsimony.



Heuristic searching used 10 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 [22].

## Results and discussion

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

Hitherto, nine proteins have been isolated that, based on the sequence of the corresponding genes, are unambiguously identified as two-domain GNA-related lectins (Table 1, see Supplementary Table 1). These lectins were purified from various tissues of *Allium sativum* (garlic), *Arum maculatum* L. (lords and ladies), *Colocasia esculenta* (taro), *Xanthosoma sagittifolium* (yautia blanco), *Dieffenbachia sequina* (mother-in-law plant), *Crocus sativus* (saffron crocus), *Crocus vernus* (dutch crocus), *Tulipa* hybrid (tulip) and *Hyacinthoides hispanica* (Spanish bluebell). In addition, lectins have been isolated from *Alocasia cucullata*, *Arisaema helleborifolium* and *Arisaema tortuosum* that most probably are closely related orthologs of the two-domain Araceae lectins. Finally, N-terminal sequencing of the 10 kDa (SAANNLMFSGEALRSESQLV) and 15 kDa (EENNVLLTGDVLETGRSLLS) subunits strongly indicates that also the previously described *Colchicum autumnale* agglutinin (CAA) [14] is 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 bacterium *Pseudomonas* sp. BW11M1. However, no lectin or carbohydrate-binding activity could be detected in putidacin L1 preparations [23].

### *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). Though 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 tandemly 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 macrorhizos*, *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 *Allium cepa* and *Zingiber 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 data show that proteins comprising two GNA domains are fairly widespread among Liliopsida (monocots) and occur in families of at least 6 different orders (Acorales, Alismatales, Asparagales, Liliales, Poales and Zingiberales). Phylogenetic analysis using Maclade/PAUP yielded a fairly complex dendrogram with several obvious phylogenetic incongruencies (Fig. 1). Three main branches can be distinguished. Branch 1 groups YucfiD1 and YucfiD2 from *Yucca filamentosa*, AllceD2 and AllceD1 from *Allium cepa*, PolmuD1 from *Polygonatum 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 clusters 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 N- and C-terminal domains of two-domain GNA-related proteins was determined (for a summary of the results see Supplementary Table 2). 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 can not be determined. Fortunately, in 8 species (*Yucca filamentosa*, *Allium cepa*, *Allium sativum*, *Hyacinthoides hispanica*, *Polygonatum multiflorum*, *Tulipa* hybrid, *Ananas comosus* and *Zingiber officinale*) belonging to 7 different families classified in 4 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 making a similar comparison for single and two-domain GNA-related lectins from the family Araceae. Since no conspecific single domain lectin could be identified in *Acorus americanus* and *Crocus* species the two-domain lectins of these species were compared to single-domain lectins that shared the highest sequence identity (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 (Fig. 1) revealed a marked intra-branch homogeneity and inter-branch heterogeneity for what concerns the interdomain sequence conservation and the residual sequence identity with the respective single-domain 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. Though the dendrogram as well as the data 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 a rather distant past whereas that leading to the garlic lectin ASA-I is of an in evolutionary terms recent past.

***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 enzyme-linked immunosorbent assay technique and surface plasmon resonance demonstrated that ASA-I binds mannose and oligomannosides with a preference for  $\alpha$ 1,2-linked mannose residues [12,24]. Man9GlcNAc2Asn, which carries several  $\alpha$ 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  $\alpha$ 1,2-linked Mannose as well as with high mannose N-glycans comprised of tri-, penta-, and hexa-mannosyl structures containing the core GlcNAc $\beta$ 1-4GlcNAc $\beta$ 1-Asn. The glycan array is a screening method and should be used cautiously in attempting to quantitative 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  $\alpha$ 1,2-linked Mannose residues (Table 3, see Supplementary Table 3) This may be related to imprecision of printing methods or could represent complexity of the lectin-carbohydrate interaction to the array, which is suggested by large difference in binding of the tri-mannosyl core with (structure **50**) and without (structure **195**) the chitobiose and the similar disparity between the mannose  $\alpha$ 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 to ASA-I, the specificity of AMA differs markedly from GNA and related single domain lectins. Preliminary studies based on inhibition assays 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 $\beta$ 1,4GlcNAc) [25]. Since at that time the three-dimensional structure of the GNA domain was not resolved yet, 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, surface plasmon resonance analysis, frontal affinity chromatography and glycan array screening.

Affinity chromatography experiments demonstrated that AMA bound to immobilized asialofetuin as well as to an invertase-Sepharose-4B matrix. In addition, AMA gave precipitation curves with both yeast mannan and asialoorosomuroid (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 $\alpha$ 1,3[Man

$\alpha$ 1,6Man $\beta$ Man > Man  $\alpha$ 1,3Man. Man $\alpha$ 1,2Man was a non-inhibitor at the highest level (25 mM) at which the former two disaccharides inhibited the precipitation reaction (Supplementary Table 4). Even at a level of 250 mM, methyl  $\alpha$ -mannoside was a non inhibitor. The best carbohydrate ligand inhibitors were non-sialylated, triantennary oligosaccharides with N-acetylglucosamine (*i.e.* Gal $\beta$ 1,4GlcNAc-) or lacto-N-biose (*i.e.* Gal $\beta$ 1,3GlcNAc-) groups at the three nonreducing 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 asialoorosomucoid 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.

Surface plasmon resonance 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 Man nor the dimannosides Man $\alpha$ 1,2Man, Man $\alpha$ 1,3Man, Man $\alpha$ 1,4Man and Man $\alpha$ 1,6Man had any effect (see Supplementary Fig. 2). 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  $\mu$ g ml<sup>-1</sup>).

Glycan array screening (Table 3, see Supplementary Table 3 and Fig. 1A) and FAC analysis (Fig. 2A) confirmed that AMA binds to both high mannose and complex N-linked oligosaccharide chains. In addition, FAC analysis allowed quantifying 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  $\mu$ 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  $\mu$ M), **006** (M6A; 23  $\mu$ M), **008** (M7A; 26  $\mu$ M), **009** (M7B; 21  $\mu$ M) and **012** (M8A; 24  $\mu$ M) it is evident that extension of the Man $\alpha$ 1-3Man $\beta$  branch with  $\alpha$ 1-

2Man residue substantially increases the affinity. AMA did not interact with position isomers **007** (M6C), **010** (M7D), **011** (M8B) and **013** (M8C), and **014** (M9A), all of which have  $\alpha$ 1-2Man residue in common at the non-reducing end of Man $\alpha$ 1-3Man $\alpha$ 1-6Man $\beta$ . This indicates that the addition of an  $\alpha$ 1-2Man residue to the Man $\alpha$ 1-3Man $\alpha$ 1-6Man $\beta$  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 was only low or 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  $\alpha$ 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, however, AMA was unreactive towards all tetra-antennary glycans (**107**, **323**, **413**, **414**, **418** and **420**). Core ( $\alpha$ 1-6) fucosylation enhanced the affinity (*e.g.*, **003** vs. **015**, **103** vs. **202** and **313** vs. **410**) regardless of the types of glycans, whereas bisecting GlcNAc completely abolished the affinity (*e.g.*, **405** vs. **406**). Accordingly  $\alpha$ 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 a markedly different specificity.

#### *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 to the occurrence of a mannose-binding site and an N-acetylgalactosamine-binding site, which act independently of each other [7]. Surface plasmon resonance 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 Fig. 2). However, FAC provided additional and



more quantitative data about the specificity of TxLC-I. No  $B_t$  value could not 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  $K_a$  values (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 (Fig. 2). This feature is consistent with the result obtained by SPR experiments using glycoproteins. None of high-mannose type glycans was recognized by TxLC-I except glycan **015** ( $\alpha$ 1-6 fucosylated Man<sub>3</sub>GlcNAc<sub>2</sub>,  $V-V_0 = 5.1 \mu\text{l}$ ). A comparison of the figures obtained with glycans **301-323** and **401-418** indicates that  $\alpha$ 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  $\alpha$ 1-3 branched over  $\alpha$ 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 addition of  $\beta$ 1-6GlcNAc transferred by GnT-V. In contrast to 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 $\alpha$ 1-3(Man $\alpha$ 1-6)Man1-4GlcNAc $\beta$ 1-4GlcNAc (Table 3, see Supplementary Table 3 and Fig. 1B). Though 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 to 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 *Crocus vernus* was described as an  $\alpha$ -1,3-mannosylmannose-recognizing protein [26]. Additional specificity studies using a solution phase method (fluorescence polarization) and three solid phase methods (flow injection, surface plasmon resonance, and microtiter plate binding) revealed that the lectin specifically recognizes Man3GlcNAc in the N-glycan core structure [27]. To check whether the *Crocus vernus* lectin possibly possesses a second type of binding site recognizing glycans other than these 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 closely resembles some of the above discussed two-domain GNA-related lectins.

A major conclusion to be drawn from the above described specificity studies 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 for what concerns 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 was compared to that of the mannose-binding sites of GNA. Structural analyses of GNA and several orthologs demonstrated that the exclusive specificity towards Man depends on the presence of three functional sites located at each of the three faces of the  $\beta$ -prism structure of the protomer [3,28]. Each binding site comprises four conserved residues (Q, D, N and Y) clustered in a short linear consensus sequence (**QDXNXVXY**). Three such sequences are located in subdomains I, II and III, respectively, of GNA (Figure 3, Table 4). The exclusive specificity of GNA and related lectins towards the monosaccharide Man is due to the fact that the four H-bonds that anchor the sugar to the binding site involve the equatorial O2 because of the particular orientation of the side chains of the Asp (**D**) and Asn (**N**) residues. This particular H-bonding is reinforced by a hydrophobic interaction between the pyranose ring of Man and a hydrophobic residue (usually Val) in the Man-binding site.

Virtually all single domain GNA-related lectins share the consensus sequences with GNA and accordingly contain 3 functional mannoside-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 stretches that build up the binding sites. For example, a closer examination of the sequence of AMA revealed that the canonical residues Q, D, N, V and Y required for the specific recognition of Man 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 Man-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 Man. The replacement of some of these residues by negatively (D, E) and positively (K, R) 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. Though 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 (Q, D, N, V and Y) required for the specific recognition of Man 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 Man (Table 4). For this two-domain lectin also the results of sequence alignments support the simultaneous occurrence of two distinct binding sites with a completely different specificity.

## CONCLUSIONS

Two-domain GNA-related lectins and/or corresponding genes are fairly widespread among monocotyledonous plants. Phylogenetic analysis indicate that these proteins/genes do not represent a monophylogenetic 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 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. Though circumstantial, these considerations suggest that plants developed -and are still developing two-domain GNA-related lectins for defense purposes.

### **Acknowledgements**

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, NIH Grant GM62116. Furthermore we would like to acknowledge grant GM29470 to I.J.G. We thank Drs. Y. Ito and 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 evolutionary 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 evolutionary 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., Goosens, K., Smeets, K., Van Leuven, F., Verhaert, P., 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., Bachhawat, 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., Houlès-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 Hirabayashi, J. (2004) Lectin-based structural glycomics: glycoproteomics and glycan profiling. *Glycoconjugate J.* **21**: 35-40
- 17 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
- 18 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
- 19 Blixt, O., Head, S., Mondala, T., Scanlan, C., Huflejt, M.E., Alvarez, R., Bryan, M.C., Fazio, F., Calarese, D., Stevens, J., Razi, N., Stevens, D.J., Skehel, J.J., van Die, I., Burton, D.R., Wilson, I.A., Cummings, R., Bovin, N., Wong, C.H. and Paulson, J.C. (2004) Printed covalent glycan array for ligand profiling of diverse glycan binding proteins. *Proc. Natl. Acad. Sci. U.S.A.* **101**, 17033-17038

- 20 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
- 21 Swofford, D.L. (2002) PAUP\*: Phylogenetic Analysis Using Parsimony (\* and other methods) (Sinauer Associates, Sunderland, MA), version 4
- 22 Trooskens, G., De Beule, D., Decouttere, F. and Van Criekinge, W. (2005) Phylogenetic trees: visualizing, customizing and detecting incongruence. *Bioinformatics* **21**, 3801-3802
- 23 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
- 24 Bachhawat, K., Thomas, C.J., Amutha, B., Krishnasastry, 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
- 25 Allen, A.K. (1995) Purification and characterization of an N-acetyllactosamine-specific lectin from tubers of *Arum maculatum*. *Biochim. Biophys. Acta* **1244**, 129-132
- 26 Misaki, A., Kakuta, M., Meah, Y. and Goldstein, I.J. (1997) Purification and characterization of the  $\alpha$ -1,3-mannosylmannose-recognizing lectin of *Crocus vernus* bulbs. *J. Biol. Chem.* **272**, 25455-25461
- 27 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
- 28 Chandra, N.R., Ramachandraiah, 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
- 29 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
- 30 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



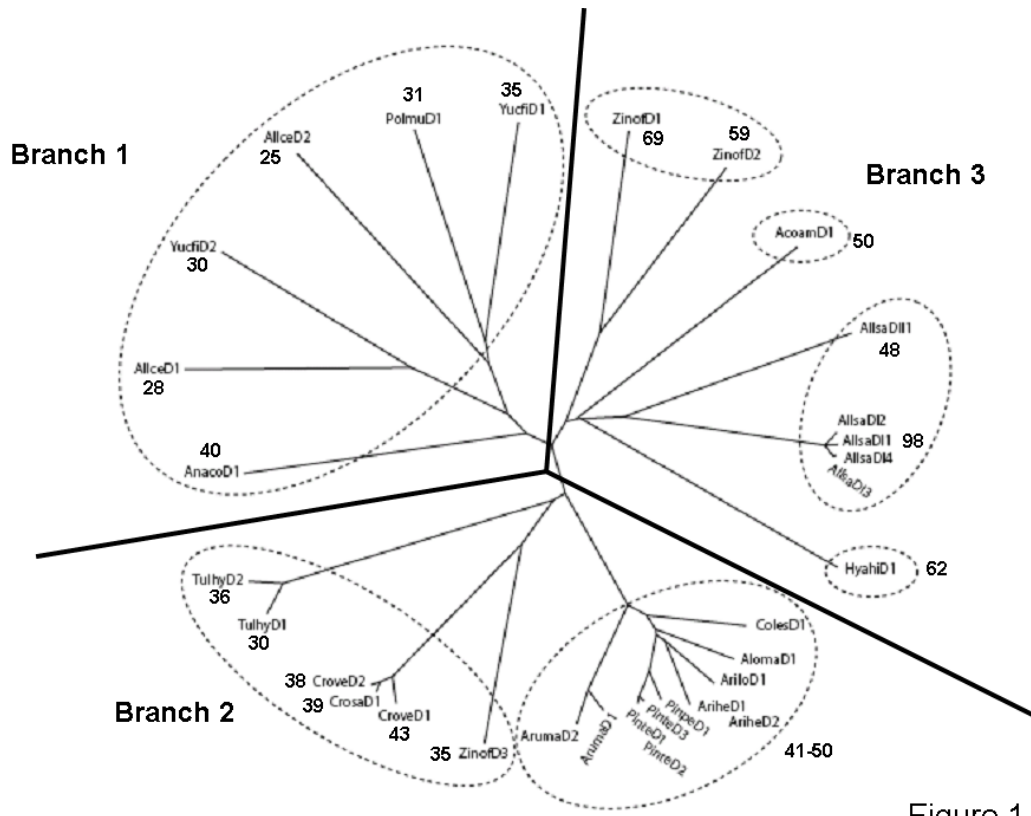
- with anti-insect activity against melon fruit fly, *Bactrocera cucurbitae* (Coquillett) and anti-cancer effect on human cancer cell lines. Arch. Biochem. Biophys. **445**, 156-165
- 31 Dhuna, V., Bains, J.S., Kamboj, S.S., Singh, J., Kamboj, S. and Saxena, A.K. (2005) Purification and characterization of a lectin from *Arisaema tortuosum* Schott having in-vitro anticancer activity against human cancer cell lines. J. Biochem. Mol. Biol. **38**, 526-532
- 32 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
- 33 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
- 34 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
- 35 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
- 36 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

## LEGENDS TO THE FIGURES

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 see Supplementary Table 1.

Figure 2. Frontal affinity chromatography 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 Fig. 3. Asterisks refer to glycans for which no data are available. Symbols used to represent pyranose rings of monosaccharides and bars used to indicate linkage are shown in the box at the bottom of the figure.

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 Val residues occurring in the carbohydrate-binding sites are not represented.



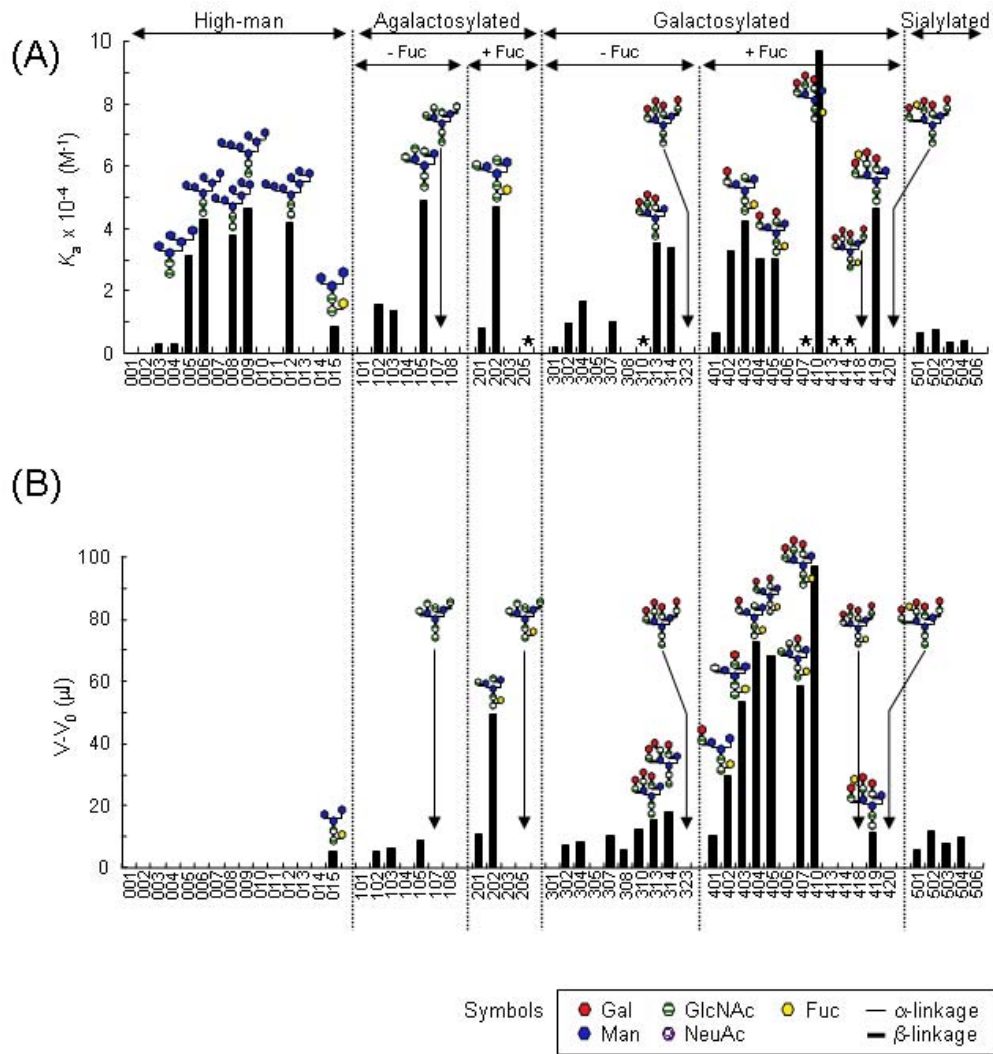
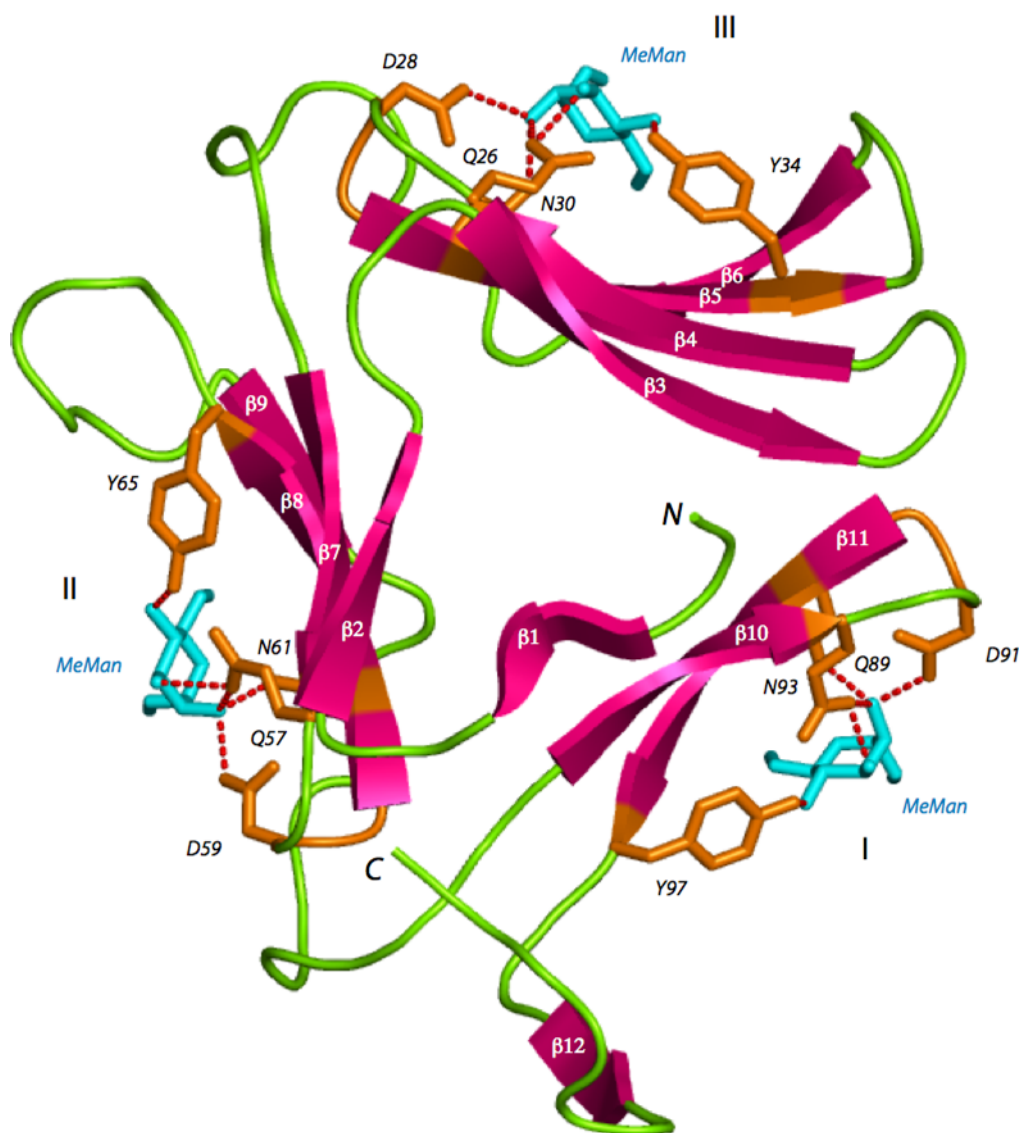


Figure 2

Figure



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

<b>Species/family</b>	<b>Lectin</b>	<b>Molecular structure</b>	<b>Specificity</b>
<b>ALLIACEAE</b>			
<i>Allium sativum</i> (ASA-I)	Bulb Natural mixture of isoforms	[(12 kDa + 12kDa)] [11]	high mannose oligosaccharide chains [24]
<b>ARACEAE</b>			
<i>Alocasia cucullata</i>	Tuber lectin	[13.5 kDa]4?? [29] [(12 kDa + 12 kDa)]2*	Not determined
<i>Arisaema helleborifolium</i>	Tuber lectin	[13.4 kDa]4?? [30] [(12 kDa + 12 kDa)]2*	Not determined
<i>Arisaema tortuosum</i>	Tuber lectin	[13.5 kDa]4?? [31] [(12 kDa + 12 kDa)]2*	Not determined
<i>Arisaema heterophyllum</i>	Recombinant	Not characterized [32]	Not determined
<i>Arum maculatum</i> (AMA)	Tuber lectin Natural mixture of isoforms	[(12 kDa + 12 kDa)]2 [8]	Thyrogobulin Asialofetuin [8]
<i>Colocasia esculenta</i> (CEA)	Tuber lectin Natural mixture of isoforms	[(12 kDa + 12 kDa)]2 [8]	Thyrogobulin Asialofetuin [8]
<i>Xanthosoma sagittifolium</i> (XSA)	Tuber lectin Natural mixture of isoforms	[(12 kDa + 12 kDa)]2 [8]	High mannose and complex N-glycans [10]
<i>Dieffenbachia sequina</i> (DSA)	Tuber lectin Natural mixture of isoforms	[(12 kDa + 14kDa)]2 [8]	Thyrogobulin Asialofetuin [8]
<b>COLCHICACEAE</b>			
<i>Colchicum autumnale</i> (CAA)	Tuber Natural mixture of isoforms	[(10 kDa + 15 kDa)]4 [14]	GalNAc [14]
<b>HYACINTHACEAE</b>			
<i>Hyacinthoides hispanica</i> (SCAFet)	Bulb Natural mixture of isoforms	[28 kDa]4 [9]	complex N-linked oligosaccharides [9]
<b>LILIACEAE</b>			
<i>Tulipa</i> hybrid (TxLC-I)	Bulb Natural mixture of isoforms	[(14 kDa + 14 kDa)]4 & [28 kDa]4 (partly) [7]	Galactosylated triantennary N-glycans
<b>IRIDACEAE</b>			
<i>Crocus sativus</i>	Bulb Natural mixture of isoforms	[(12 kDa + 12 kDa)]2 [33]	yeast mannan [33]
<i>Crocus vernus</i> (CVA)	Bulb Natural mixture of isoforms	[(12 kDa + 12 kDa)]2 [13]	Man $\alpha$ 1,3Man high mannose-type glycans [26] Man3GlcNAc [27]

?? No sequence information available; molecular structure uncertain

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

**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**

<b>Species/family</b>	<b>Code</b>	<b>Nucleic acid</b>
<b>ACORACEAE</b>		
<i>Acorus americanus</i>	AcoamD1	mRNA (E) [U]
<b>ARACEAE</b>		
<i>Alocasia macrorrhizos</i>	AlomaD1	mRNA [U]
<i>Arisaema lobatum</i>	AriLoD1	Gen DNA [U]
<i>Arisaema heterophyllum</i>	AriheD1 AriheD2	Gen DNA [32] mRNA [U]
<i>Pinellia pedatisecta</i>	PinpeD1	Gen DNA [U]
<i>Pinellia ternata</i>	PinteD1 PinteD2 PinteD3	Gen DNA [34] Gen DNA [U] mRNA [35]
<b>AGAVACEAE</b>		
<i>Yucca filamentosa</i>	YucfiD1 YucfiD2	mRNA (E) [U] mRNA (E) [U]
<b>ALLIACEAE</b>		
<i>Allium cepa</i>	AllceD1 AllceD2	mRNA (E) [U] mRNA (E) [U]
<i>Allium sativum</i>	AllsaDII1	mRNA [11]
<b>RUSCACEAE</b>		
<i>Polygonatum multiflorum</i>	PolmuD1	mRNA [36]
<b>BROMELIACEAE</b>		
<i>Ananas comosus</i>	AnacoD	mRNA (E) [U]
<b>ZINGIBERACEAE</b>		
<i>Zingiber officinale</i>	ZinofD1 ZinofD2 ZinofD3	mRNA (E) [U] mRNA (E) [U] mRNA (E) [U]

(E): EST

[U]: unpublished; sequences retrieved in NCBI database



**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 % of Maximum RFU* Bound				
		CAA	TxLC-I	AMA	ASA	CVA
<b>High-mannose N-glycans</b>						
50	Mana1-3(Mana1-6)Manβ1-4GlcNAcβ1-4GlcNAcβ-Gly	77	83	37	99	85
189	Mana1-2Mana1-2Mana1-3Mana-Sp9	<10	<10	<10	100	<10
190	Mana1-2Mana1-3(Mana1-2Mana1-6)Mana-Sp9	<10	<10	<10	35	<10
192	Mana1-6(Mana1-2Mana1-3)Mana1-6(Mana2Mana1-3)Manβ1-4GlcNAcβ1-4GlcNAcβ-N	26	<10	49	47	26
193	Mana1-2Mana1-6(Mana1-3)Mana1-6(Mana2Mana2Mana1-3)Manβ1-4GlcNAcβ1-4GlcNAcβ-N	69	<10	61	40	<10
194	Mana1-2Mana1-2Mana1-3(Mana1-2Mana1-3(Mana1-2Mana1-6)Mana1-6)Manβ1-4GlcNAcβ1-4GlcNAcβ-N	56	<10	45	25	<10
195	Mana1-3(Mana1-6)Mana-Sp9	63	<10	<10	<10	100
196	Mana1-3(Mana1-2Mana1-2Mana1-6)Mana-Sp9	0	<10	<10	57	<10
197	Mana1-6(Mana1-3)Mana1-6(Mana2Mana1-3)Manβ1-4GlcNAcβ1-4GlcNAcβ-N	49	<10	54	91	49
198	Mana1-6(Mana1-3)Mana1-6(Mana1-3)Manβ1-4GlcNAcβ1-4GlcNAcβ-N	86	<10	41	82	90
199	Man5_9mix N	83	<10	60	93	98
<b>Complex N-glycans</b>						
51	GlcNAcβ1-2Mana1-3(GlcNAcβ1-2Mana1-6)Manβ1-4GlcNAcβ1-4GlcNAcβ-Gly	53	74	100	<10	15
52	Galβ1-4GlcNAcβ1-2Mana1-3(Galβ1-4GlcNAcβ1-2Mana1-6)Manβ1-4GlcNAcβ1-4GlcNAcβ-Gly	81	89	63	<10	<10
53	Neu5Aca2-6Galβ1-4GlcNAcβ1-2Mana1-3(Neu5Aca2-6Galβ1-4GlcNAcβ1-2Mana1-6)Manβ1-4GlcNAcβ1-4GlcNAcβ-G	80	87	<10	<10	<10
54	Neu5Aca2-6Galβ1-4GlcNAcβ1-2Mana1-3(Neu5Aca2-6Galβ1-4GlcNAcβ1-2Mana1-6)Manβ1-4GlcNAcβ1-4GlcNAcβ-S	69	88	<10	<10	<10
186	GlcAβ1-6Galβ-Sp8	28	<10	<10	<10	<10
201	Neu5Aca2-3(Galβ1-3GalNAcβ1-4)Galβ1-4Glcβ-Sp0	15	<10	<10	<10	<10
203	NeuAca2-8NeuAca2-8NeuAca2-8NeuAca2-3(GalNAcβ1-4)Galβ1-4Glcβ-Sp0	32	<10	<10	<10	<10
232	Neu5Aca2-3Galβ1-4(Fuca1-3)GlcNAcβ1-3Galβ-Sp8	25	<10	<10	<10	<10
264	Neu5Gca-Sp8	18	<10	<10	<10	<10
234	Neu5Aca2-3Galβ1-4GlcNAcβ1-3Galβ1-4(Fuca1-3)GlcNAc-Sp0	17	<10	<10	<10	<10
219	Neu5Aca2-3Galβ1-3(Neu5Aca2-3Galβ1-4)GlcNAcβ-Sp8	16	<10	<10	<10	<10
156	GlcNAca1-3Galβ1-4GlcNAcβ-Sp8	16	<10	<10	<10	<10

\*RFU: Relative Fluorescence Units

**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 (\*).

Lectin/Domain	Sub-domain III	Sub-domain II	Sub-domain I
GNA <sup>a</sup>	<b>Q D N V Y</b>	<b>Q D N V Y</b>	<b>Q D N V Y</b>
ASAL <sup>b</sup>	<b>Q D N V Y</b>	<b>Q D N V Y</b>	<b>Q D N V Y</b>
ASAI-N	<b>Q D N V Y</b>	<b>Q D N V Y</b>	<b>Q D N V Y</b>
ASAI-C	<b>Q D N V Y</b>	<b>Q D N V Y</b>	<b>Q D N V Y</b>
AmokoS1 <sup>c</sup>	<b>Q D N V Y</b>	<b>Q D N V Y</b>	<b>Q D N V Y</b>
AMA-N	<b>Q D N V Y</b>	<u>T</u> <u>F</u> <u>E</u> * <b>V</b> <u>K</u> *	<u>H</u> <u>E</u> * <u>R</u> * <b>V</b> <b>Y</b>
AMA-C	<b>Q D N V Y</b>	<u>T</u> <u>K</u> * <u>E</u> * <b>V</b> <u>K</u> *	<b>Q D</b> <u>L</u> <u>I</u> <b>Y</b>
TxLM-II <sup>b</sup>	<b>Q D N V Y</b>	<b>Q D N V Y</b>	<b>Q D N V Y</b>
TxLC-I-N	<u>L</u> <u>K</u> * <u>Q</u> <u>S</u> <u>S</u>	<u>N</u> <u>N</u> <u>H</u> <u>I</u> <u>N</u>	<u>R</u> * <u>A</u> <u>D</u> * <u>A</u> <b>Y</b>
TxLC-I-C		<u>D</u> * <u>R</u> * <u>H</u> <u>S</u> <u>L</u>	<u>E</u> * <u>G</u> <u>A</u> <u>Y</u> <b>Y</b>

<sup>a</sup>Residues in sub-domain III, II and I correspond to **Q<sup>26</sup> D<sup>28</sup> N<sup>30</sup> V<sup>32</sup> Y<sup>34</sup>, Q<sup>57</sup> D<sup>59</sup> N<sup>61</sup> V<sup>63</sup> Y<sup>65</sup>** and **Q<sup>89</sup> D<sup>91</sup> N<sup>93</sup> V<sup>95</sup> Y<sup>97</sup>** respectively, of mature GNA.

<sup>b</sup>ASAL and TxLM-II are conspecific single domain homologs of the two-domain lectins ASA-I and TxLC-I, respectively.

<sup>c</sup>AmokoS is the only single domain homolog of AMA found in the family Araceae