



## Structural basis for sugar recognition, including the Tn carcinoma antigen, by the lectin SNA-II from *Sambucus nigra*

Laurent Maveyraud, Hideaki Niwa, Valérie Guillet, Dmitri I Svergun, Peter V Konarev, Rex A Palmer, Willy J Peumans, Pierre Rougé, Els J M van Damme, Colin D Reynolds, et al.

### ► To cite this version:

Laurent Maveyraud, Hideaki Niwa, Valérie Guillet, Dmitri I Svergun, Peter V Konarev, et al.. Structural basis for sugar recognition, including the Tn carcinoma antigen, by the lectin SNA-II from *Sambucus nigra*. *Proteins - Structure, Function and Bioinformatics*, 2008, 75, pp.89 - 103. 10.1002/prot.22222 . hal-03003386

**HAL Id: hal-03003386**

**<https://hal.science/hal-03003386>**

Submitted on 20 Nov 2020

**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.

# Structural basis for sugar recognition, including the Tn carcinoma antigen, by the lectin SNA-II from *Sambucus nigra*

Laurent Maveyraud,<sup>1†\*</sup> Hideaki Niwa,<sup>2†</sup> Valérie Guillet,<sup>1</sup> Dmitri I. Svergun,<sup>3</sup> Peter V. Konarev,<sup>3</sup> Rex A. Palmer,<sup>2</sup> Willy J. Peumans,<sup>4</sup> Pierre Rougé,<sup>5</sup> Els J. M. Van Damme,<sup>4</sup> Colin D. Reynolds,<sup>6</sup> and Lionel Mourey<sup>1\*</sup>

<sup>1</sup> Institut de Pharmacologie et de Biologie Structurale (IPBS), UMR 5089, Université Paul Sabatier Toulouse III/CNRS, Toulouse, France

<sup>2</sup> School of Crystallography, Birkbeck College, University of London, London, United Kingdom

<sup>3</sup> European Molecular Biology Laboratory, Hamburg Outstation, EMBL c/o DESY, Hamburg, Germany; and Institute of Crystallography, Russian Academy of Sciences, Moscow, Russia

<sup>4</sup> Department of Molecular Biotechnology, Ghent University, Ghent, Belgium

<sup>5</sup> Surfaces Cellulaires et Signalisation chez les Végétaux, UMR-CNRS 5546, Pôle de Biotechnologie Végétale, Toulouse, France

<sup>6</sup> School of Biomolecular Sciences, Liverpool John Moores University, Liverpool, United Kingdom

## ABSTRACT

Bark of elderberry (*Sambucus nigra*) contains a galactose (Gal)/N-acetylgalactosamine (GalNAc)-specific lectin (SNA-II) corresponding to slightly truncated B-chains of a genuine Type-II ribosome-inactivating protein (Type-II RIPs, SNA-V), found in the same species. The three-dimensional X-ray structure of SNA-II has been determined in two distinct crystal forms, hexagonal and tetragonal, at 1.90 Å and 1.35 Å, respectively. In both crystal forms, the SNA-II molecule folds into two linked  $\beta$ -trefoil domains, with an overall conformation similar to that of the B-chains of ricin and other Type-II RIPs. Glycosylation is observed at four sites along the polypeptide chain, accounting for 14 saccharide units. The high-resolution structures of SNA-II in complex with Gal and five Gal-related saccharides (GalNAc, lactose,  $\alpha$ 1-methylgalactose, fucose, and the carcinoma-specific Tn antigen) were determined at 1.55 Å resolution or better. Binding is observed in two saccharide-binding sites for most of the sugars: a conserved aspartate residue interacts simultaneously with the O3 and O4 atoms of saccharides. In one of the binding sites, additional interactions with the protein involve the O6 atom. Analytical gel filtration, small angle X-ray scattering studies and crystal packing analysis indicate that, although some oligomeric species are present, the monomeric species predominate in solution.

Proteins 2009; 75:89–103.  
© 2008 Wiley-Liss, Inc.

**Key words:** lectin; glycoprotein; isolated ricin B-chain; sugar binding; galactose-binding protein; protein–ligand complexes.

## INTRODUCTION

Plants belonging to the genus *Sambucus* express a wide range of lectins, with various saccharide-binding specificities.<sup>1</sup> Many of these lectins, and this is especially the case in the bark, are Type-II ribosome-inactivating proteins (Type-II RIPs) or are related to them. RIPs are plant proteins that are able to specifically remove an adenine residue from a highly conserved loop in the large subunit of ribosomal RNA, thereby interrupting protein synthesis in target cells.<sup>2</sup> These proteins are distributed among Type-I (single chain) and Type-II (two chains) RIPs.<sup>3</sup> The latter, exemplified by the well-known toxin ricin, are built up of an A-subunit, responsible for the N-glycosidase activity, that is covalently attached to a lectinic B-subunit. The linkage of both subunits involves a disulphide bridge between cysteine residues at Position 4 of the B-subunit and at Position 259 of the A-subunit (ricin numbering).<sup>4</sup> The lectin subunit usually specifically recognizes galactose (Gal), N-acetylgalactosamine (GalNAc) or related sugars.<sup>1</sup> It anchors the toxin

<sup>†</sup>Laurent Maveyraud and Hideaki Niwa contributed equally to the work.

Hideaki Niwa's current address is Systems and Structural Biology Centre, RIKEN Yokohama Institute, Yokohama 230-0045, Japan.

Grant sponsor: Centre National de la Recherche Scientifique (UMR 5089); Grant sponsor: Scientific Research-Flanders and the European Community; Grant number: RII3/CT/2004/5060008 (Research Infrastructure Action under the FP6, Structuring the European Research Area Programme).

\*Correspondence to: Laurent Maveyraud or Lionel Mourey, Institut de Pharmacologie et de Biologie Structurale (IPBS), UMR 5089, Université Paul Sabatier Toulouse III/CNRS, 205 route de Narbonne, 31077 Toulouse Cedex, France.

E-mail: laurent.maveyraud@ipbs.fr or lionel.mourey@ipbs.fr

Received 9 April 2008; Revised 27 June 2008; Accepted 18 July 2008

Published online 17 September 2008 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/prot.22222

to the target cell membrane, facilitating its entry in the cell.<sup>5</sup> Type-I RIPs are made of the A-subunit only. The toxicity of RIPs toward animal cells suggests that they might have a role in plant defence.<sup>1</sup> However, as they are also abundantly found in storage organs, they may also be considered as storage proteins.<sup>6</sup>

Type-II RIPs are found in many different tissues in *Sambucus* species. Interestingly, in the bark of *S. nigra* and *S. sieboldiana*, lectins corresponding to the B-subunit of Type-II RIPs are also found.<sup>7,8</sup> In the bark of *S. nigra*, the lectin SNA-II corresponds to the B-subunit of the genuine Type-II RIP SNA-V found in the same tissue,<sup>9</sup> depleted of the eight N-terminal residues, including the cysteine residue involved in the association with the A-subunit.<sup>7</sup> SNA-II is, therefore, unable to associate with an A-subunit. A similar situation occurs in the bark of *S. sieboldiana*, where two lectins, SSA-b-3 and SSA-b-4, are related to the Type-II RIP sieboldin-b.<sup>8</sup> Earlier biochemical data suggested that SNA-II is a dimeric glycosylated protein, which preferentially recognizes Gal and GalNAc,<sup>7</sup> as well as Gal-related saccharides such as the carcinoma Tn antigen (P. Rougé, unpublished work).

The glycosylation pattern of cancer cells is known to differ from normal cells,<sup>10</sup> often resulting in the presence of accessible GalNAc residues at the surface.<sup>11</sup> The Tn epitope (Ser-O-GalNAc) is among the most specific human tumor-associated determinant. It is naturally present but remains buried in most normal cells, whereas it is uncovered at the cell surface in the vast majority of carcinomas,<sup>12</sup> and, hence, serves as a specific marker of the tumor development process.

In this study, we determined the structure of the native glycosylated SNA-II lectin purified from the bark of *S. nigra*, as well as the high-resolution structures of the complexes of SNA-II with Gal and five-related saccharides, including the Tn antigen. Furthermore, the solution quaternary structure of SNA-II was studied using analytical gel filtration and small angle X-ray scattering (SAXS).

## METHODS

### Production and purification

The SNA-II protein was purified directly from the bark of *S. nigra* as previously described.<sup>7</sup>

### Analytical methods

Analytical gel filtration of SNA-II was performed on a Pharmacia Superdex 75 HR column, using a running buffer made of 50 mM NaCl, 100 mM sodium acetate pH 5.0, and 100 mM Gal to prevent binding of SNA-II to the column. Molecular mass (MM) references were ribonuclease A (15.6 kDa), chymotrypsinogen A (19.4 kDa), ovalbumin (47.6 kDa), and albumin (62.9 kDa). Their elution volume was determined on separate runs.

### Small angle X-ray scattering

The X-ray scattering data were collected on the X33 camera<sup>13,14</sup> at the European Molecular Biology Laboratory (EMBL) on the storage ring DORIS III of the Deutsches Elektronen Synchrotron (DESY) using a gas-filled detector with delay line readout<sup>15</sup> and Mar345 Image Plate detector (Marresearch GmbH, Norderstedt, Germany). The scattering patterns from the solutions of SNA-II at protein concentrations 2, 5, 8, 13, 18, and 25 mg/mL (50 mM NaCl, 100 mM sodium acetate pH 5.0) were recorded at a sample-detector distance of 2.4 m covering the range of momentum transfer  $0.13 < s < 3.4 \text{ nm}^{-1}$  ( $s = 4\pi \sin(\theta)/\lambda$ , where  $2\theta$  is the scattering angle, and  $\lambda = 0.15 \text{ nm}$  is the X-ray wavelength). Similar data were collected with samples complemented with 100 mM Gal, in order to reproduce the conditions used in the gel filtration studies.

For the linear gas detector, the data collected in 15 successive 1-min frames were analyzed for the absence of radiation damage and the successive frames were averaged. For the Image Plate, the collection time was 5 min (no radiation damage was detected by comparison with the linear detector data), and the data were radially averaged by the program Mar-Primus.<sup>16</sup> All the data analysis steps were performed using the program PRIMUS.<sup>17</sup>

The forward scattering  $I(0)$  and the radius of gyration  $R_g$  were evaluated using the Guinier approximation<sup>18</sup> assuming that at very small angles ( $s < 1.3/R_g$ ) the intensity is represented as  $I(s) = I(0) \exp(-(sR_g)^2/3)$ . These parameters were also computed from the entire scattering pattern using GNOM.<sup>19,20</sup> The MM of the solute was evaluated by comparison of the forward scattering with that from a reference solution of bovine serum albumin (MM = 66 kDa).

The scattering from the crystallographic model of monomeric SNA-II and from the putative dimers and tetramer was calculated using CRY SOL.<sup>21</sup> The volume fractions in the monomer-dimer or monomer-tetramer mixtures best fitting the experimental data were evaluated using the computed scattering curves from monomeric, dimeric, and tetrameric SNA-II by OLIGOMER.<sup>17</sup>

### Crystallization and data collection

Crystals were obtained by the hanging drop vapour diffusion method at 20°C, using ammonium sulphate as precipitant. A 6  $\mu\text{L}$  droplet, made by mixing 3  $\mu\text{L}$  of a solution of SNA-II (16 mg mL<sup>-1</sup> in 20 mM Tris HCl buffer at pH 7.5, 20 mM NaCl) and 3  $\mu\text{L}$  of reservoir solution, was equilibrated against 500  $\mu\text{L}$  of 2.0 M ammonium sulphate in 100 mM sodium acetate buffer, pH 4.5–5.0. Under these conditions, two different crystal forms were obtained: hexagonal crystals appeared within 2 weeks, whereas tablet-shaped crystals grew more slowly. Both crystal forms needed to be stabilized by increasing

the ammonium sulphate concentration up to 3.0 M before they could be handled without damage. Cryoprotection was achieved by a 2 min immersion in the stabilizing solution complemented with 5% (v/v) ethylene glycol. Crystals could then withstand being transferred into a gaseous nitrogen stream at 100 K.

Crystals of SNA-II complexed to Gal and 5 Gal derivatives [GalNAc, lactose (Lac),  $\alpha$ 1-methylgalactose (MeGal), fucose (Fuc), and the Tn antigen] were obtained by adding 5 mM of the selected saccharide to the protein solution before setting up the crystallization experiment. In these conditions, only tablet-shaped crystals were obtained. Crystals were cryoprotected by a brief immersion in the crystallization condition supplemented with 20% (w/v) of the appropriate sugar, except for crystals obtained with the Tn antigen for which 5% ethylene glycol was used.

All diffraction data were collected at the European Synchrotron Radiation Facility (ESRF, Grenoble), processed using MOSFLM,<sup>22</sup> scaled with SCALA,<sup>23</sup> and, subsequently, handled with the CCP4 package.<sup>24</sup> Data collection and processing statistics are summarized in Table I.

Data collected up to 1.9 Å with the hexagonal crystals displayed systematic absences along 00l, which together with the results of the self rotation function, indicated that they belong either to spacegroup  $P6_222$  or  $P6_422$ , with cell parameters  $a = 120.2$  Å and  $c = 177.3$  Å. Tablet-shaped native crystals diffracted at least to 1.3 Å. Merging statistics and systematic absences suggested that they belong to the spacegroup  $I4_122$  with cell parameters  $a = 126.1$  Å and  $c = 76.0$  Å. Crystals obtained in the presence of saccharides belong to the same spacegroup and display similar cell parameters (Table I). There is one molecule of SNA-II in the asymmetric unit of both crystal forms.

### Structure determination and refinement

The structure of the hexagonal crystal form was solved by molecular replacement, using the structure of the B-chain of ricin (PDB entry 2AAI,<sup>25</sup>) as the search model. The molecular replacement solution found with MOLREP,<sup>26</sup> in spacegroup  $P6_422$ , was first refined using CNS,<sup>27</sup> and, when the refined model was considered sufficiently reliable ( $R_{\text{cryst}} = 0.291$ ,  $R_{\text{free}} = 0.300$ ), it was used as a search model for the molecular replacement of the tetragonal crystal form, with MOLREP. From that point, REFMAC<sup>28</sup> was used to refine all structures, using an explicit bulk solvent correction, inclusion of hydrogen atoms in their riding positions and the conjugate gradient method to minimize the maximum likelihood target function. Models and sigmaA weighted electron density maps<sup>29</sup> were inspected using COOT.<sup>30</sup> Water molecules were automatically introduced into the model using WARP.<sup>31</sup> In the last refinement cycles, individual anisotropic B factors were refined for all structures in the tet-

ragonal spacegroup, which resulted in a similar improvement of the  $R_{\text{cryst}}$  and  $R_{\text{free}}$  values.

The final structure of native SNA-II in the hexagonal crystal form includes 257 protein residues, 14 saccharide units, 342 water molecules, 1 acetate, and 7 sulphate anions. In the tetragonal crystal form of native SNA-II, the final model comprises 257 protein residues, 9 saccharide units, 253 water molecules, 1 acetate, and 4 sulphate anions. Refinement statistics for all structures are summarized in Table I.

Secondary structure elements were assigned using STRIDE.<sup>32</sup> The structures were superimposed using Dali-Lite.<sup>33</sup>

The atomic coordinates of refined structures and the experimental data have been deposited with the Protein Data Bank.<sup>34</sup> The accession numbers for the atomic coordinates and experimental data are indicated in Table I.

### Protein sequence analysis

Sequences were aligned using CLUSTALW.<sup>35</sup> The alignment was manually optimized and displayed with ESPrpt.<sup>36</sup>

## RESULTS AND DISCUSSION

### Overall structure

The structure of SNA-II was determined in two crystal forms. Both are very similar, with a root mean square deviation (rmsd) of 0.55 Å for the 257 C $\alpha$  atoms. It is organized in two similar domains [Fig. 1(a)], Domain I (Residues 1–129) and Domain II (Residues 130–257), each adopting a  $\beta$ -trefoil fold<sup>37,38</sup> [Fig. 1(b)]. In this fold, a central  $\beta$ -barrel built from six antiparallel  $\beta$ -strands delineates a hydrophobic core. At the bottom of the  $\beta$ -barrel, short connecting loops are found, not exceeding four residues in length, whereas at the top, three much longer loops are found, each including two  $\beta$ -strands arranged as a  $\beta$ -hairpin and an occasional short helical segment. The fold displays a pseudo three-fold symmetry when viewed along the axis of the  $\beta$ -barrel, which results from the organization of three similar subdomains of about 40 residues, named  $\alpha$ ,  $\beta$ , and  $\gamma$  [Fig. 1(b)]. Each subdomain contains 4  $\beta$ -strands, the first and fourth contributing to the central  $\beta$ -barrel. The average sequence identity between these subdomains is 18%, which confirms that they have evolved from a common ancestral gene [Fig. 2(a)].<sup>39,40</sup>

The  $\beta$ -trefoil fold is widely found in protein structures, either associated to a different fold as in the inositol-triphosphate receptor Type-1 binding core,<sup>41</sup> in clostridium neurotoxins,<sup>42</sup> and in the pore-forming lectin from *Laetiporus sulphureus*,<sup>43</sup> or isolated as in the fibroblast growth factors<sup>44</sup> and interleukin-1.<sup>45</sup> In most proteins, a single occurrence of the  $\beta$ -trefoil fold is found.

**Table 1**  
Data Collection and Crystallographic Refinement Statistics<sup>a</sup>

	Native		Galactose		GalNAc		Lactose		Me galactose		Fucose		Tn antigen	
Beamline	ID14-EH1	ID14-EH1	BM30A	BM30A	BM30A	BM30A	BM30A	BM30A	BM30A	BM30A	BM30A	BM30A	ID14-EH2	ID14-EH2
Wavelength (Å)	0.937	0.937	0.937	0.937	0.937	0.937	0.937	0.937	0.937	0.937	0.937	0.937	0.933	0.933
Spacegroup	P6 <sub>4</sub> 22	14,22	14,22	14,22	14,22	14,22	14,22	14,22	14,22	14,22	14,22	14,22	14,22	14,22
Cell parameters (Å)	120.20 120.20 177.34	126.12 126.12 76.04	126.31 126.31 75.41	126.49 126.49 75.41	126.75 126.75 75.50	126.75 126.75 75.50	126.75 126.75 75.50	126.75 126.75 75.50	126.75 126.75 75.50	126.75 126.75 75.50	126.75 126.75 75.50	126.75 126.75 75.50	127.04 127.04 75.49	127.04 127.04 75.49
Resolution limits (Å)	51.30–1.90 (2.00–1.90)	64.55–1.35 (1.39–1.35)	34.74–1.55 (1.64–1.55)	34.74–1.55 (1.63–1.55)	34.79–1.54 (1.63–1.55)	34.79–1.54 (1.63–1.55)	34.79–1.54 (1.63–1.55)	34.79–1.54 (1.63–1.55)	34.78–1.55 (1.64–1.55)	34.68–1.55 (1.64–1.55)	34.68–1.55 (1.64–1.55)	34.68–1.55 (1.64–1.55)	45.41–1.40 (1.44–1.40)	45.41–1.40 (1.44–1.40)
No. of observations	287,126 (12,647)	305,842 (16,291)	188,533 (13,275)	193,325 (13,178)	198,082 (14,080)	198,082 (14,080)	198,082 (14,080)	198,082 (14,080)	201,518 (15,040)	248,493 (18,781)	248,493 (18,781)	248,493 (18,781)	378,099 (18,366)	378,099 (18,366)
No. unique reflections	57,308 (5,977)	66,690 (4,901)	42,644 (4,945)	43,195 (5,240)	43,092 (5,002)	43,092 (5,002)	43,092 (5,002)	43,092 (5,002)	42,458 (4,882)	43,501 (5,890)	43,501 (5,890)	43,501 (5,890)	60,175 (4,332)	60,175 (4,332)
Multiplicity	5.0 (2.1)	4.6 (3.3)	4.4 (2.7)	4.5 (2.5)	4.6 (2.8)	4.6 (2.8)	4.6 (2.8)	4.6 (2.8)	4.7 (3.1)	5.7 (3.2)	5.7 (3.2)	5.7 (3.2)	6.3 (4.2)	6.3 (4.2)
Completeness (%)	95.5 (70.0)	99.6 (99.7)	96.6 (78.0)	97.4 (82.6)	96.7 (78.4)	96.7 (78.4)	96.7 (78.4)	96.7 (78.4)	96.9 (77.2)	98.8 (92.9)	98.8 (92.9)	98.8 (92.9)	99.5 (98.4)	99.5 (98.4)
R <sub>sym</sub> <sup>b</sup>	0.083 (0.324)	0.075 (0.395)	0.052 (0.135)	0.051 (0.166)	0.041 (0.142)	0.041 (0.142)	0.041 (0.142)	0.041 (0.142)	0.058 (0.216)	0.044 (0.095)	0.044 (0.095)	0.044 (0.095)	0.069 (0.180)	0.069 (0.180)
Average I/σ(I)	13.7 (2.1)	14.0 (2.3)	21.5 (7.1)	21.6 (5.5)	23.0 (7.6)	23.0 (7.6)	23.0 (7.6)	23.0 (7.6)	17.2 (4.3)	32.4 (11.5)	32.4 (11.5)	32.4 (11.5)	17.6 (6.4)	17.6 (6.4)
No. of protein atoms	2027	2101	2153	2100	2091	2091	2091	2091	2091	2079	2079	2079	2128	2128
Nonprot. atoms	551	400	543	553	557	557	557	557	524	521	521	521	514	514
Res range (Å)	20.00–1.95	20.00–1.35	20.00–1.55	20.00–1.55	20.00–1.55	20.00–1.55	20.00–1.55	20.00–1.55	20.00–1.55	20.00–1.55	20.00–1.55	20.00–1.55	20.00–1.40	20.00–1.40
R <sub>cryst</sub> <sup>c</sup>	0.173 (52,296)	0.161 (63,280)	0.119 (40,471)	0.125 (40,982)	0.133 (40,807)	0.133 (40,807)	0.133 (40,807)	0.133 (40,807)	0.135 (40,281)	0.125 (41,276)	0.125 (41,276)	0.125 (41,276)	0.143 (57,048)	0.143 (57,048)
R <sub>free</sub> <sup>d</sup>	0.194 (2,762)	0.199 (3,366)	0.163 (2,131)	0.173 (2,179)	0.179 (2,155)	0.179 (2,155)	0.179 (2,155)	0.179 (2,155)	0.180 (2,125)	0.167 (2,186)	0.167 (2,186)	0.167 (2,186)	0.187 (3,038)	0.187 (3,038)
Crickshand DPI (Å)	0.085	0.054	0.077	0.076	0.082	0.082	0.082	0.082	0.084	0.074	0.074	0.074	0.057	0.057
Average B (Å <sup>2</sup> )	41.7	24.1	14.2	14.3	17.4	17.4	17.4	17.4	25.2	12.7	12.7	12.7	13.8	13.8
Wilson B (Å <sup>2</sup> )	29.5	16.2	15.5	15.6	15.3	15.3	15.3	15.3	17.4	15.4	15.4	15.4	15.2	15.2
Protein B (Å <sup>2</sup> )	54.2	21.1	16.6	17.0	14.0	14.0	14.0	14.0	21.8	14.5	14.5	14.5	16.1	16.1
Solvent B (Å <sup>2</sup> )	58.8	55.8	32.4	32.9	29.0	29.0	29.0	29.0	47.2	30.2	30.2	30.2	30.0	30.0
Glycan B (Å <sup>2</sup> )	71.0	34.1	43.7	37.8	38.2	38.2	38.2	38.2	37.4	39.1	39.1	39.1	44.8	44.8
Bound sugar B (Å <sup>2</sup> )	—	—	15.7	14.6	21.2	21.2	21.2	21.2	23.4	27.0	27.0	27.0	18.8	18.8
Rmsd bond lengths (Å)	0.021	0.021	0.020	0.021	0.019	0.019	0.019	0.019	0.021	0.021	0.021	0.021	0.021	0.021
Rmsd bond angles (°)	1.907	1.917	1.925	1.857	1.781	1.781	1.781	1.781	1.825	2.026	2.026	2.026	1.922	1.922
PDB code	3CA0	3C9Z	3CA1	3CA3	3CA4	3CA4	3CA4	3CA4	3CA5	3CAH	3CAH	3CAH	3CA6	3CA6

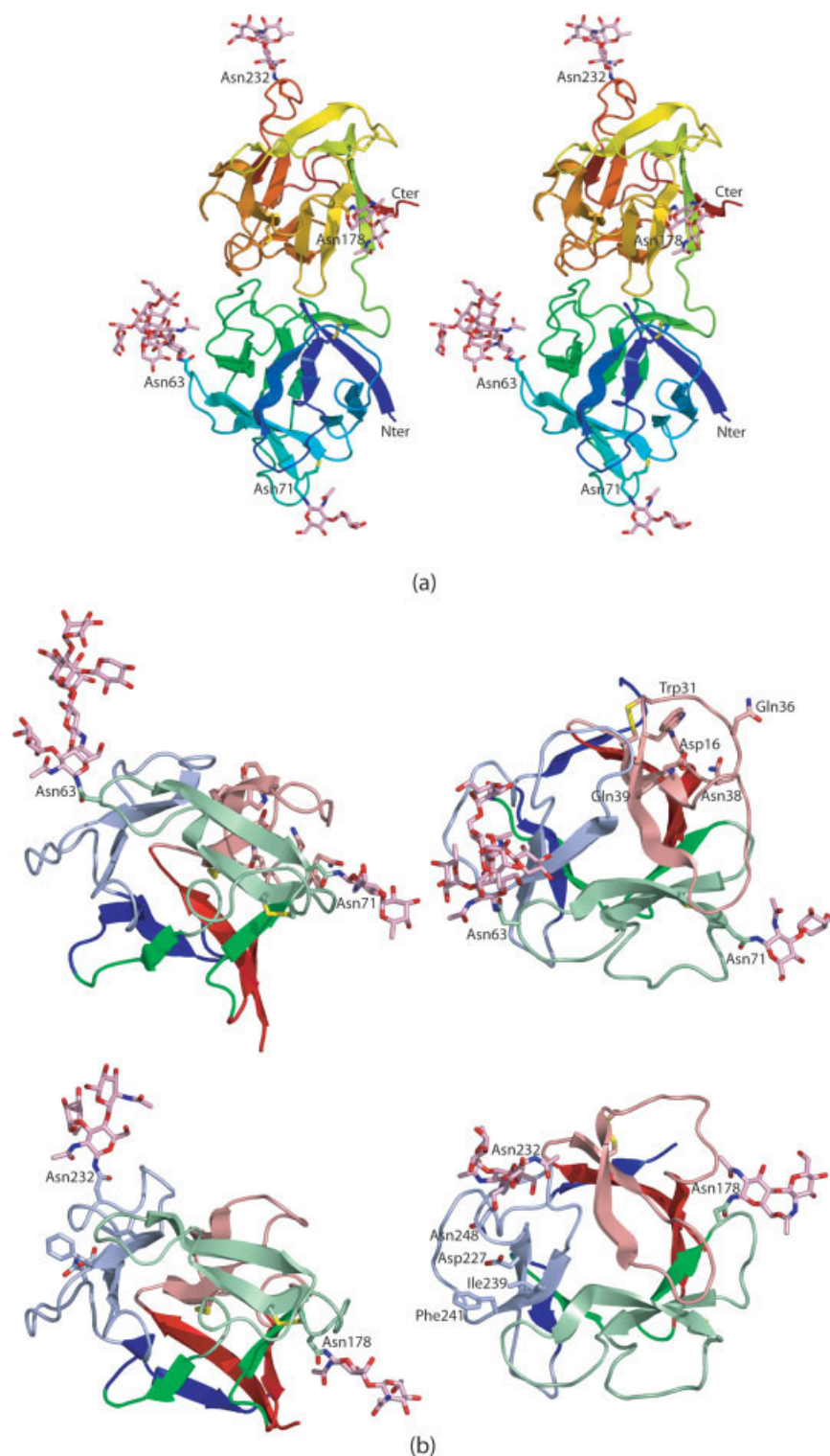
<sup>a</sup>Numbers in parentheses concern the highest resolution shell.

<sup>b</sup> $R_{\text{sym}} = \sum I_{\text{hkl}} - \langle I_{\text{hkl}} \rangle / \sum \langle I_{\text{hkl}} \rangle$ , where  $I_{\text{hkl}}$  is the scaled intensity of the hkl reflection, and  $\langle I_{\text{hkl}} \rangle$  is the mean value of the hkl reflections.

<sup>c</sup> $R_{\text{cryst}} = \sum |F_{\text{obs}} - F_{\text{calc}}| / \sum F_{\text{obs}}$ , where  $F_{\text{obs}}$  and  $F_{\text{calc}}$  are the observed and calculated structure factor amplitudes of the hkl reflection. The number of reflection used for the calculation of  $R_{\text{cryst}}$  is indicated in parenthesis.

<sup>d</sup> $R_{\text{free}}$  is calculated in the same manner as  $R_{\text{cryst}}$  but using the test set of reflections. The size of the free set is indicated in parenthesis.



**Figure 1**

Ribbon representation of the structure of SNA-II. (a) Stereoview of the overall structure of SNA-II colored from blue (N terminus) to red (C terminus) as observed in the hexagonal crystal form. The four glycosylation sites are labeled, and the glycans are represented as pink sticks. Disulphide bridges are represented as sticks. (b) View of Domains I (top) and II (bottom) in two orthogonal orientations. In each domain, subdomains  $\alpha$  are colored in red,  $\beta$  in green, and  $\gamma$  in blue. The first and fourth  $\beta$ -strands of each subdomain forming the central  $\beta$ -barrel are indicated with vivid colors. Glycosylation sites and saccharide-binding residues are labeled and represented as sticks. Glycans are represented as pink sticks.

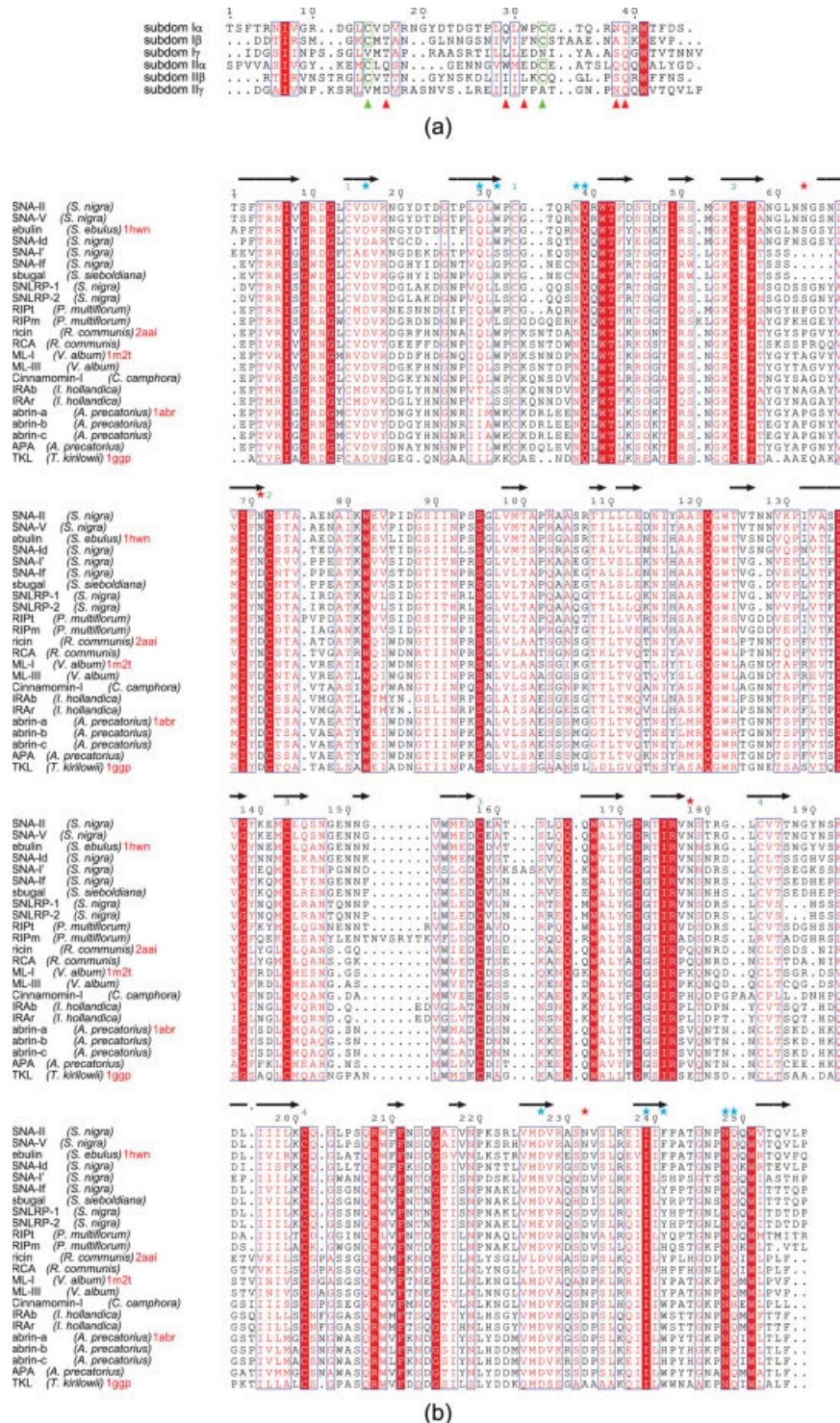
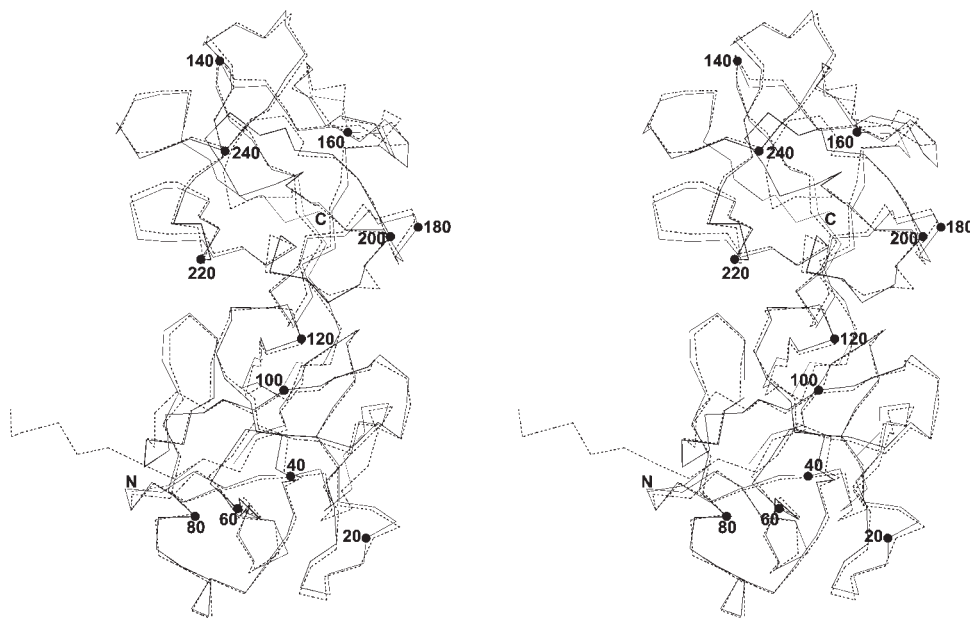


Figure 2

Sequence analysis of SNA-II and of related proteins. (a) Sequence alignment of the six subdomains of SNA-II. Cysteine residues involved in disulphide bridges are indicated with a green triangle, and saccharide-binding residues are indicated with a red triangle. Strictly conserved residues are on a red background, and conserved positions are indicated with red letters. (b) Sequence alignment of the B-chains of Type-II RIPs. The secondary structure elements identified in SNA-II are indicated above the sequences ( $\beta$ -strands are depicted with an arrow). Saccharide-binding residues and glycosylation sites are identified with a blue and a red star, respectively. Disulphide bridges present in SNA-II are numbered in green. Strictly conserved residues are on a red background, and conserved positions are indicated with red letters. The PDB code is indicated with the sequence name for available structures.



**Figure 3**

Structural comparison of SNA-II and of the ricin B-chain Stereoview of the C $\alpha$  trace of SNA-II (solid lines) superimposed on that of the ricin B-chain (dotted lines). Every 20th C $\alpha$  is indicated with a black dot and labeled.

However, a few examples exist where such folds are duplicated, as in plant cytotoxin B-chains, or even quadruplicated, as observed in the human actin-crosslinking protein fascin.<sup>46</sup> However, although the  $\beta$ -trefoil folds are similar when considered individually (with an average rmsd of 2.1 Å for 114 common C $\alpha$  when 57  $\beta$ -trefoil domain-containing proteins, as identified by SCOP,<sup>47</sup> are considered), the organization of the double or quadruple repeats are markedly different from what is observed in SNA-II. A global similarity is only evidenced with the plant cytotoxin B-chains. These include the B-chains of ricin (2AAI, rmsd of 0.8 Å for 244 equivalent C $\alpha$  atoms when compared with SNA-II) (see Fig. 3),<sup>25,48</sup> abrin (1ABR, rmsd of 0.8 Å for 249 equivalent C $\alpha$  atoms),<sup>49</sup> *Abrus precatorius* agglutinin (2Q3N, rmsd of 0.8 Å for 253 equivalent C $\alpha$  atoms),<sup>50</sup> ebulin (1HWM, rmsd of 0.7 Å for 246 equivalent C $\alpha$  atoms),<sup>51</sup> Mistletoe lectin ML-I (1M2T, rmsd of 0.8 Å for 249 equivalent C $\alpha$  atoms),<sup>52</sup> and the Type-II RIP from *Trichosanthes kirilowii* (1GGP, rmsd of 1.0 Å for 235 equivalent C $\alpha$  atoms).<sup>53</sup> B-chains of Type-II RIPs from other *Trichosanthes* species also display a similar fold.<sup>54,55</sup> Hence, this specific arrangement of two  $\beta$ -trefoil domains can be considered a characteristic of the B-chains of Type-II RIPs and of related proteins, such as SNA-II.

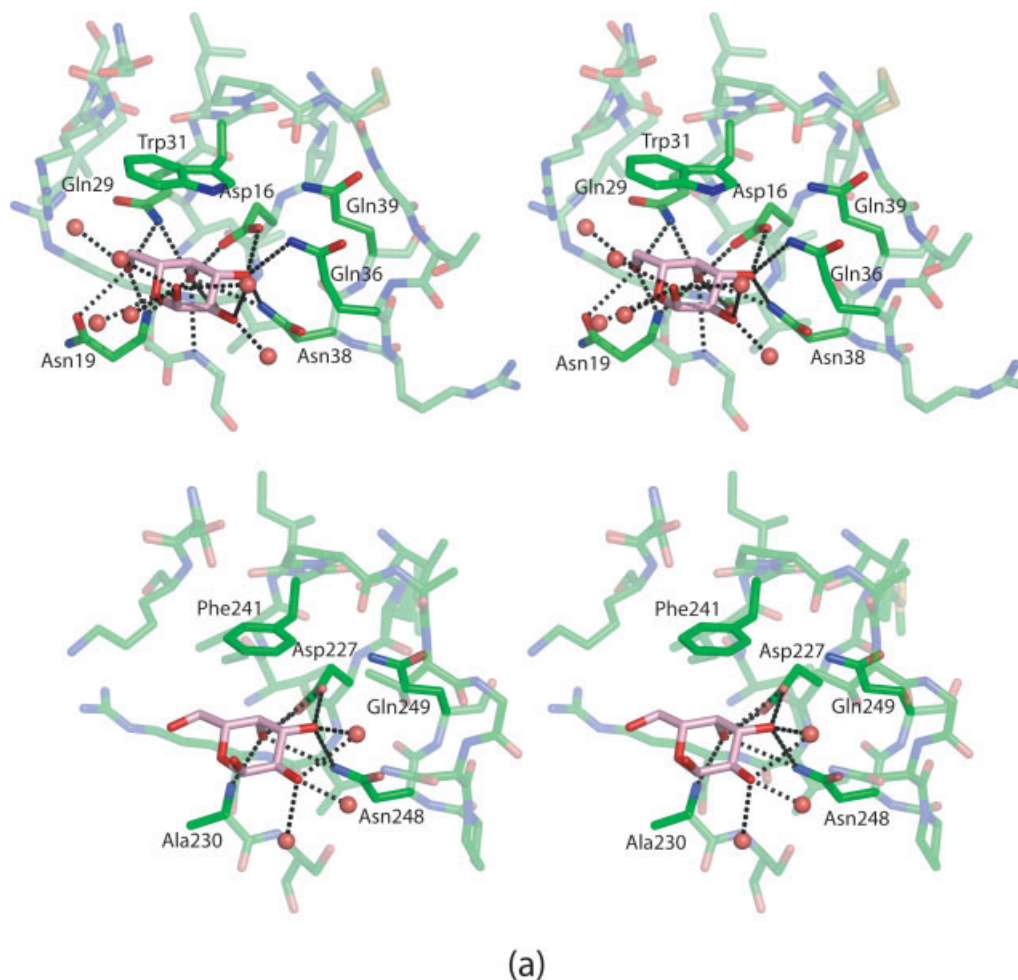
Disulphide bridges are found in subdomains I $\alpha$  (Cys14-Cys33), I $\beta$  (Cys55-Cys72), II $\alpha$  (Cys143-Cys158), and II $\beta$  (Cys184-Cys201). These disulphide bridges occur

at equivalent positions in the four subdomains [Fig. 2(a)] and link the N-terminal extremity of the second  $\beta$ -strand to the C terminus of the third  $\beta$ -strand of each subdomain. The Cys14-Cys33 disulphide bridge is absent from some Type-II RIPs found in *V. album* species, whereas the Cys184-Cys201 disulphide bridge is missing in the B-chain of RIPm, the Type-II RIP from *Polygonatum multiflorum*.<sup>56</sup> Finally, Cys55-Cys72 and Cys143-Cys158 are probably the only conserved disulphide bridges among all B-chains of Type-II RIPs [Fig. 2(b)].

### Glycosylation and structure of the N-glycans

Previous biochemical studies indicated that SNA-II is a glycoprotein containing 7.8% (w/w) neutral sugars,<sup>7</sup> which correspond to  $\sim 14$  monosaccharide units. In addition, there are four putative N-glycosylation sites (Asn63-Gly-Ser, Asn71-Cys-Ser, Asn178-Ser-Thr, and Asn232-Val-Ser) in the sequence of the mature SNA-II [Fig. 2(b)]. Glycosylation could clearly be distinguished at the four sites in both crystal forms. In the tetragonal crystal form, nine monosaccharide units were fitted into the electron density map. In the hexagonal crystal form, the N-glycan linked to Asn63 protrudes from the protein and reaches the sugar-binding site of a symmetry-related molecule (see later). The decreased flexibility resulting from this interaction induces a well-defined electron density. According to assignments of the sugars in the elec-



**Figure 4**

X-ray analysis of saccharide binding to SNA-II. (a) Stereoview of the saccharide-binding sites of Domain I (top) and of Domain II (bottom), with bound galactose (pink). Saccharide-binding residues are labeled, and water molecules are indicated with a red sphere. Hydrogen bonds are represented with dotted lines. (b) Comparison of the binding of the different saccharides used in this study (pink), except galactose, in the saccharide-binding site of Domain I. The orientation is identical as in (a).

tron density map, the N-glycan attached to Asn63 comprises three mannoses (Man), 2 *N*-acetylglucosamines (GlcNAc), one Fuc, and one xylose (Xyl) and is similar in structure to the oligosaccharide chains of other plant glycoproteins (including the B-chain of ricin)<sup>48</sup>:

Man( $\alpha$ 1,6)

\

Man( $\alpha$ 1,3)—Man( $\beta$ 1,4)—D-GlcNAc( $\beta$ 1,4)—D-GlcNAc( $\beta$ 1)—Asn63

/

Xyl( $\beta$ 1,2)

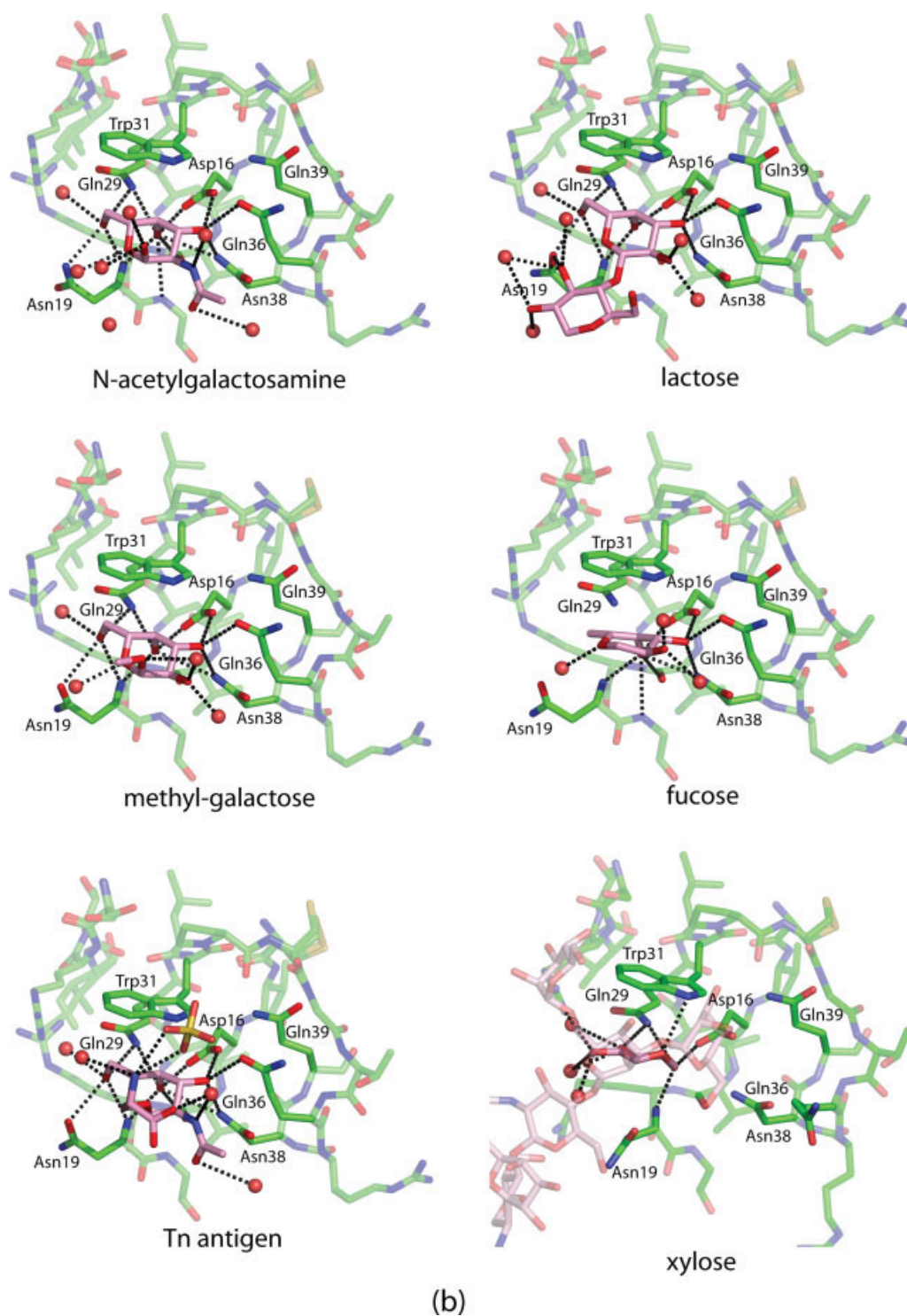
L-Fuc( $\alpha$ 1,3)

The assigned oligosaccharides linked to the three other glycosylation sites are much smaller. An L-Fuc( $\alpha$ 1,3)-D-GlcNAc disaccharide could be distinguished at Asn71, whereas a D-GlcNAc( $\beta$ 1,4)-D-GlcNAc disaccharide and a

branched D-GlcNAc( $\beta$ 1,4(L-Fuc $\alpha$ 1,3))D-GlcNAc trisaccharidic chain were found to be attached to Asn178 and Asn232, respectively. Thus, a total number of 14 sugar units with a total mass of  $\sim$ 1760 Da were identified and modeled into the structure of SNA-II, which may account for all the neutral sugars reported.<sup>7</sup>

### Saccharide-binding sites

It is generally accepted that the molecular organization of the ricin B-domain with its typical  $\beta$ -trefoil fold results from the triplication of an ancestral gene encoding a polypeptide chain of  $\sim$ 40 amino acid residues.<sup>39</sup> As this ancestral module contained a single active carbohydrate-binding site, triplication can yield a  $\beta$ -trefoil fold with up to three carbohydrate-binding sites, as it is illustrated by the lectin domain of the *Streptomyces*

**Figure 4**

(Continued)

$\beta$ -xylanases.<sup>57,58</sup> However, it appears that after the triplication event, some of the modules have lost their carbohydrate-binding capacity. Hence, most of the  $\beta$ -trefoil

folds contain a single active site, as has been demonstrated for the carbohydrate-binding domain of the mouse macrophage mannose receptor.<sup>59</sup> In some cases,

**Table II**  
Sugar–Protein Interactions Observed in Each Saccharide–Binding Sites

interactions (Å)	Galactose ( $\alpha/\beta$ ) <sup>a</sup>	GalNAc ( $\alpha/\beta$ ) <sup>a</sup>	Lactose	Me Galactose	Fucose	Tn antigen
Domain I						
O3–Gln36NE2	3.1/2.7	3.1/2.8	2.9	3.0	2.9	2.9
O3–Asn38 ND2	2.6/3.2	2.5/3.2	2.8	2.9	3.0	3.0
O3–Asp16OD1	2.7/2.6	2.6/2.6	2.6	2.6	2.6	2.6
O4–Asp16OD1	2.6/2.7	2.6/2.8	2.7	2.6	2.6	2.6
O4–Gln29NE2	3.3/3.1	3.3/3.1	3.2	3.2	3.6	3.2
O4–Asn19N	2.8/3.1	2.8/3.1	2.8	3.0	2.9	2.9
O6–Gln29NE2	3.0/3.0	3.0/3.0	3.0	3.0	C6 3.0	3.0
O6–Asn19N	2.5/3.3	2.7/3.4	3.2	3.1	C6 3.2	3.0
O6–Asn19OD1	3.0/3.3	2.9/3.3	2.7	3.3	C6 3.8	3.1
O6–Wat	3.0/2.6	2.9/2.6	2.8	2.7	C6 2.6	2.8
GlcO3–Asn19OD1 2.8						
Domain II						
O3–Asn248ND2	3.0/2.7	2.9/3.2	2.8	3.0		
O3–Asp227OD2	2.5/2.5	2.5/2.7	2.6	2.6		
O3–Wat	2.2/2.9	2.9/2.5	2.6	2.6		
O4–Asp227OD1	2.6/2.5	2.7/2.6	2.6	2.6		
O4–Ala230N	3.0/2.8	2.7/3.0	2.8	2.8		
O7–Ser231OG 3.0/2.5						

<sup>a</sup>The two distances correspond to epimers  $\alpha$  and  $\beta$ , respectively.

however, there is still some uncertainty about the number of active carbohydrate-binding sites. This is illustrated by the B-chain of ricin: although there is no doubt that the sites located within subdomains I $\alpha$  and II $\gamma$  definitely possess carbohydrate-binding activity, the occurrence of a third active site in subdomain I $\beta$  is still debated.<sup>60–62</sup> It is also noteworthy that a marked divergence has occurred in the specificity of the carbohydrate-binding modules. Some of ricin-like B-domains found in bacteria bind Xyl, whereas their animal counterparts recognize mannose, sulphated carbohydrate ligands,<sup>59</sup> or Gal and GalNAc. Most plant ricin-like B-domains bind Gal or GalNAc, as illustrated by the sugar-binding specificity of the majority of Type-II RIPs.<sup>1</sup>

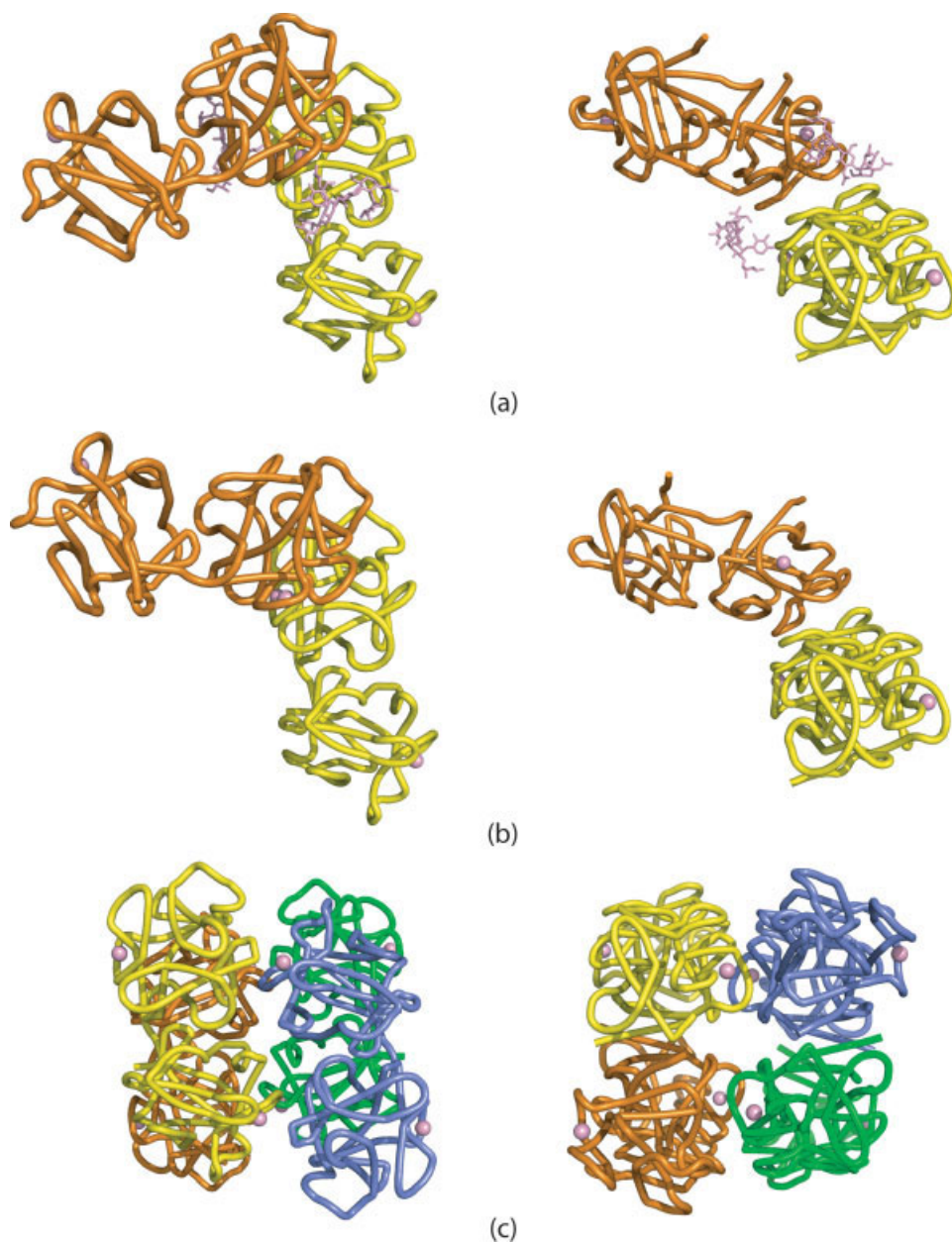
Superimposition of the structures revealed that the two canonical saccharide-binding sites of the ricin B-chain are conserved in SNA-II. These sites are located in subdomains I $\alpha$  and II $\gamma$  on opposite faces at both ends of SNA-II (see Fig. 1). Each site is built around a small set of conserved residues. In subdomain I $\alpha$ , these residues are Asp16, Gln29, Trp31, Gln36, Asn38, and Gln39, whereas in subdomain II $\gamma$  they correspond to Asp227, Ile239, Phe241, and Asn248 [Fig. 4(a)]. These residues occupy equivalent positions both in the sequence and in terms of spatial arrangement [Figs. 2(a) and 4(a)].

### High-resolution X-ray analysis of saccharide binding

SNA-II was reported to bind Gal and related saccharides with about 100 to 2500-fold better affinity than for glucose, with a marked preference for GalNAc versus Fuc, Gal, and MeGal.<sup>7</sup> The structures of SNA-II bound

to Gal and to various Gal derivatives were determined at 1.55 Å resolution. Furthermore, binding of the Tn carcinoma general autoantigen (Ser-O-GalNAc) was characterized at 1.40 Å resolution (Table I).

Hydrogen bonds (H-bonds) discussed in the text are identified upon the observed geometry of the structure. Binding of Gal in both saccharide-binding sites relies on the H-bonds formed with the hydroxyl groups at Positions 3 and 4 of the sugar ring [Table II, Fig. 4(a)]. In the saccharide-binding site of subdomain I $\alpha$ , the O3 atom of Gal interacts with the carboxylic group of Asp16, and with the side-chain amide group of both Gln36 and Asn38. The O4 atom is H-bonded to the carboxylic group of Asp16, to the side-chain amide group of Gln29 and to the main-chain nitrogen atom of Asn19. Similar interactions occur in the saccharide-binding site of domain II $\gamma$ , where the O3 atom is H-bonded to the carboxylic group of Asp227, to the side-chain amide group of Asn248 and to a water molecule, whereas the O4 atom interacts with Asp227 and with the main-chain nitrogen atom of Ala230. In the saccharide-binding site of subdomain I $\alpha$ , additional interactions involve the O6 atom of the sugar and both the main-chain nitrogen atom and the amide function of Asn19, the side-chain amide group of Gln29 and a water molecule. The side-chain of Asn19 is significantly displaced from its position in the saccharide free structure in order to interact with the incoming sugar. In both binding sites, an aromatic residue is found at van der Waals distance from the sugar ring to the C6 atom (Trp31 and Phe241). The O1 atom of the sugar does not contribute to binding, and both anomers could be identified in the electron density maps at both binding sites.



**Figure 5**

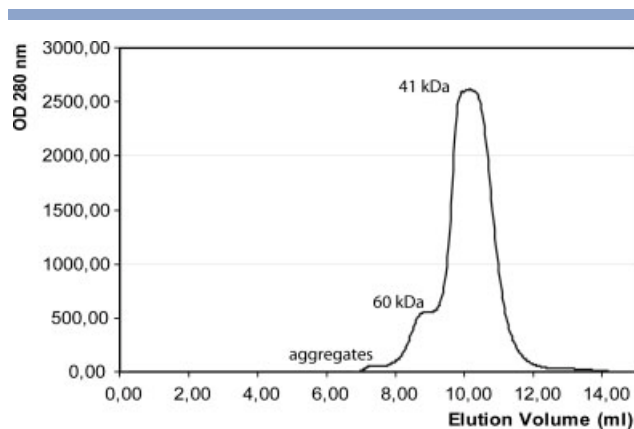
Orthogonal view of the dimeric and tetrameric assemblies observed in the crystals formed by SNA-II. (a) Dimer observed in the hexagonal crystal form. The glycan chains participating in the dimer formation (attached to Asn63) are indicated in pink. (b) Dimer observed in the tetragonal crystal form. (c) Tetrameric assembly identified in both crystal forms. Chain A, colored yellow, is represented in the same orientation in each case. The positions of the saccharide-binding sites are indicated with violet spheres.

Five Gal derivatives were also studied: GalNAc, Lac, MeGal, Fuc, and the Tn antigen. In the case of the two latter saccharides, the initial electron density maps clearly indicated that binding did only occur in the saccharide-binding site of subdomain I $\alpha$ , an acetate ion being found in the saccharide-binding site of subdomain II $\gamma$ . All studied Gal derivatives display the common core of interac-

tions with SNA-II found in the case of Gal [Table II, Fig. 4(b)], with additional direct interactions in the case of GalNAc in the saccharide-binding site of subdomain II $\gamma$  (Ser231OG-O7) and in the case of Lac in the saccharide-binding site of subdomain I $\alpha$  (Asn19OD1-GlcO3).

The Tn antigen is only bound in the saccharide-binding site of the first  $\beta$ -trefoil domain. The GalNAc part of





**Figure 6**

Analytical gel filtration elution pattern. The concentrated SNA-II solution was loaded on a Pharmacia Superdex 75 HR column, in presence of 100 mM galactose. The indicated apparent molecular weights were determined from the calibration curve (see text).

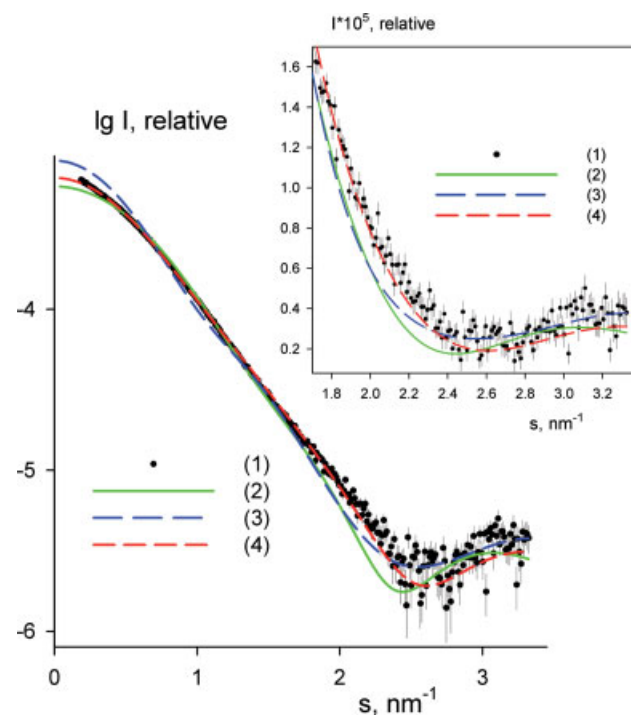
the antigen displays identical interactions as observed for the binding of GalNAc [Fig. 4(b), Table II]. The serine residue does not directly interact with the protein, and the electron density suggests that it is rather mobile. The main-chain nitrogen atom of the serine residue interacts with a sulphate anion, which itself is H-bonded to the indole nitrogen atom of Trp131. Additionally, a water molecule bridges the nitrogen atom of the serine to the guanidinium group of Arg107. Solvent-mediated interactions were also observed in other structures of lectin complexed to the Tn antigen.<sup>63,64</sup>

In the hexagonal native crystal form, the saccharide-binding site of subdomain I $\alpha$  is occupied by a  $\beta$ -D-xylopyranose unit of the glycan chain attached to Asn63 of a symmetry-related molecule [Fig. 4(b)].<sup>65</sup> H-bonds are formed between the O2 atom of Xyl and the NH1 atom of Arg107 (3.0 Å), between the O3 of Xyl and the side-chain nitrogen atom of Gln29 (2.8 Å), and between the O4 of Xyl and both the main-chain nitrogen atom of Asn19 (2.7 Å) and the carboxylate group of Asp16 (2.6 Å), and additional H-bonds are solvent mediated. These interactions markedly differ from the interactions of Gal and related sugars described earlier. Specific binding of Xyl and Gal was described in the saccharide-binding site of the E86 xylanase of *Streptomyces olivaceoviridis*.<sup>66</sup> This saccharide-binding site occurs in a  $\beta$ -trefoil structure highly similar to the  $\beta$ -trefoil domain found in ricin-like B-chains, and it displays a comparable topology. Binding of Xyl and Gal was shown to involve H-bonds between two hydroxyl groups of the sugar (O2 and O3 for Xyl and O3 and O4 for Gal) and the carboxylate group of Asp325, equivalent to Asp16 or Asp227 found in the saccharide-binding sites of SNA-II. Binding of Gal is remarkably similar to what is observed with SNA-II, and

it would be expected for Xyl to bind in a comparable manner. Hence, it is likely that the binding of Xyl to SNA-II, as observed in the crystal structure reported here, is nonspecific.

### Oligomerization

In previous studies, SNA-II was suggested to exist predominantly as a dimeric entity of 51–60 kDa in solution, as evidenced by gel permeation chromatography.<sup>7,9</sup> However, crystal contacts observed in the two crystal forms reported here suggest that in addition to monomeric species, SNA-II might well exist also as dimers or tetramers displaying internal 222 symmetry (see Fig. 5). This prompted us to further investigate the behavior of SNA-II in solution, by analytical gel filtration chromatography and SAXS. The elution profile of SNA-II on a superdex 75 HR column clearly indicated the existence of at least two distinct oligomeric species, in a concentration range of 1–10 mg/mL. At the highest concentration, traces of aggregated material also occur. The major and minor nonaggregated species displayed an apparent MM of 41 and 60 kDa, respectively (see Fig. 6). The apparent MM of the minor species correlates rather well with a dimer of SNA-II, for which a MM of 62.8 kDa is



**Figure 7**

X-ray scattering pattern of SNA-II and the scattering computed from the models. Dots with error bars: experimental data; Curves 2, 3, 4 are calculated scattering patterns from, respectively, monomeric protein, dimer from the hexagonal crystals, and the best fit from the mixture of the two species.

**Table III**

Summary of Structural Parameters of SNA-II in Solution Computed from the Scattering Data

Model/parameter	$R_g$ (nm)	$D_{max}$ (nm)	MM (kDa)	$\chi_s^a$	$\chi_M^b$
Experiment	$2.5 \pm 0.1$	$9.5 \pm 0.1$	$40 \pm 5$	—	—
Monomer	1.9	7.0	31	2.4	—
Dimer A–C, tetragonal	3.0	10.0	63	4.4	1.6
Dimer A–C, hexagonal	2.9	9.8	63	5.3	0.9
Dimer A–C from the tetramer	2.6	8.8	63	6.7	1.1
Dimer A–G from the tetramer	2.7	9.1	63	7.0	1.0
Tetramer	3.2	9.1	126	19.0	1.2

<sup>a</sup> $\chi$  value indicative of the quality of fit between the experimental diffusion curve and the theoretical curve computed from the corresponding structure alone.<sup>b</sup> $\chi$  value indicative of the quality of fit between the experimental diffusion curve and the theoretical curve computed from a mixture of the corresponding structure in mixture with the monomeric form of SNA-II.

expected. Hence, the major species which displays a longer retention time would correspond to a monomer, although the apparent MM of 40 kDa is significantly higher than 31.4 kDa, MM of a SNA-II monomer. Although these results clearly indicate that more than one oligomeric state exists in solution, they do not allow the unambiguous determination of their stoichiometry: the overall shape of a SNA-II protomer is elongated and might lead to uncertainties in the molecular weight determination.

SAXS experiments were performed to further elucidate the oligomeric composition of SNA-II in solution. The SAXS pattern from SNA-II is presented in Figure 7, and the calculated parameters are summarized in Table III. The experimental radius of gyration ( $R_g$ ) and maximum size of SNA-II in solution ( $2.5 \pm 0.1$  nm and  $9.5 \pm 0.1$  nm, respectively) exceed the values computed from the crystallographic monomer (1.9 nm and 7.0 nm, respectively). The estimated MM of the solute ( $40 \pm 5$  kDa) is also larger than what is expected for the monomer (31 kDa) but still much smaller than that of the dimeric protein (63 kDa). It agrees with the major species identified in the gel filtration experiment (40 kDa). These overall parameters suggest that SNA-II in solution may be a mixture of monomers with higher oligomers.

Theoretical scattering patterns computed from the monomeric protein and from the possible dimers and tetramer found in both crystal forms failed to fit the experiment, with the best agreement provided by the monomer (Curve 2, Fig. 7) still displaying significant systematic deviations ( $\chi = 2.4$ ). The fit was clearly improved when using a mixture of monomer and higher oligomeric species, with  $\chi$  values ranging between 0.9 and 1.6 (Table III). A mixture of 70% monomer and 30% dimer best describes the SAXS experiment (Curve 4, Fig. 7), with the effective MM and  $R_g$  of 42 kDa and 2.4 nm, respectively. It is difficult to unequivocally determine, which oligomeric assembly coexists with the

monomers (given the limited presence of the oligomers in solution, most of the scattering originates from the monomeric fraction). However, these results indicate that SNA-II, independently of the presence of Gal, is predominantly monomeric in solution, in contrast to the previously published results<sup>7,9</sup> suggesting largely dimeric species.

The potential oligomerization of SNA-II can have a significant effect on the accessibility of the various saccharide-binding sites. Some of them might be masked at the interfaces, as exemplified by the tetrameric species that only display four accessible binding sites (see Fig. 5). However, the results of the cocrystallization experiments indicate that the two saccharide-binding sites of SNA-II are accessible in solution. Sugar binding is indeed observed at both sites in four cases (Gal, GalNAc, Lac, and MeGal). Thus, in the conditions used throughout our study, SNA-II is likely to be predominantly monomeric.

## CONCLUSIONS

In this study, we report on the high-resolution structures of SNA-II, a Gal/GalNAc-specific lectin found in the bark of *Sambucus nigra* (black elderberry). SNA-II is a glycosylated polypeptidic chain, homologous to the B-chain of the highly toxic heterodimeric ricin protein, and related Type-II RIPs. The overall shape of SNA-II is similar to that of the Type-II RIPs B-chain, although it is not covalently attached to an A-chain. Glycosylation was observed at four sites, and the electron density maps allowed the identification of 14 saccharide residues. Two saccharide-binding sites, built around conserved Asp16 and Asp227, are described.

In addition, the binding of Gal and five Gal derivatives (GalNAc, Lac, MeGal, Fuc, and the carcinoma-specific Tn antigen) is described at 1.55 Å resolution or better. All six saccharides are observed in the first binding site, whereas Fuc and Tn-antigen binding is not observed in the second binding site. This, and the protein–sugar interactions observed in both sites, suggests that the latter displays a weaker affinity.

The quaternary structure of SNA-II was investigated using analytical gel filtration and SAXS. It is shown that the monomeric species is predominant in solution, although some higher oligomeric species, likely dimers, are detected.

## ACKNOWLEDGMENTS

The authors thank Dr. S. D. Wood for his help at the early stages of this research. They also thank the scientific staff of the European Synchrotron Radiation Facility (Grenoble, France) and EMBL/DESY (Hamburg, Germany) for the use of their excellent data collection facilities.

## REFERENCES

1. Van Damme EJM, Peumans WJ, Barre A, Rougé P. Plant lectins: a composite of several distinct families of structurally and evolutionarily related proteins with diverse biological roles. *Crit Rev Plant Sci* 1998;17:575–692.
2. Endo Y, Mitsui K, Motizuki M, Tsurugi K. The mechanism of action of ricin and related toxic lectins on eukaryotic ribosomes. The site and the characteristics of the modification in 28 S ribosomal RNA caused by the toxins. *J Biol Chem* 1987;262:5908–5912.
3. Van Damme EJ, Hao Q, Chen Y, Barre A, Vandenbussche F, Desmyter S, Rougé P, Peumans WJ. Ribosome-inactivating proteins: a family of plant proteins that do more than inactivate ribosomes. *Crit Rev Plant Sci* 2001;20:395–465.
4. Olsnes S, Pihl A. Different biological properties of the two constituent peptide chains of ricin, a toxic protein inhibiting protein synthesis. *Biochemistry* 1973;12:3121–3126.
5. Sandvig K, van Deurs B. Endocytosis and intracellular transport of ricin: recent discoveries. *FEBS Lett* 1999;452:67–70.
6. Van Damme EJ, Barre A, Barbieri L, Valbonesi P, Rougé P, Van Leuven F, Stirpe F, Peumans WJ. Type 1 ribosome-inactivating proteins are the most abundant proteins in iris (*Iris hollandica* var. Professor Blaauw) bulbs: characterization and molecular cloning. *Biochem J* 1997;324:963–970.
7. Kaku H, Peumans WJ, Goldstein IJ. Isolation and characterization of a second lectin (SNA-II) present in elderberry (*Sambucus nigra* L.) bark. *Arch Biochem Biophys* 1990;277:255–262.
8. Rojo MA, Kaku H, Ishii-Minami N, Minami E, Yato M, Hisajima S, Yamaguchi T, Shibuya N. Characterization and cDNA cloning of monomeric lectins that correspond to the B-Chain of a type 2 ribosome-inactivating protein from the bark of Japanese elderberry (*Sambucus sieboldiana*). *J Biochem (Tokyo)* 2004;135:509–516.
9. Van Damme EJ, Barre A, Rougé P, Van Leuven F, Peumans WJ. 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* 1996;237:505–513.
10. Hakomori S. Tumor-associated carbohydrate antigens. *Annu Rev Immunol* 1984;2:103–126.
11. Brooks SA, Hall DM, Buley I. GalNAc glycoprotein expression by breast cell lines, primary breast cancer and normal breast epithelial membrane. *Br J Cancer* 2001;85:1014–1022.
12. Springer GF. Immunoreactive T and Tn epitopes in cancer diagnosis, prognosis, and immunotherapy. *J Mol Med* 1997;75:594–602.
13. Koch MHJ, Bords J. X-ray diffraction and scattering on disordered systems using synchrotron radiation. *Nucl Instrum Method* 1983;208:461–469.
14. Boulin CJ, Kempf R, Gabriel A, Koch MHJ. Data acquisition systems for linear and area X-ray detectors using delay readout. *Nucl Instrum Methods A* 1988;269:312–320.
15. Gabriel A, Dauvergne F. The localization method used at EMBL. *Nucl Instrum Methods* 1982;201:223–224.
16. Roessle MW, Klaering R, Ristau U, Robrahn B, Jahn D, Gehrmann T, Konarev PV, Round A, Fiedler S, Hermes S, Svergun DI. Upgrade of the small angle X-ray scattering beamline X33 at the European Molecular Biology Laboratory, Hamburg. *J Appl Cryst* 2007;40: s190–s194.
17. Konarev PV, Volkov VV, Sokolova AV, Koch MHJ, Svergun DI. PRIMUS: a Windows PC-based system for small-angle scattering data analysis. *J Appl Cryst* 2003;36:1277–1282.
18. Guinier A. La diffraction des rayons X aux très petits angles: application à l'étude de phénomènes ultramicroscopiques. *Ann Phys (Paris)* 1939;12:161–237.
19. Svergun DI. A direct indirect method of small-angle scattering data treatment. *J Appl Cryst* 1993;26:258–267.
20. Svergun DI. Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *J Appl Cryst* 1992;25:495–503.
21. Svergun DI, Barberato C, Koch MHJ. CRY SOL—a program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. *J Appl Cryst* 1995;28:768–773.
22. Leslie AGW. Profile fitting. In: Helliwell JR, Machin PA, Papiz MZ, editors. Computational aspects of protein crystal analysis. Proceedings of the Daresbury Study Weekend. Daresbury: Daresbury Laboratory; 1987. pp 39–50.
23. Evans PR. Data reduction. In: Sawyer L, Isaacs N, Bayley S, editors. Data collection and processing. Proceedings of the CCP4 study weekend. Daresbury: Daresbury Laboratory; 1993. pp 114–122.
24. CCP4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr D Biol Crystallogr* 1994;50:760–763.
25. Rutenber E, Katzin BJ, Ernst S, Collins EJ, Mlsna D, Ready MP, Robertus JD. Crystallographic refinement of ricin to 2.5 Å. *Proteins* 1991;10:240–250.
26. Vagin AA, Teplyakov A. MOLREP: an automated program for molecular replacement. *J Appl Cryst* 1997;30:1022–1025.
27. Brünger AT, Adams PD, Clore GM, DeLano WL, Gros P, Grosse-Kunstleve RW, Jiang JS, Kuszewski J, Nilges M, Pannu NS, Read RJ, Rice LM, Simonson T, Warren GL. Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr D Biol Crystallogr* 1998;54:905–921.
28. Murshudov G, Vagin AA, Dodson EJ. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr* 1997;53:240–255.
29. Read RJ. Improved Fourier coefficients for maps using phases from partial structures with errors. *Acta Crystallogr A* 1986;42:140–149.
30. Emsley P, Cowtan K. Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* 2004;60:2126–2132.
31. Perrakis A, Sixma TK, Wilson KS, Lamzin VS. wARP: improvement and extension of crystallographic phases by weighted averaging of multiple-refined dummy atomic models. *Acta Crystallogr D Biol Crystallogr* 1997;53:448–455.
32. Frishman D, Argos P. Knowledge-based protein secondary structure assignment. *Proteins* 1995;23:566–579.
33. Holm L, Park J. DaliLite workbench for protein structure comparison. *Bioinformatics* 2000;16:566–567.
34. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. The Protein Data Bank. *Nucl Acids Res* 2000;28:235–242.
35. Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucl Acids Res* 1994;22:4673–4680.
36. Gouet P, Courcelle E, Stuart DI, Metoz F. ESPript: analysis of multiple sequence alignments in PostScript. *Bioinformatics* 1999;15: 305–308.
37. Murzin AG, Lesk AM, Chothia C.  $\beta$ -Trefold fold. Patterns of structure and sequence in the Kunitz inhibitors interleukins-1 $\beta$  and 1 $\alpha$  and fibroblast growth factors. *J Mol Biol* 1992;223:531–543.
38. Wright CS. New folds of plant lectins. *Curr Opin Struct Biol* 1997;7:631–636.
39. Robertus JD, Ready MP. Ricin B chain and discoidin I share a common primitive protein fold. *J Biol Chem* 1984;259:13953–13956.
40. Rutenber E, Ready M, Robertus JD. Structure and evolution of ricin B chain. *Nature* 1987;326:624–626.
41. Bosanac I, Alattia JR, Mal TK, Chan J, Talarico S, Tong FK, Tong KI, Yoshikawa F, Furuichi T, Iwai M, Michikawa T, Mikoshiba K, Ikura M. Structure of the inositol 1,4,5-trisphosphate receptor binding core in complex with its ligand. *Nature* 2002;420:696–700.
42. Segelke B, Knapp M, Kadkhodayan S, Balhorn R, Rupp B. Crystal structure of *Clostridium botulinum* neurotoxin protease in a product-bound state: evidence for noncanonical zinc protease activity. *Proc Natl Acad Sci USA* 2004;101:6888–6893.
43. Mancheno JM, Tateno H, Goldstein IJ, Martinez-Ripoll M, Hermoso JA. Structural analysis of the *Laetiporus sulphureus* hemolytic pore-forming lectin in complex with sugars. *J Biol Chem* 2005;280: 17251–17259.

44. Eriksson AE, Cousens LS, Weaver LH, Matthews BW. Three-dimensional structure of human basic fibroblast growth factor. *Proc Natl Acad Sci USA* 1991;88:3441–3445.
45. Finzel BC, Clancy LL, Holland DR, Muchmore SW, Watenpau K, Einspahr HM. Crystal structure of recombinant human interleukin-1 $\beta$  at 2.0 Å resolution. *J Mol Biol* 1989;209:779–791.
46. Fedorov AA, Fedorov EV, Ono S, Matsumura F, Almo SC. Structure determination of human fascin, an actin-crosslinking protein. American Crystallography Association: Arlington, TX; 1998.
47. Murzin AG, Brenner SE, Hubbard T, Chothia C. SCOP: a structural classification of proteins database for the investigation of sequences and structures. *J Mol Biol* 1995;247:536–540.
48. Rutenber E, Robertus JD. Structure of ricin B-chain at 2.5 Å resolution. *Proteins* 1991;10:260–269.
49. Tahirov TH, Lu TH, Liaw YC, Chen YL, Lin JY. Crystal structure of abrin-a at 2.14 Å. *J Mol Biol* 1995;250:354–367.
50. Bagaria A, Surendranath K, Ramagopal UA, Ramakumar S, Karande AA. Structure-function analysis and insights into the reduced toxicity of *Abrus precatorius* agglutinin I in relation to abrin. *J Biol Chem* 2006;281:34465–34474.
51. Pascal JM, Day PJ, Monzingo AF, Ernst SR, Robertus JD, Iglesias R, Perez Y, Ferreras JM, Citores L, Girbes T. 2.8-Å crystal structure of a nontoxic type-II ribosome-inactivating protein, ebulin I. *Proteins* 2001;43:319–326.
52. Krauspenhaar R, Rypniewski W, Kalkura N, Moore K, DeLucas L, Stoeva S, Mikhailov A, Voelter W, Betzel C. Crystallisation under microgravity of mistletoe lectin I from *Viscum album* with adenine monophosphate and the crystal structure at 1.9 Å resolution. *Acta Crystallogr D Biol Crystallogr* 2002;58:1704–1707.
53. Li M, Chai J-J, Wang Y-P, Wang K-Y, Bi R-C. Crystal structure of *Trichosanthes kirilowii* lectin-1 and its relation to the type 2 ribosomes inactivating proteins. *Protein Pept Lett* 2001;8:81–87.
54. Manoj N, Jeyaprakash AA, Pratap JV, Komath SS, Kenoth R, Swamy MJ, Vijayan M. Crystallization and preliminary X-ray studies of snake gourd lectin: homology with type II ribosome-inactivating proteins. *Acta Crystallogr D Biol Crystallogr* 2001;57:912–914.
55. Dharkar PD, Anuradha P, Gaikwad SM, Suresh CG. Crystallization and preliminary characterization of a highly thermostable lectin from *Trichosanthes dioica* and comparison with other *Trichosanthes* lectins. *Acta Crystallogr F Struct Biol Cryst Commun* 2006;62:205–209.
56. Van Damme EJ, Hao Q, Charels D, Barre A, Rouge P, Van Leuven F, Peumans WJ. Characterization and molecular cloning of two different type 2 ribosome-inactivating proteins from the monocotyledonous plant *Polygonatum multiflorum*. *Eur J Biochem* 2000;267:2746–2759.
57. Fujimoto Z, Kuno A, Kaneko S, Yoshida S, Kobayashi H, Kusakabe I, Mizuno H. Crystal structure of *Streptomyces olivaceoviridis* E-86  $\beta$ -xylanase containing xylan-binding domain. *J Mol Biol* 2000;300:575–585.
58. Notenboom V, Boraston AB, Williams SJ, Kilburn DG, Rose DR. High-resolution crystal structures of the lectin-like xylan binding domain from *Streptomyces lividans* xylanase 10A with bound substrates reveal a novel mode of xylan binding. *Biochemistry* 2002;41:4246–4254.
59. Liu Y, Chirino AJ, Misulovin Z, Leteux C, Feizi T, Nussenzweig MC, Bjorkman PJ. Crystal structure of the cysteine-rich domain of mannose receptor complexed with a sulfated carbohydrate ligand. *J Exp Med* 2000;191:1105–1116.
60. Fu T, Burbage C, Tagge EP, Brothers T, Willingham MC, Frankel AE. Ricin toxin contains three lectin sites which contribute to its in vivo toxicity. *Int J Immunophar* 1996;18:685–692.
61. Frankel AE, Burbage C, Fu T, Tagge E, Chandler J, Willingham MC. Ricin toxin contains at least three galactose-binding sites located in B chain subdomains 1 $\alpha$ , 1 $\beta$ , and 2 $\gamma$ . *Biochemistry* 1996;35:14749–14756.
62. Mishra V, Bilgrami S, Sharma RS, Kaur P, Yadav S, Krauspenhaar R, Betzel C, Voelter W, Babu CR, Singh TP. Crystal structure of himalayan mistletoe ribosome-inactivating protein reveals the presence of a natural inhibitor and a new functionally active sugar-binding site. *J Biol Chem* 2005;280:20712–20721.
63. Babino A, Tello D, Rojas A, Bay S, Osinaga E, Alzari PM. The crystal structure of a plant lectin in complex with the Tn antigen. *FEBS Lett* 2003;536:106–110.
64. Kulkarni KA, Sinha S, Katiyar S, Surolia A, Vijayan M, Suguna K. Structural basis for the specificity of basic winged bean lectin for the Tn-antigen: a crystallographic, thermodynamic and modelling study. *FEBS Lett* 2005;579:6775–6780.
65. Palmer RA, Niwa H. X-ray crystallographic studies of protein–ligand interactions. *Biochem Soc Trans* 2003;31:973–979.
66. Fujimoto Z, Kuno A, Kaneko S, Kobayashi H, Kusakabe I, Mizuno H. Crystal structures of the sugar complexes of *Streptomyces olivaceoviridis* E-86 xylanase: sugar binding structure of the family 13 carbohydrate binding module. *J Mol Biol* 2002;316:65–78.