Glycoconj J (2014) 31:345–354
DOI 10.1007/s10719-014-9527-9

Comparative analysis of carbohydrate binding properties of *Sambucus nigra* lectins and ribosome-inactivating proteins

Chenjing Shang · Els J. M. Van Damme

Received: 20 February 2014 / Revised: 29 April 2014 / Accepted: 2 May 2014 / Published online: 23 May 2014

Abstract In the past three decades a lot of research has been done on the extended family of carbohydrate-binding proteins from *Sambucus nigra*, including several so-called type 2 RIPs as well as hololectins. Although all these proteins have been studied for their carbohydrate-binding properties using hapten inhibition assays, detailed carbohydrate specificity studies have only been performed for a few *Sambucus* proteins. In particular SNA-I, has been studied extensively. Because of its unique binding characteristics this lectin was developed as an important tool in glycoconjugate research to detect sialic acid containing glycoconjugates. At present much less information is available with respect to the detailed carbohydrate binding specificity of other *S. nigra* lectins and RIPs, and as a consequence their applications remain limited. In this paper we report a comparative analysis of several lectins from *S. nigra* using the glycan microarray technology. Ultimately a better understanding of the ligands for each lectin can contribute to new/more applications for these lectins in glycoconjugate research. Furthermore, the data from glycan microarray analyses combined with the previously obtained sequence information can help to explain how evolution within a single lectin family eventually yielded a set of carbohydrate-binding proteins with a very broad specificity range.

Keywords Carbohydrate · Elderberry · Lectin · *Sambucus nigra* · Specificity

C. Shang · E. J. M. Van Damme (✉)
Department of Molecular Biotechnology,
Lab of Biochemistry and Glycobiology, Ghent University,
Coupure Links 653, 9000 Ghent, Belgium
e-mail: ElsJM.VanDamme@UGent.be

Introduction

Elderberry species played an important role in studies related to the biochemistry, molecular biology and physiology of both plant lectins and ribosome-inactivating proteins (RIPs). The first report of an elderberry lectin describing the isolation and partial characterization of a lectin from bark tissue of *Sambucus nigra* dates back to 1984 [1]. A few years later it turned out that this lectin (now referred to as SNA-I) exhibits a unique specificity towards NeuAc(α2-6)Gal/GalNAc [2] and accordingly can be used as a powerful tool in glycoconjugate research. Despite detailed biochemical studies the structure of SNA-I remained unclear until cloning of the corresponding cDNA revealed that the lectin is in fact a type-2 RIP structurally and evolutionary related to ricin [3].

In the meantime ample evidence has been presented for the occurrence of a fairly extended lectin/RIP family not only in *Sambucus nigra* but also in other elderberry species like *Sambucus ebulus* L., *Sambucus sieboldiana* Blume ex Graebn., and *Sambucus racemosa* L.[4]. Of all elderberry species, the lectin/RIP family from *S. nigra* is the best characterized. It includes several so-called type 2 RIPs as well as hololectins. For historical reasons the lectins are usually referred to by the abbreviation SNA (from *S. nigra* agglutinin) followed by Roman numbers attributed in chronological order of their discovery. Hitherto, SNA-I, SNA-II, SNA-III, SNA-IV and SNA-V have been described [1, 3, 5–10]. In addition to the genuine (*i.e.* agglutinating) lectins/RIP, a closely related type 2 RIP was identified in 1997 that exhibited no agglutination activity and did not bind to any immobilized carbohydrate or glycoprotein and accordingly was named SNLRP, which stands for *S. nigra* lectin related protein [11]. Lectins/RIPs occur in all tissues of elderberry but are particularly abundant in bark tissue, in which they represent >95 % of the total protein content and are believed to play a role in defense against herbivores.

Analysis of the purified proteins revealed marked differences in the molecular structure of the different lectins. Some lectins (SNA-II, SNA-III and SNA-IV) are homodimeric proteins composed of identical subunits, whereas others (SNA-I, SNA-V and SNLRP) are built up of two different subunits (Table 1). Molecular cloning and sequence analyses eventually revealed the relationships between the different lectins at the genetic level, and allowed classifying the different lectins into the family of ricin-related lectins or type 2 RIPs. SNA-I, SNA-V and SNLRP are type 2 RIPs, which are basically chimeric proteins built up of an enzymatically active A-chain and a lectinic B-chain held together by an inter-chain disulfide boud. Both chains are derived from a single precursor by a complex post-translational processing involving several proteolytic events [3, 8, 11]. In the case of the precursor of SNA-V an alternative processing takes place whereby not only the linker between the A- and B-chain but a sequence of 272 amino acid residues covering the complete A-chain, the linker and the first 8 amino acids of the B-chain are excised giving rise to a protein consisting solely of a slightly truncated B-chain. SNA-V and SNA-II are a rare example of two plant proteins with a totally different molecular structure that are derived from the very same precursor through a differential post-translational processing [8]. SNA-IV closely resembles SNA-II for what concerns its molecular structure. However, the origin of the SNA-IV subunits is completely different because it is not derived from a type 2 RIP precursor but from the primary translation product of a gene in which the complete A-chain, the linker sequence and the first 8 amino acid residues of the B-chain are deleted [9]. The final quaternary structure of the elderberry RIPs and lectins depends on the degree of oligomerization of the protomers. Like in all other plant lectins and type 2 RIPs non-covalent interactions mediate the oligomerization of the elderberry proteins except for SNA-I. In the latter type 2 RIP an extra Cys residue at position 47 of the B-chain allows the formation of inter-protomer disulfide bonds [3].

The availability of the purified proteins also allowed studying the biological activities of the elderberry type 2 RIPs (SNA-I, SNA-V and SNLRP) and lectins (SNA-II and SNA-IV). All three type 2 RIPs exhibit a high N-glycosidase activity *in vitro* demonstrating that they possess a catalytically active A-chain. However, their cytotoxicity towards animal and human cells is rather low and can in no way be compared with that of the genuine toxic type 2 RIPs, like ricin [4]. Most probably the low cytotoxicity of the type 2 RIP from *S. nigra* (as well as these from other elderberry species) results from a low uptake/internalization, which itself might be linked to the cell-binding properties and eventually to the sugar-binding specificity. Apart from SNLRP all elderberry type 2 RIPs and lectins agglutinate human erythrocytes and can be purified by affinity chromatography on an immobilized carbohydrate or glycoprotein [5]. Hapten inhibition assays revealed the gross-specificity and allowed classifying the different *Sambucus* lectins according to their carbohydrate-binding specificity. The hololectins SNA-IV and SNA-II, and the type 2 RIP SNA-V exhibited specificity towards Gal/GalNAc and Gal/GalNAc containing glycan structures [6, 8, 12] whereas binding assays with SNA-I revealed specificity for NeuAc(α2-6)Gal/GalNAc [2]. Since until now SNA-I (and the nearly identical orthologs from *S. canadensis* and *S. sieboldiana* [13]) is the only documented lectin that

Table 1 Overview of type 2 RIPs and lectins from *S. nigra*

Name	Source	Structure	Molecular weight	Reported carbohydrate binding specificity	Protein synthesis inhibition*	Agglutination activity**	Reference
SNA-I	Bark	Tetramer $[A\text{-}s\text{-}s\text{-}B]_4$	240 kDa (4 × 60 kDa)	Neu5Ac(α2-6)Gal/GalNAc	+	+++	[1]
SNA-I'	Bark	Dimer $[A\text{-}s\text{-}s\text{-}B]_2$	120 kDa (2 × 60 kDa)	Neu5Ac(α2-6)Gal/GalNAc	+	++	[32]
SNA-V	Bark	Dimer $[A\text{-}s\text{-}s\text{-}B]_2$	120 kDa (2 × 60 kDa)	GalNAc > Gal	+	+	[8]
SNLRP	Bark	Monomer [A-s-s-B]	62 kDa	–	+	–	[11]
SNA-II	Bark	Dimer $[B]_2$	60 kDa (2 × 30 kDa)	GalNAc > Gal	–	+	[8]
SNA-III	Seeds	Dimer $[B]_2$	60 kDa (2 × 30 kDa)	GalNAc > Gal	–	+++	[5]
SNA-IV	Fruit	Dimer $[B]_2$	64 kDa (2 × 32 kDa)	Gal/GalNAc	–	+++	[7, 9]

*, results based on hapten inhibition assays

**, +++, ++ and + indicate very strong, strong and weak agglutination activity of rabbit red blood cells; −, designates lack of activity

interacts exclusively with glycoconjugates carrying the Neu5Ac(α2-6)Gal/GalNAc linkage SNA-I was developed into an efficient analytical reagent to detect sialic acid containing glycans. Much less information is available with respect to the detailed carbohydrate binding specificity of other *S. nigra* lectins and RIPs, and as a consequence their applications remain limited. It was reported that the Gal/GalNAc binding lectin SNA-II can recognize the carcinoma Tn epitope (Ser-O-GalNAc) [14]. Since galactose-containing glycans on the cell surface are important molecules *e.g.* for the interaction between cells, or between cells and pathogens [15, 16], it is important to have some more detailed knowledge on the carbohydrate binding properties of the lectins. Therefore a comparative analysis was made of the glycan binding specificities of the type 2 RIP B-chain and lectins from *S. nigra* using glycan arrays [17]. A major advantage of the glycan microarray is the availability of many glycans that can be analysed in a single analysis using limited amounts of the immobilized glycans and of the fluorescently labelled carbohydrate-binding proteins. Ultimately a better understanding of lectin ligands can also contribute to new applications for these lectins in glycoconjugate research. In addition, glycan microarray analyses combined with the previously obtained sequence information can help to explain how evolution within a single lectin family eventually yielded a set of carbohydrate-binding proteins with such a broad specificity range.

Materials and methods

Lectins and RIPs

Apart from SNLRP all *S. nigra* proteins were purified by a combination of affinity chromatography and gel filtration as described previously [1, 7–9]. SNA-I SNA-II and SNA-V were isolated from lyophilized *S. nigra* bark whereas SNA-IV was purified from fruits. SNLRP was isolated from a lectin-depleted bark extract by ion-exchange chromatography and gel filtration [11].

Glycan array screening

Glycan microarrays were printed as described previously [18]. The printed glycan array contains a library of natural and synthetic glycan sequences representing major glycan structures of glycoproteins and glycolipids. Array version 2.0 and 2.1 with 264 and 303 glycan targets, respectively, were used for the analyses with *S. nigra* RIPs and lectins (http://www.functionalglycomics.org/ glycomics/publicdata/selectedScreens.jsp). The purified *S. nigra* proteins were labeled using the Alexa Fluor® 488 Protein Labeling Kit (Invitrogen, California, USA) following the manufacturer's instructions. The labeled proteins were applied to separate microarray slides and incubated for 60 min under a cover slip in a dark, humidified chamber at room temperature. After 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 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, USA) to quantify fluorescence. The data are reported as average relative fluorescence units (RFU) of six replicates for each glycan present on the array after removing the highest and lowest values.

Sequence alignment and phylogenetic analysis

ClustalW [19] was used for alignment of the eldeberry lectins/RIP protein sequences. The phylogenetic tree of the sequences was constructed using the constraint-based alignment tool-COBALT (http://www.ncbi.nlm.nih.gov/tools/cobalt/) [20].

Results and discussion

Carbohydrate-binding properties of type 2 RIPs B-chain and lectins from *S. nigra*

To corroborate the biological activity of the different *S. nigra* lectins and B-chains of type 2 RIPs, detailed carbohydrate-binding studies were performed using a screening of the lectins on a glycan array. As shown in Fig. 1, all *S. nigra* proteins exhibited a strong reaction towards a well-defined set of carbohydrates/glycans of the glycan array. For each analysis the 30 most reactive glycans were analyzed in detail to determine the preferred glycan motifs for each protein (Table 2).

The glycan array result for SNA-I (Fig. 1a) showed strong binding to Neu5Ac(α2-6)Gal/GalNAc, which is in agreement with previous reported publications [2]. All strongly interacting glycans contain at least one terminal sialic acid residue bound in the α2-6 linkage. Recent reports also showed a comparative analysis of SNA-I and several sialic acid binding proteins using a specific array containing sialylated glycans [21–23]. These data also clearly indicated that SNA-I is unique with respect to its carbohydrate-binding properties. SNA-I showed stronger binding to 2-keto-3-deoxy-D-glycero-D-galactononic acid (Kdn) and Kdn derivatives than the derivatives of more common Neu5Ac and Neu5Gc [22]. Of all *Sambucus nigra* lectins under study SNA-I showed the strongest binding to the glycan array which

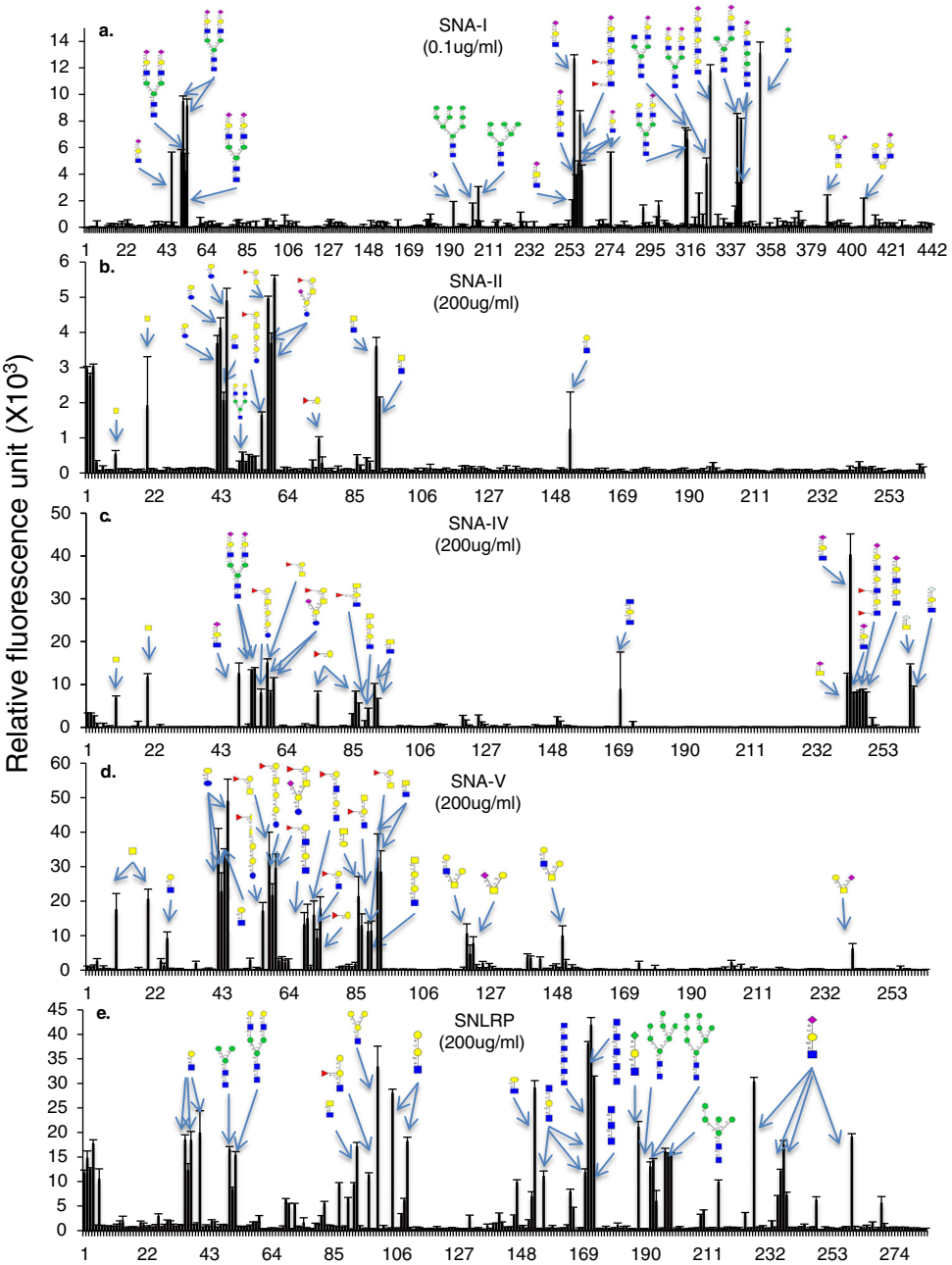


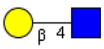
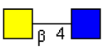
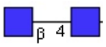
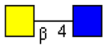
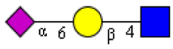
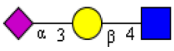
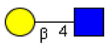
Fig. 1 Glycan array analysis of type 2 RIPs and lectins from *S. nigra*. The Consortium for Functional Glycomics website (http://www.functionalglycomic.org) supports the complete raw data for all the proteins. Sugar code used: green circles represent mannose residues; yellow circles and squares indicate galactose and *N*-Acetylgalactosamine residues, respectively; blue squares indicate GlcNAc residues, red triangles show fucose, purple diamonds indicate NeuAc and white diamonds indicate NeuGc

is in agreement with its very strong interaction with cells and glycoproteins.

In contrast to SNA-I none of the other *S. nigra* lectins tested exhibited any preference for sialic acid containing glycans. SNA-II, SNA-IV and SNA-V all clearly interacted with Gal/GalNAc containing glycans as well as with the single GalNAc residue. In the case of SNA-II and SNA-V all reactive glycans possess a terminal Gal or GalNAc residue, but substitution with (α1-2)Fuc is tolerated. Unlike SNA-II and SNA-V, SNA-IV also recognizes Gal residues in sialylated complex glycans. This sialic acid residue can occur in the α2-6 as well as α2-3 linkage. Furthermore, SNA-IV is the only protein tested that reacted with *N*-glycolylneuraminic acid (Neu5Gc) (α2-6)-linked to GalNAc [NeuGc(α2-6)GalNAc]. Although Gal and Neu5Gc are also present on the array they do not show up as interactors with any of the lectins under study.

Though the glycan array analyses of SNA-I, SNA-II, SNA-IV and SNA-V revealed previously unknown carbohydrate-binding properties there was no conflict with the "old" results based on hapten-inhibition assays. In contrast, analysis of SNLRP on the glycan array revealed that this protein can no longer be considered a "carbohydrate-binding-defective lectin related protein" since it strongly interacts with GlcNAc-oligomers (pentamer, hexamer, trimer with decreasing affinity) as well as with many glycan structures containing GlcNAc

Table 2 Glycan array data analysis. Top three glycan motifs that reacted with *S. nigra* type 2 RIPs and lectins

Protein		1. Glycan motifs	2. Glycan motifs	3. Glycan motifs
Type 2 RIP	SNA-I	Neu5Ac(α2-6)Gal(β1-4)GlcNAc 	Gal(β1-4)GlcNAcβ 	Gal(β1-4)Glcβ 
	SNA-V	Gal(β1-4)Glcβ 	Fuc(α1-2)Gal(β1-4)GlcNAcβ 	GalNAc(β1-4)GlcNAcβ 
	SNLRP	GlcNAc(β1-4)GlcNAcβ 	Gal(β1-4)GlcNAcβ 	
Lectin	SNA-II	Fuc(α1-2)Gal(β1-4)GalNAcβ 	Gal(β1-4)-Glcβ 	GalNAc(β1-4)GlcNAcβ 
	SNA-IV	Neu5Ac(α2-6)Gal(β1-4)GlcNAc  Neu5Ac(α2-3)Gal(β1-4)GlcNAc 	Gal(β1-4)GlcNAcβ 	Fuc(α1-2)Gal(β1-4)GalNAcβ 

residues substituted with Gal residues. Taken into account that SNLRP reacts with GlcNAc oligomers as well as both high mannose and complex N-glycans it seems likely that SNLRP reacts with the core structure of N-glycans. Apparently the protein does not require oligomers of GlcNAc since also glycan structures containing a single GlcNAc residue (*e.g.* Galβ1-4GlcNAc) are recognized. However, the monomeric GlcNAc was not reactive on the array.

A comparison of the glycans reacting with SNA-I, SNA-IV and SNLRP revealed that they all interact preferentially with type-2 lactosamine structures (Galβ1-4GlcNAc) (Table 2), located internally or at the terminal branches of the glycans, some of which are also sialylated. These glycan motifs are responsible for blood group determination, cell-cell recognition and adhesion processes in the higher animals [24].

Based on the results of the glycan array results and in particular on the top 30 most reactive glycans the *Sambucus* lectins can be divided into three specificity groups (Table 2). The first group comprises the lectins that recognize Gal/GalNAc containing glycans, being SNA-II, SNA-V and SNA-IV. SNA-I forms the second group and specifically reacts with glycans containing terminal sialic acid residues (α2-6) linked to Gal/GalNAc. Finally, SNLRP can be considered a GlcNAc binding lectin, representing a third specificity group.

Two other *S. nigra* lectins, which have been isolated in the past (namely SNA-III and SNA-I') were not analyzed on the glycan microarrays because the purified proteins were no longer available. However, the results of preliminary specificity studies indicate that the specificity of SNA-III is similar to that of SNA-IV whereas SNA-I' exhibits the same exclusive specificity towards Neu5Ac(α2-6)Gal/GalNAc as SNA-I.

Carbohydrate-binding sites within the ricin B domain

Lectins of the ricin B-chain type consist of two β-trefoil domains, which are binding domains found in various kinds of proteins and show a pseudo threefold symmetry formed by an α, β and γ-subdomain. It is commonly accepted that triplication of an ancestral gene [25] encoding a conserved amino acid sequence gave rise to the β-trefoil domains composing the ricin B-chain [26, 27]. The specific amino acid residues that form the actual sugar-binding site in ricin have been determined by cocrystalization of ricin with several carbohydrate and glycan structures [28]. It has been reported that both the α-subdomain of the first β-trefoil domain and the γ-subdomain of the second domain can interact with glycans. No interaction with carbohydrate molecules has been shown for the β-subdomain of the β-trefoil [29].

Sequence alignment of the full-length amino acid sequences deduced from the cDNA clones encoding lectins and the B-chains of the type 2 RIPs from *S. nigra* revealed a considerable degree of sequence conservation. Furthermore alignment with the ricin sequence also enabled a comparative

analysis of the amino acid residues in the carbohydrate binding sites (Fig. 2, Table 3).

Three out of five residues (corresponding to Asp15, Gln27 and Gln40 in ricin) forming the carbohydrate binding site in the α-subdomain of the β-trefoil of ricin are conserved between ricin and all *Sambucus* lectins under study. Trp residue 29 in the first binding site of ricin is also present in the Gal/GalNAc binding proteins SNA-II, SNA-IV and SNA-V but has been replaced by an Arg residue in SNA-I and a Leu in SNLRP (Table 3). In addition, residue Asn39 of the ricin binding site was replaced in SNLRP, but was conserved in all other *Sambucus* proteins.

Similarly three out of five amino acids (Ile238, Asn248 and Gln249 in ricin) in the carbohydrate binding site of the γ-subdomain of the β-trefoil of ricin are conserved among all sequences shown in Fig. 2. Asp227 from ricin is also present in all *Sambucus* lectins, except for SNLRP where it is replaced by a Glu residue. Tyr241 is conserved between ricin, SNA-I and SNLRP, but was replaced by a Phe residue in the Gal/GalNAc binding proteins SNA-II, SNA-IV and SNA-V.

At present SNA-II is the only lectin from *S. nigra* for which information with respect to its three-dimensional structure is available. In 2009 Maveyraud *et al.* [14] reported the X-ray structure of SNA-II in complex with Gal and five Gal-related saccharides (GalNAc, lactose, alpha1-methylgalactose, fucose, and the carcinoma-specific Tn antigen) at 1.55 Å. These structural data confirm that the Gal/GalNAc-binding residues of SNA-II are present around the conserved Asp16 in subdomain I α and Asp227 in subdomain II γ. Three-dimensional structures have also been reported for the SNA-V ortholog found in *S. ebulus* (called ebulin-I) [30]. The structures of both SNA-II and the B-chain of ebulin-I appear to be identical.

At present no structural data are available for SNA-I. However, based on the results of mutational analysis of SSA-I (the SNA-I ortholog from *S. sieboldiana*), which shares 94 % sequence identity with SNA-I and exhibits the same specificity) it was concluded that Ser197, Ala 233 and Gln234 in the B-chain of SSA-I are crucial for binding to the Neu5Ac(α2-6)Gal/GalNAc sequences [31]. If so, one has to conclude that this binding site is not confined to the γ-subdomain of the second ricin domain but covers also (part of) the β-subdomain (harboring Ser197). Though the conclusions drawn for SSA are supported by the presence of the very same amino acid residues in the sequence of the B-chain of SNA-I (being Ser197, Ala233 and Gln234) they are contradicted by the fact that in SNA-I', which binds Neu5Ac(α2-6)Gal/GalNAc almost equally well as SNA-I [32], the presumed critical residues Ser197 and Ala233 are replaced by an Asn and Asp residue, respectively. Irrespective of these contradictory results the Neu5Ac(α2-6)Gal/GalNAc binding site can be localized with certainty at the C-terminus of the B-chain because it has been demonstrated that a 22 kDa polypeptide corresponding to the C-terminal part of the B-chain exhibits the same binding properties as the parent type 2 RIP [10].

At present only few sialic acid binding lectins of plant origin have been reported. A ricin-like lectin specific for glycans terminating with the sequence Neu5Ac(α2-6)Gal was also purified from the fruiting bodies of *Polyporus squamosus* [33, 34]. Similar to SSA-I a Serine residue was shown to be a key residue for the interaction between the lectin and the Neu5Ac residue. Interestingly a novel sialic acid binding lectin was created from a naturally occurring Gal binding ricin-like lectin designated EW29Ch (C-terminal domain of earthworm 29-kDa lectin). Using random mutagenesis by error-prone PCR the lactose binding pocket of the scaffold protein EW29Ch was modified into an extended binding site for α2-6 sialic acid, confirming the close relationship between Gal binding and sialic acid binding lectin of the ricin type [35].

Another interesting observation from the glycan array data is that the different *Sambucus* lectins react with sulfated glycans, mainly 6-*O*-sulfated galactoses. Recently a few papers have shown that some basic amino acids in the ricin-like lectins are crucial for the recognition of 6-*O*-sulfated glycans. Wang *et al.* [36] have shown that modification of the terminal Galβ has a significant effect on the interaction with the *Ricinus communis* agglutinin 120, RCA120. Sulfation at the 6-*O*- or 2-*O*- positions of the terminal Galβ enhanced the binding activity of RCA120 whereas sulfation at the 4-*O*-position abolished the activity when compared with the non-sulfated residues. Similarly, Hu *et al.* [37] have shown that mutants of EW29Ch in which a Glu residue at position 20 was replaced by an Arg or Lys residue acquired the ability to recognize 6-*O*-sulfated galactose. Sequence alignment of all *Sambucus* lectin sequences (Fig. 2) indicated that basic amino acids (Arg or Lys, corresponding to Arg17 in ricin) are present in subdomain Iα of all *Sambucus* lectins under study. In addition, these basic residues are also present in subdomain II γ of most *Sambucus* lectins (except SNA-I, SNA-I' and SSA). These data are in good agreement with the observation that all *Sambucus* lectins analysed on the glycan array reacted with sulfated glycans. SNA-I, SNA-II, SNA-IV and SNA-V showed good interaction with 6-*O*-sulfated galactose residues whereas SNLRP showed interaction with 6-*O*- sulfated, 4-*O*-sulfated as well as 3-*O*-sulfated galactoses.

The complex mixture of type 2 RIP and related lectins covers an unsually broad and unique range of carbohydrate-binding specificities

Though the occurrence of multiple type 2 RIP genes within a single plant species is the rule rather than the exception the RIP/lectin family in elderberry species seems to be unique for several reasons. First, even in the absence of a completed genome sequence the total number of documented protein and cDNA sequences of *S. nigra* exceeds that found in any

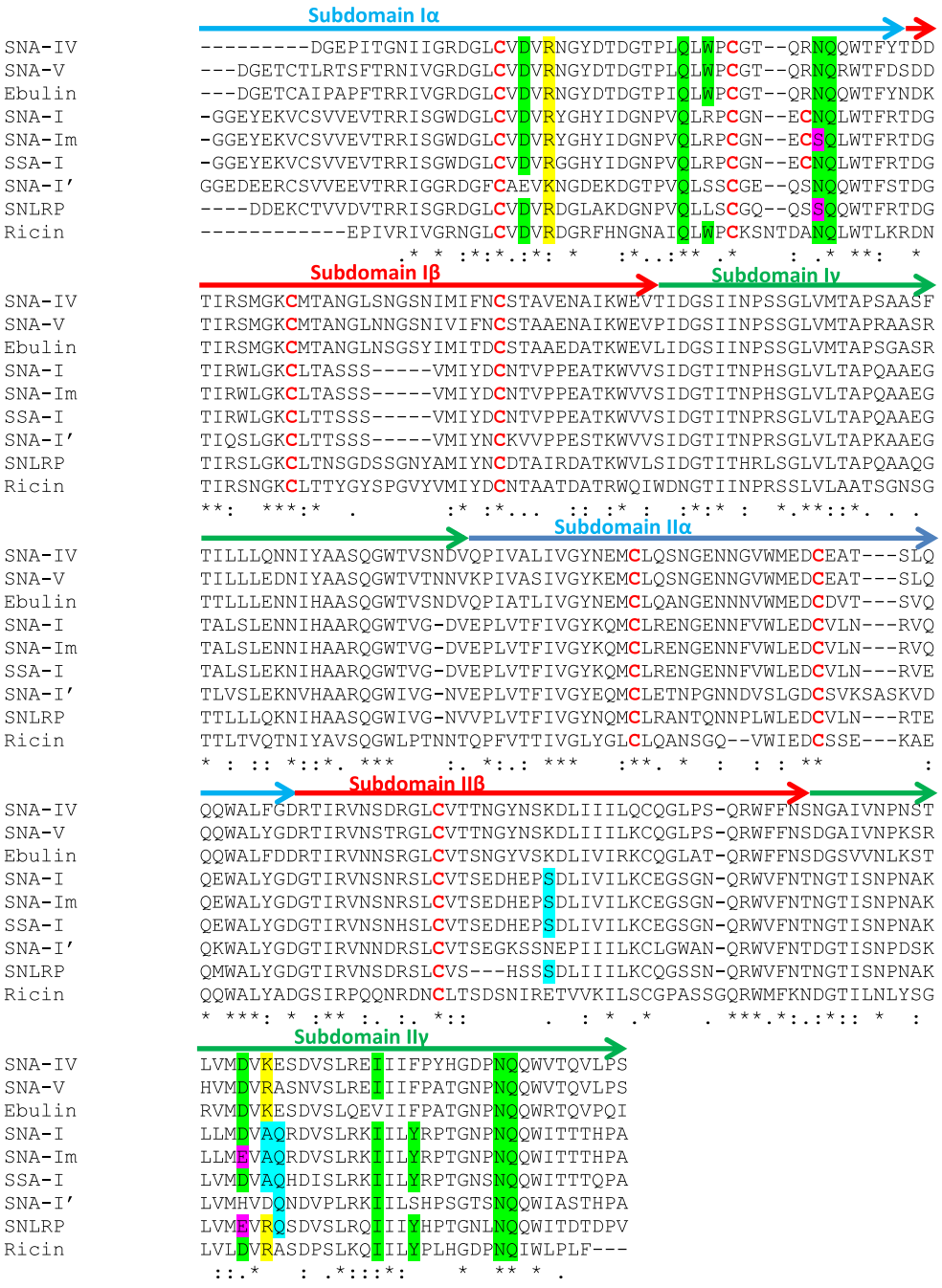


Fig. 2 Sequence alignment of the B-chains of the *S. nigra* type 2 RIP/lectins and ricin. '*' Means that the amino acids are identical in all sequences; ':' means conserved conversions (amino acids with the same shape, charge and other properties), and '.' semi-conserved substitutions (properties not the same but still similar). Amino acid residues forming the binding sites in ricin are shown in *green*; Amino acid residues replaced in SNA-I by site-directed mutagenesis to mimic the conversion of the highly active B-chain of SNA into the completely inactive B-chain of SNLRP are shown in *purple*; Residues reported to be critical for the binding to the sialic acid in Neu5A(α2-6)Gal/GalNAc sequence of 2-6-sialyllactose (according to Kaku *et al.* [31]) are indicated in *blue*. Basic residues for 6*S*-Gal binding according to Hu *et al.* [37] are highlighted in *yellow*. Cys residue involved disulphide bridges are shown in bold. Homologous subdomains (α, β, γ) are indicated by arrows

Table 3 Comparative analysis of the residues forming the carbohydrate binding sites of ricin and the type 2 RIP B-chains and lectins from *S. nigra*

RIP - B chain/lectin	Subdomain- Iα	Subdomain - IIγ
SNA-I	Asp26, Gln39, Arg41, Asn48, Gln49	Asp231, Ile243, Tyr245, Asn252, Gln253
SNLRP	Asp23, Gln36, Leu38, Ser45, Gln46	Glu230, Ile242, Tyr244, Asn251, Gln252
SNA-II	Asp16, Gln29, Trp31, Asn38, Gln39	Asp227, Ile239, Phe241, Asn248, Gln249
SNA-V	Asp24, Gln37, Trp39, Asn46, Gln47	Asp235, Ile247, Phe249, Asn256, Gln257
SNA-IV	Asp18, Gln31, Trp33, Asn40, Gln41	Asp229, Ile241, Phe243, Asn250, Gln251
Ricin	Asp15, Gln27, Trp29, Asn39, Gln40	Asp227, Ile238, Tyr241, Asn248, Gln249

other species. Secondly, *S. nigra* is the only species identified thus far in which both type 2 RIPs and lectins built up of subunits sharing a very high sequence identity with the B-chain of the type RIP have been identified. Third, the documented range of carbohydrate-binding specificities covered by the mixture of *S. nigra* RIP/lectins is much broader than that of any set of type 2 RIPs from any other species (*e.g. Ricinus communis*, *Viscum album*). This taken together with the availability of ample sequence information and detailed specificity analyses using glycan microarrays raised the question how evolution of a gene family within a single species eventually resulted in a battery of proteins capable of interacting with an extended range of carbohydrate structures.

Sequence alignments revealed a high degree of sequence identity between the B-chains of the different elderberry RIP/lectins leaving no doubt that they are all members of a single protein family. Moreover, the elderberry proteins share also a high sequence identity with ricin, and according to X-ray crystallographic studies and molecular modelling have a nearly identical overall fold. Accordingly, the very detailed information at the atomic level of the structure of the carbohydrate-binding sites in ricin was exploited to explain and interpret the results of the specificity studies of the elderberry lectins obtained by hapten inhibition assays. Though this approach was perfectly applicable to the Gal/GalNAc-binding properties of *e.g.* SNA-II, SNA-IV, SNA-V and ebulin-I, as well as for the apparent lack of sugar-binding activity of SNLRP it failed to explain the Neu5Ac(α2-6)Gal/GalNAc binding properties of SNA-I and SSA-I. Moreover, the introduction of the highly performing glycan microarray technique revealed that the specificity of ricin is not primarily directed against Gal or GalNAc but against more complex glycan structures, which implies that the fine specificity (especially in terms of recognition of complex glycans) of ricin homologs can hardly been inferred from the conservation of the "canonical" residues occurring in the two binding sites of the ricin-B-chain. This limitation is clearly illustrated by the differences in specificity between SNA-II/SNA-V and SNA-IV. Though these proteins share all "canonical" amino acid residues in both the α-subdomain of the first β-trefoil domain and the γ-subdomain of the second domain SNA-IV is capable of interacting with Gal residues in sialylated complex glycans and [NeuGc(α2-6)GalNAc], whereas SNA-II/SNA-V is not. Most probably these obvious differences in specificity rely on differences in the amino acid sequence in the respective γ-subdomains of the second domain.

SNLRP represents another example of the limitations inherent to predictions made on the basis of the presence/absence of presumed critical residues in the binding sites. Molecular modelling and docking studies with Gal and GalNAc confirmed that the presumed inactive B-chain of this type-2 RIP possessed only defective binding sites [11]. However, glycan microarray analysis revealed that SNLRP strongly interacts with GlcNAc-oligomers as well as with the core structure of N-glycans, and accordingly should be considered a genuine lectin with a specificity unprecedented among type 2 RIPs.

Since it is evident that the carbohydrate-binding properties of the different elderberry lectins are determined by the atomic structure of the binding sites and thus eventually rely on the amino acid sequence of the B-chains it seemed worthwhile to check whether and if so to what extent differences in specificity are reflected in the overall phylogeny of the RIP/lectin family. Analyses of the B-chain of the *S. nigra* type 2 RIPs and lectins allowed to investigate the sequence similarity and evolutionary relationships between the proteins under study. Furthermore it allowed checking whether the grouping of the *Sambucus* lectins in different clades could also be correlated with their carbohydrate-binding properties. Therefore, a phylogenetic tree was built of the sequences of the *S. nigra* RIP/lectins. Besides the RIP/lectins discussed here a few more *Sambucus* type-2 RIPs as well as ricin were included. As shown in Fig. 3 the dendrogram yields two major clades for the *Sambucus* protein. The first clade groups the Gal/GalNAc specific binding lectins SNA-V and SNA-IV whereas the second comprises SNA-I, SNA-Im (an inactive mutant of SNA-I in which two amino acid residues were replaced [38]), SSA-I, SNA-I' (a paralog of SNA-I missing the Cys residue involved in the inter-protomer disulfide bond formation [32]) and SNLRP [11]. Ricin forms a separate branch.

Two major conclusions can be drawn from the dendrogram. First, there is no straight forward relationship between sequence identity/similarity and carbohydrate-binding specificity. This can easily be explained by the fact that replacement of one or two amino acids within a binding site can profoundly alter the activity and specificity of a lectin but has little consequences for its phylogeny, as is illustrated by SNA-I/SNA-Im. Secondly, the sequence of the B-chain of ricin forms a separate branch indicating that all elderberry sequences

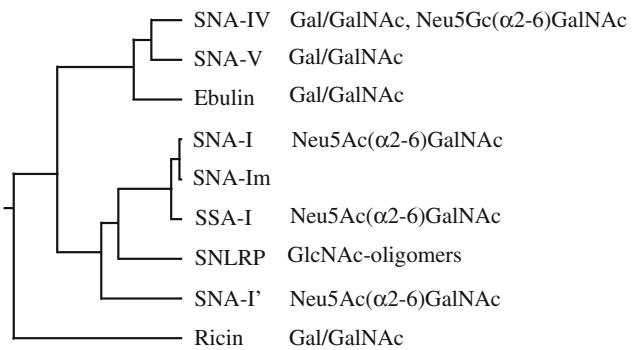


Fig. 3 Phylogenetic tree of sequences encoding type 2 RIPs and lectins from *S. nigra* and ricin. The dendrogram was made by using constraint-based alignment tool (COBALT). Accession numbers: Ricin (XP_002534649.1), SNA-I (U27122), SNA-I' (U66191), SNA-II (U41299), SNA-IV (U76523), SNA-V (U41299), SNLRP (U58357), SSA-I (D25317), Ebulin (AJ400822)

represent a monophyletic group and thus evolved from a single ancestral gene through multiple gene duplication events. It is most likely that gene duplication and/or excision events have occurred in the ancestral lectin gene. In addition amino acid changes in the carbohydrate binding sites resulted in differences in the carbohydrate-binding specificity for the different *Sambucus* lectins [39, 40]. Our data also agree with the hypothesis that sialic acid binding lectins have emerged during evolution from the Gal binding lectins, as evidenced by the fact that some trace features of Gal recognition still remain in the sialic acid binding lectins such as SNA-I [35].

Since no complete genome sequences are available the exact composition of the *S. nigra* RIP/lectin gene family cannot be determined yet. However, cDNA sequencing revealed that most -if not all- of the currently identified genes occur in multiple copies, and provided also indications for the occurrence of RIPs and lectins that have not been identified yet. Though only a preliminary analysis can be made of the phylogeny and the evolution it seems likely that (as was suggested previously by Kaku *et al.* [31]) a gene encoding a Gal/GalNAc-binding lectin was at the basis of the current elderberry RIP/lectin gene family. Genome and transcriptome sequencing programs provided evidence for the occurrence of (extended) type-2 RIP gene families in other species (*e.g. Ricinus communis*, *Viscum album*). However, the range of carbohydrate-binding specificites covered by the whole set of type-2 RIP and related lectins is apparently far more narrow than in elderberry.

Evidently, the simultaneous occurrence of large amounts of three different type 2 RIPs raises the question of the physiological importance of these proteins for elderberry. At present, one can only speculate about the evolutionary advantage of possessing such a mixture of carbohydrate-binding proteins. Most likely, a synergistic effect between the different RIP/lectins eventually results in an increased resistance to phytophagous insects and/or herbivorous animals. Indeed, over the years several experiments have shown that several RIPs and lectins from *S. nigra* play an important role in plant defense against pathogens [41] and insects [42, 43], or show a more or less selective cytotoxicity towards tumor cells [44]. Mutational analysis with SNA-I have shown that especially the B-chain of the protein, and thus the carbohydrate binding properties of the protein play an important role in the toxicity to cells and organisms. The variety of glycan-binding specificities from B-chains of type 2 RIPs and lectins from *S. nigra* indicates that they will play an important role in the protein-carbohydrate recognition taken place at the surface of cells. For instance first evidence already showed interaction of different *Sambucus* lectins with insect and animal cells [45, 46] and proved the importance of the carbohydrate binding activity in the interaction with the cell surface [42, C. Shang, unpublished data]. More research is needed to identify the interacting proteins for the *Sambucus* lectins on the cell and unravel how the lectins are internalized. Surely, a better knowledge of the glycans interacting with the proteins will enable to decipher the biological activity of these proteins and will contribute to future applications of the lectins in glycoconjugate research.

Acknowledgments This work was funded primarily by the Fund for Scientific Research-Flanders (FWO grants G.0022.08) and the Research Council of Ghent University (project BOF10/GOA/003). Chenjing Shang is recipient of China Scholarship Council and receives doctoral co-funding from the Special Research Council of Ghent University. The authors want to thank the Consortium for Functional Glycomics funded by the NIGMS GM62116 for the glycan array analysis. Special thanks to Willy J. Peumans for purification of the proteins and critical reading of the manuscript.

References

1. Broekaert, W.F., Nsimba-Lubaki, M., Peeters, B., Peumans, W.J.: A lectin from elder (*Sambucus nigra* L.) bark. Biochem. J. **221**, 163–169 (1984)
2. Shibuya, N., Goldstein, I.J., Broekaert, W.F., Nsimba-Lubaki, M., Peeters, B., Peumans, W.J.: The elderberry (*Sambucus nigra* L.) bark lectin recognizes the NeuAc(α2,6)-Gal/GalNAc sequence. J. Biol. Chem. **262**, 1596–1601 (1987)
3. Van Damme, E.J.M., Barre, A., Rougé, P., Van Leuven, F., Peumans, W.J.: The NeuAc(a-2,6) Gal/Gal-NAc-binding lectin from elderberry (*Sambucus nigra*) bark, a type-2 ribosome-inactivating protein with an unusual specificity and structure. Eur. J. Biochem. **235**, 128–137 (1996)
4. Girbés, T., Ferreras, J.M., Arias, F.J., Muñoz, R., Iglesias, R., Jimenez, P., Rojo, M.A., Arias, Y., Perez, Y., Benitez, J., Sanchez, D., Gayoso, M.J.: Non-toxic type 2 ribosome-inactivating proteins (RIPs) from Sambucus: occurrence, cellular and molecular activities and potential uses. Cell. Mol. Biol. **49**, 537–545 (2003)
5. Peumans, W.J., Kellens, J.T.C., Allen, A.K., Van Damme, E.J.M.: Isolation and characterization of a seed lectin from elderberry (*Sambucus nigra*) and its relationship to the bark lectins. Carbohydr. Res. **213**, 7–17 (1991)
6. Kaku, H., Peumans, W.J., Goldstein, I.J.: Isolation and characterization of a second lectin (SNA-II) present in elderberry (*Sambucus nigra* L.) bark. Arch. Biochem. Biophys. **277**, 255–262 (1990)
7. Mach, L., Waltraud, S., Ammann, M., Poetsch, J., Bertsch, W., Marz, L., Glossl, J.: Purification and partial characterization of a novel lectin from elder (*Sambucus nigra* L.) fruit. Biochem. J. **278**, 667–671 (1991)
8. Van Damme, E.J.M., Barre, A., Rougé, P., Leuven, F.V., Peumans, W.J.: Characterization and molecular cloning of *Sambucus nigra* agglutinin V (nigrin b), a GalNAc-specific type-2 ribosome-inactivating protein from the bark of elderberry (*Sambucus nigra*). Eur. J. Biochem. **237**, 505–513 (1996)
9. Van Damme, E.J.M., Roy, S., Rarre, A., Rougé, P., Van Leuven, F., Peumans, W.J.: The major elderberry fruit protein is a lectin derived from a truncated type 2 ribosome-inactivating protein. Plant J. **12**, 1251–1260 (1997)
10. Peumans, W.J., Roy, S., Barre, A., Rougé, P., Van Leuven, F., Van Damme, E.J.: Elderberry (*Sambucus nigra*) contains truncated Neu5Ac(α-2,6)Gal/GalNAc-binding type 2 ribosome-inactivating proteins. FEBS Lett. **425**, 35–39 (1998)
11. Van Damme, E.J.M., Barre, A., Rougé, P., Van Leuven, F., Peumans, W.J.: Isolation and molecular cloning of a novel type 2 ribosome-

inactivating protein with an inactive B-chain from elderberry (*Sambucus nigra*) bark. J. Biol. Chem. **272**, 8353–8360 (1997)
12. Mach, L., Kerschbaumer, R., Schwihla, H., Glossl, J.: Elder (*Sambucus nigra* L.) fruit lectin (SNA-IV) occurs in monomeric, dimeric and oligomeric isoforms. Biochem. J. **315**, 1061 (1996)
13. Shibuya, N., Tazaki, K., Song, Z.W., Tarr, G.E., Goldstein, I.J., Peumans, W.J.: A comparative study of bark lectins from three elderberry (*Sambucus*) species. J. Biochem. **106**, 1098–1103 (1989)
14. Maveyraud, L., Niwa, H., Guillet, V., Svergun, D.I., Konarev, P.V., Palmer, R.A., Peumans, W.J., Rouge, P., Van Damme, E.J.M., Reynolds, C.D., Mourey, L.: Structural basis for sugar recognition, including the Tn carcinoma antigen, by the lectin SNA-II from *Sambucus nigra*. Proteins **75**, 89–103 (2009)
15. Rhodes, J.M., Campbell, B.J., Yu, L.G.: Lectin-epithelial interactions in the human colon. Biochem. Soc. Trans. **36**, 1482–1486 (2008)
16. Boscher, C., Dennis, J.W., Nabi, I.R.: Glycosylation, galectins and cellular signaling. Curr. Opin. Cell Biol. **23**, 383–392 (2011)
17. Van Damme, E.J.M., Smith, D.F., Cummings, R., Peumans, W.J.: Glycan arrays to decipher the specificity of plant lectins. Adv. Exp. Med. Biol. **705**, 757–768 (2011)
18. 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., Paulson, J.C.: Printed covalent glycan array for ligand profiling of diverse glycan binding proteins. Proc. Natl. Acad. Sci. U. S. A. **101**, 17033–17038 (2004)
19. Thompson, J.D., Higgins, D.G., Gilson, T.J.: CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. **22**, 4673–4680 (1994)
20. Papadopoulos, J.S., Agarwala, R.: COBALT: constraint-based alignment tool for multiple protein sequences. Bioinformatics **23**, 1073–1079 (2007)
21. Smith, D.F., Song, X., Cummings, R.D.: Use of glycan microarrays to explore specificity of glycan-binding proteins. Methods Enzymol. **480**, 417–444 (2010)
22. Song, X., Yu, H., Chen, X., Lasanajak, Y., Tappert, M.M., Air, G.M., Tiwari, V.K., Cao, H., Chokhawala, H.A., Zheng, H., Cummings, R.D., Smith, D.F.: A sialylated glycan microarray reveals novel interactions of modified sialic acids with proteins and viruses. J. Biol. Chem. **286**, 31610–31622 (2011)
23. Padler-Karavani, V., Song, X., Yu, H., Hurtado-Ziola, H., Huang, S., Muthana, S., Chokhawala, H.A., Cheng, J., Verhagen, A., Langereis, M.A., Kleene, R., Schachner, M., De Groot, R.J., Lasanajak, Y., Matsuda, H., Schwab, R., Chen, X., Smith, D.F., Cummings, R.D., Varki, A.: Cross-comparison of protein recognition of sialic acid diversity on two novel sialoglycan microarrays. J. Biol. Chem. **287**, 22593–22608 (2012)
24. Stanley, P., Cummings, R.D.: Structures common to different glycans. In: Varki, A., Cummings, R.D., Esko, J.D., Freeze, P., Stanley, P., Bertozzi, C.R., Hart, G.W., Etzler, M.E. (eds.) Essentials of Glycobiology, 2nd edn, pp. 175–198. Cold Spring Harbor Laboratory Press, NY (2009)
25. Hazes, B.: The $(QxW)_3$ domain: A flexible lectin scaffold. Protein Sci. **5**, 1490–1501 (1996)
26. Robertus, J.D., Ready, M.P.: Ricin B-chain and discoidin I share a common primitive protein fold. J. Biol. Chem. **259**, 13953–13956 (1984)
27. Rutenber, E., Ready, M., Robertus, J.: Structure and evolution of ricin B-chain. Nature **326**, 624–626 (1987)
28. Montfort, W., Villafranca, J.E., Monzingo, A.F., Ernst, S.R., Katzin, B., Rutenber, E., Xuong, N.H., Hamlin, R., Robertus, J.D.: The three-dimensional structure of ricin at 2.8 Å. J. Biol. Chem. **262**, 5398–5403 (1987)
29. Rutenber, E., Roberus, J.D.: Structure of ricin B-chain at 2.5 Å resolution. Proteins **10**, 240–250 (1991)
30. Pascal, J.M., Day, P.J., Monzingo, A.F., Ernst, S.R., Robertus, J.D., Iglesias, R., Pérez, Y., Férreras, J.M., Citores, L., Girbés, T.: 2.8-Å Crystal structure of a nontoxic type-II ribosome-inactivating protein, ebulin l. proteins: structure, function. Genetics **43**, 319–326 (2001)
31. Kaku, H., Kaneko, H., Minamihara, N., Iwata, K., Jordan, E.T., Rojo, M.A., Minami-Ishii, N., Minami, E., Hisajima, S., Shibuya, N.: Elderberry bark lectins evolved to recognize Neu5Acα2,6Gal/GalNAc sequence from a Gal/GalNAc binding lectin through the substitution of amino-acid residues critical for the binding to sialic acid. Biochem. J. **142**, 393–401 (2007)
32. Van Damme, E.J.M., Roy, S., Barre, A., Citores, L., Mostafapous, K., Rougé, P., Van Leuven, F., Girbés, T., Goldstein, I.J., Peumans, W.J.: Elderberry (*Sambucus nigra*) bark contains two structurally different Neu5Ac(α-2,6)Gal/GalNAc-binding type 2 ribosome-inactivating proteins. Eur. J. Biochem. **245**, 648–655 (1997)
33. Mo, H., Winter, H.C., Goldstein, I.J.: Purification and characterization of a Neu5Ac α2-6Galβ1-4Glc/GlcNAc-specific lectin from the fruiting body of the polypore mushroom *Polyporus squamosus*. J. Biol. Chem. **275**, 10623–10629 (2000)
34. Kadirvelraj, R., Grant, O.C., Goldstein, I.J., Winter, H.C., Tateno, H., Fadda, E., Woods, R.J.: Structure and binding analysis of *Polyporus squamosus* lectin in complex with the Neu5Acα2-6 Galβ1-4GlcNAc human-type influenza receptor. Glycobiology **21**, 973–984 (2011)
35. Yabe, R., Suzuki, R., Kuno, A., Fujimoto, Z., Jigami, Y., Hirabayashi, J.: Tailoring a novel sialic acid-binding lectin from a ricin-B chain-like galactose-binding protein by natural evolution-mimicry. J. Biochem. **141**, 389–399 (2007)
36. Wang, Y., Yu, G., Han, Z., Yang, B., Hu, Y., Zhao, X., Wu, J., Lv, Y., Chai, W.: Specificities of *Ricinus communis* agglutinin 120 interaction with sulfated galactose. FEBS Lett. **585**, 3927–3934 (2011)
37. Hu, D., Tateno, H., Kuno, A., Yabe, R., Hirabayashi, J.: Directed evolution of lectins with sugar-binding specificity for 6-sulfogalactose. J. Biol. Chem. **287**, 20313–20320 (2012)
38. Chen, Y., Rougé, P., Peumans, W.J., Van Damme, E.J.M.: Mutational analysis of carbohydrate-binding activity of the NeuAc(α-2,6)Gal/GalNAc-specific type 2 ribosome-inactivating protein from elderberry (*Sambucus nigra*) fruits. Biochem. J. **364**, 587–592 (2002)
39. Van Damme, E.J.M., Peumans, W.J., Barre, A., Rougé, P.: Plant lectins: a composite of several distinct families of structurally and evolutionary related proteins with diverse biological roles. Crit. Rev. Plt. Sci. **17**, 575–692 (1998)
40. Ferreras, J.M., Citores, L., Iglesias, R., Jiménez, P., Girbés, T: *Sambucus* ribosome-inactivating proteins and lectins. (Lord, J.M., Hartley, M,R., Eds.), pp.122-123, Chapter 6. Toxic Plant Proteins, Plant Cell Monographs **18**, Springer-Verlag Berlin Heidelberg (2010)
41. Chen, Y., Peumans, W.J., Van Damme, E.J.M.: The *Sambucus nigra* type-2 ribosome-inactivating protein SNA-I' exhibits in planta antiviral activity in transgenic tobacco. FEBS J. **516**, 27–30 (2002)
42. Shahidi Noghabi, S., Van Damme, E.J.M., Smagghe, S.: Carbohydrate-binding activity of the type-2 ribosome-inactivating protein SNA-I from elderberry (*Sambucus nigra*) is a determining factor for its insecticidal activity. Phytochemistry **69**, 2972–2978 (2008)
43. Shahidi-Noghabi, S., Van Damme, E.J.M., Smagghe, G.: Expression of *Sambucus nigra* agglutinin (SNA-I') from elderberry bark in transgenic tobacco plants results in enhanced resistance to different insect species. Transgenic Res. **18**, 249–259 (2009)
44. Battelli, M., Citores, L., Buonamici, L., Ferreras, J.M., De Benito, F.M., Girbés, T., Stirpe, F.: Toxicity and cytotoxicity of nigrin b, a two-chain ribosome-inactivating protein from *Sambucus nigra* : comparison with ricin. Arch. Toxicol. **71**, 360–364 (1997)
45. Shahidi-Noghabi, S., Van Damme, E.J.M., De Vos, W.H., Smagghe, G.: Internalization of *Sambucus nigra* agglutinins I and II in insect midgut cells. Arch. Insect Biochem. Physiol. **76**, 211–222 (2011)
46. Shang, C., Madej, M., De Vos, W., Van Damme, E.J.M.: The ribosome-inactivating proteins (RIPs) from elderberry (*Sambucus nigra*) in the battle against cancer?? Glycoconj. J. **30**, 355 (2013)