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


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,1,2,3* Hideaki Niwa,2,3 Valérie Guillet,1 Dmitri I. Svergun,3 Peter V. Konarev,3 Rex A. Palmer,2 Willy J. Peumans,4 Pierre Rougeé,5 Els J. M. Van Damme,4 Colin D. Reynolds,6 and Lionel Mourey1,*

1 Institut de Pharmacologie et de Biologie Structurale (IPBS), UMR 5089, Université Paul Sabatier Toulouse III/CNRS, Toulouse, France
2 School of Crystallography, Birkbeck College, University of London, London, United Kingdom
3 European Molecular Biology Laboratory, Hamburg Outstation, EMBL c/o DESY, Hamburg, Germany; and Institute of Crystallography, Russian Academy of Sciences, Moscow, Russia
4 Department of Molecular Biotechnology, Ghent University, Gent, Belgium
5 Surfaces Cellulaires et Signalisation chez les Végétaux, UMR-CNRS 5546, Pôle de Biotechnologie Végétale, Toulouse, France
6 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 β-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, α1-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.

INTRODUCTION

Plants belonging to the genus Sambucus express a wide range of lectins, with various saccharide-binding specificities.1 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 disrupting protein synthesis in target cells.2 These proteins are distributed among Type-I (single chain) and Type-II (two chains) RIPs.3 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).4 The lectin subunit usually specifically recognizes galactose (Gal), N-acetylgalactosamine (GalNAc) or related sugars.5 It anchors the toxin

1Laurent 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: R13/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. 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. However, as they are also abundantly found in storage organs, they may also be considered as storage proteins.

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. 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, depleted of the eight N-terminal residues, including the cysteine residue involved in the association with the A-subunit. 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. Earlier biochemical data suggested that SNA-II is a dimeric glycosylated protein, which preferentially recognizes Gal and GalNAc, 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, often resulting in the presence of accessible GalNAc residues at the surface. 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, 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.

**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 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 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 nm⁻¹ (*s* = 4π sin(θ)/λ, where θ is the scattering angle, and λ = 0.15 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. All the data analysis steps were performed using the program PRIMUS.

The forward scattering I(0) and the radius of gyration Rg were evaluated using the Guinier approximation assuming that at very small angles (<3/Rg) the intensity is represented as I(0) = I(0) exp[−(sRg)²/3]. These parameters were also computed from the entire scattering pattern using GNOM. 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 CRYSOLO. 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.

**Crystallization and data collection**

Crystals were obtained by the hanging drop vapour diffusion method at 20°C, using ammonium sulphate as precipitant. A 6 μL droplet, made by mixing 3 μL of a solution of SNA-II (16 mg mL⁻¹ in 20 mM Tris HCl buffer at pH 7.5, 20 mM NaCl) and 3 μL of reservoir solution, was equilibrated against 500 μL 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), 5-1-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% (v/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, scaled with SCALA and, subsequently, handled with the CCP4 package. 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 to spacegroup P6122 or P6522, 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 I4122 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) as the search model. The molecular replacement solution found with MOLREP in spacegroup P6522, was first refined using CNS and, when the refined model was considered sufficiently reliable (Rcryst = 0.291, Rfree = 0.300), it was used as a search model for the molecular replacement of the tetragonal crystal form, with MOLREP. From that point, REFMAC 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 were inspected using COOT. Water molecules were automatically introduced into the model using wARP. In the last refinement cycles, individual anisotropic B factors were refined for all structures in the tetragonal spacegroup, which resulted in a similar improvement of the Rcryst and Rfree 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. The structures were superimposed using DaliLite.

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

Protein sequence analysis

Sequences were aligned using CLUSTALW. The alignment was manually optimized and displayed with ESPript.

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 Ca 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 β-trefoil fold [Fig. 1(b)]. In this fold, a central β-barrel built from six antiparallel β-strands delineates a hydrophobic core. At the bottom of the β-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 β-strands arranged as a β-hairpin and an occasional short helical segment. The fold displays a pseudo three-fold symmetry when viewed along the axis of the β-barrel, which results from the organization of three similar subdomains of about 40 residues, named α, β, and γ [Fig. 1(b)]. Each subdomain contains 4 β-strands, the first and fourth contributing to the central β-barrel. The average sequence identity between these subdomains is 18%, which confirms that they have evolved from a common ancestral gene [Fig. 2(a)].

The β-trefoil fold is widely found in protein structures, either associated to a different fold as in the inositol-triphosphate receptor Type-1 binding core, in clostridium neurotoxins, and in the pore-forming lectin from Laetiporus sulphurous, or isolated as in the fibroblast growth factors and interleukin-1. In most proteins, a single occurrence of the β-trefoil fold is found.
### Table I
Data Collection and Crystallographic Refinement Statistics

<table>
<thead>
<tr>
<th></th>
<th>Native</th>
<th>Galactose</th>
<th>GalNAc</th>
<th>Lactose</th>
<th>Me galactose</th>
<th>Fucose</th>
<th>Tn antigen</th>
</tr>
</thead>
<tbody>
<tr>
<td>Beamline ID14-EH1</td>
<td>0.937</td>
<td>0.937</td>
<td>0.979</td>
<td>0.979</td>
<td>0.979</td>
<td>0.979</td>
<td>0.933</td>
</tr>
<tr>
<td>Beamline ID14-EH2</td>
<td>0.933</td>
<td>0.933</td>
<td>0.933</td>
<td>0.933</td>
<td>0.933</td>
<td>0.933</td>
<td>0.933</td>
</tr>
<tr>
<td>Spacegroup P6_22</td>
<td>14,22</td>
<td>14,22</td>
<td>14,22</td>
<td>14,22</td>
<td>14,22</td>
<td>14,22</td>
<td>14,22</td>
</tr>
<tr>
<td>Resolution limits</td>
<td>51.30–1.90 (2.00–1.90)</td>
<td>64.55–1.35 (1.39–1.35)</td>
<td>34.74–1.55 (1.64–1.55)</td>
<td>34.74–1.55 (1.64–1.55)</td>
<td>34.79–1.54 (1.64–1.55)</td>
<td>34.78–1.55 (1.64–1.55)</td>
<td>45.41–1.40 (1.44–1.40)</td>
</tr>
<tr>
<td>No. of observations</td>
<td>287,126 (12,647)</td>
<td>305,842 (16,291)</td>
<td>188,533 (13,275)</td>
<td>193,325 (13,178)</td>
<td>198,082 (14,080)</td>
<td>201,518 (15,040)</td>
<td>248,493 (18,781)</td>
</tr>
<tr>
<td>No. unique reflections</td>
<td>57,308 (5,977)</td>
<td>66,690 (4,901)</td>
<td>42,644 (4,945)</td>
<td>43,195 (5,240)</td>
<td>43,092 (5,002)</td>
<td>42,458 (4,882)</td>
<td>43,501 (5,890)</td>
</tr>
<tr>
<td>Multiplicity</td>
<td>5.0 (2.1)</td>
<td>4.6 (3.3)</td>
<td>4.4 (2.7)</td>
<td>4.5 (2.5)</td>
<td>4.6 (2.8)</td>
<td>4.7 (3.1)</td>
<td>5.7 (3.2)</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>95.5 (70.0)</td>
<td>96.6 (78.0)</td>
<td>97.4 (82.6)</td>
<td>96.7 (78.4)</td>
<td>96.9 (77.2)</td>
<td>98.8 (92.9)</td>
<td>99.5 (98.4)</td>
</tr>
<tr>
<td>R_{sym} b</td>
<td>0.083 (0.324)</td>
<td>0.075 (0.395)</td>
<td>0.052 (0.135)</td>
<td>0.041 (0.126)</td>
<td>0.058 (0.216)</td>
<td>0.044 (0.095)</td>
<td>0.069 (0.180)</td>
</tr>
<tr>
<td>Average I/[σ(I)]</td>
<td>13.7 (2.1)</td>
<td>14.0 (2.3)</td>
<td>21.5 (7.1)</td>
<td>21.6 (5.5)</td>
<td>23.0 (7.6)</td>
<td>17.2 (4.3)</td>
<td>32.4 (11.5)</td>
</tr>
<tr>
<td>No. of protein atoms</td>
<td>2027</td>
<td>2101</td>
<td>2153</td>
<td>2100</td>
<td>2091</td>
<td>2091</td>
<td>2128</td>
</tr>
<tr>
<td>Nonprot. atoms</td>
<td>551</td>
<td>400</td>
<td>543</td>
<td>553</td>
<td>557</td>
<td>524</td>
<td>521</td>
</tr>
<tr>
<td>Res range (Å)</td>
<td>20.00–1.95</td>
<td>20.00–1.35</td>
<td>20.00–1.55</td>
<td>20.00–1.55</td>
<td>20.00–1.55</td>
<td>20.00–1.55</td>
<td>20.00–1.40</td>
</tr>
<tr>
<td>R_{cryst} c</td>
<td>0.173 (52.296)</td>
<td>0.161 (63.280)</td>
<td>0.119 (40.471)</td>
<td>0.125 (40.982)</td>
<td>0.133 (40.807)</td>
<td>0.135 (40.281)</td>
<td>0.125 (41.276)</td>
</tr>
<tr>
<td>R_{free} d</td>
<td>0.194 (2.762)</td>
<td>0.198 (3.366)</td>
<td>0.162 (2.131)</td>
<td>0.173 (2.179)</td>
<td>0.179 (2.155)</td>
<td>0.180 (2.125)</td>
<td>0.167 (2.166)</td>
</tr>
<tr>
<td>Cruickshand DPI (Å)</td>
<td>41.7</td>
<td>24.1</td>
<td>14.2</td>
<td>14.3</td>
<td>17.4</td>
<td>25.2</td>
<td>12.7</td>
</tr>
<tr>
<td>Wilson B (Å)</td>
<td>29.5</td>
<td>16.2</td>
<td>15.5</td>
<td>15.6</td>
<td>15.3</td>
<td>17.4</td>
<td>15.4</td>
</tr>
<tr>
<td>Protein B (Å)</td>
<td>54.2</td>
<td>21.1</td>
<td>16.6</td>
<td>17.0</td>
<td>14.0</td>
<td>21.8</td>
<td>14.5</td>
</tr>
<tr>
<td>Solvent B (Å)</td>
<td>58.8</td>
<td>59.8</td>
<td>32.4</td>
<td>32.9</td>
<td>29.0</td>
<td>37.3</td>
<td>30.2</td>
</tr>
<tr>
<td>Glycan B (Å)</td>
<td>71.0</td>
<td>34.1</td>
<td>43.7</td>
<td>37.8</td>
<td>38.2</td>
<td>37.4</td>
<td>39.1</td>
</tr>
<tr>
<td>Bound sugar B (Å)</td>
<td>—</td>
<td>15.7</td>
<td>14.6</td>
<td>21.2</td>
<td>23.4</td>
<td>27.0</td>
<td>18.8</td>
</tr>
<tr>
<td>Rmsd bond lengths (Å)</td>
<td>0.021</td>
<td>0.021</td>
<td>0.020</td>
<td>0.021</td>
<td>0.019</td>
<td>0.021</td>
<td>0.021</td>
</tr>
<tr>
<td>Rmsd bond angles (°)</td>
<td>1.907</td>
<td>1.917</td>
<td>1.925</td>
<td>1.957</td>
<td>1.891</td>
<td>1.825</td>
<td>2.026</td>
</tr>
<tr>
<td>PDB code</td>
<td>3CA0</td>
<td>3C9Z</td>
<td>3CA1</td>
<td>3CA3</td>
<td>3CA4</td>
<td>3CA5</td>
<td>3CA6</td>
</tr>
</tbody>
</table>

aNumbers in parentheses concern the highest resolution shell.
b$R_{sym} = \Sigma |I_{hkl} - \langle I_{hkl} \rangle| / \Sigma |I_{hkl}|$, where $I_{hkl}$ is the scaled intensity of the hkl reflection, and $\langle I_{hkl} \rangle$ is the mean value of the hkl reflections.
c$R_{cryst} = \Sigma|F_{obs} - F_{calc}| / \Sigma|F_{obs}|$, where $F_{obs}$ and $F_{calc}$ are the observed and calculated structure factor amplitude of the hkl reflection. The number of reflection used for the calculation of $R_{cryst}$ is indicated in parenthesis.
d$R_{free}$ is calculated in the same manner as $R_{cryst}$ 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 α are colored in red, β in green, and γ in blue. The first and fourth β-strands of each subdomain forming the central β-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. Cystine 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 (β-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.
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.46 However, although the β-trefoil folds are similar when considered individually (with an average rmsd of 2.1 Å for 114 common Ca atoms when 57 β-trefoil domain-containing proteins, as identified by SCOP, 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 Ca atoms when compared with SNA-II) (see Fig. 3),25,48 abrin (1ABR, rmsd of 0.8 Å for 249 equivalent Ca atoms),49 Abrus precatorius agglutinin (2Q3N, rmsd of 0.8 Å for 253 equivalent Ca atoms),50 ebulin (1HWM, rmsd of 0.7 Å for 246 equivalent Ca atoms),51 Mistletoe lectin ML-I (1M2T, rmsd of 0.8 Å for 249 equivalent Ca atoms),52 and the Type-II RIP from Trichosanthes kirilowii (1GGP, rmsd of 1.0 Å for 235 equivalent Ca atoms).53 B-chains of Type-II RIPS from other Trichosanthes species also display a similar fold.54,55 Hence, this specific arrangement of two β-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 Iα (Cys14-Cys33), Iβ (Cys55-Cys72), IIα (Cys143-Cys158), and IIβ (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 β-strand to the C terminus of the third β-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.56 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,7 which correspond to ~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-
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)\textsuperscript{48}:

\[
\text{Man(}\alpha1,6)\backslash
\text{Man(}\alpha1,3)\rightarrow\text{Man(}\beta1,4)\rightarrow\text{D-GlcNAc(}\beta1,4)\rightarrow\text{D-GlcNAc(}\beta1)\rightarrow\text{Asn63}
\]

\[
\text{Xyl(}\beta1,2)\rightarrow\text{L-Fuc(}\alpha1,3)
\]

The assigned oligosaccharides linked to the three other glycosylation sites are much smaller. An L-Fuc(\alpha1,3)\rightarrow\text{D-GlcNAc disaccharide could be distinguished at Asn71, whereas a D-GlcNAc(}\beta1,4)\rightarrow\text{D-GlcNAc disaccharide and a}

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).

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 \(~\text{~}40\) amino acid residues.\textsuperscript{39} 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 \textit{Streptomyces}
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. In some cases,
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 IA and IIy definitely possess carbohydrate-binding activity, the occurrence of a third active site in subdomain IB is still debated.60–62 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,59 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.1

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 IA and IIy 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 IA, these residues are Asp16, Gln29, Trp31, Gln36, Asn38, and Gln39, whereas in subdomain IIy 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)].

**Table II**

| Sugar–Protein Interactions Observed in Each Saccharide–Binding Sites |
|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|
| interactions (Å)       | Galactose (α/β)⁴       | GalNAc (α/β)⁴          | Lactose               | Me Galactose           | Fucose                 |
| Domain I               |                        |                        |                        |                        |                        |
| 03–Gln36NE2            | 3.1/2.7                | 3.1/2.8                | 2.9                    | 3.0                    | 2.9                    |
| 03–Asn38 ND2           | 2.6/3.2                | 2.6/3.2                | 2.8                    | 2.9                    | 3.0                    |
| 03–Asp160D1            | 2.7/2.6                | 2.6/2.6                | 2.6                    | 2.6                    | 2.6                    |
| 04–Asp160D1            | 2.6/2.7                | 2.6/2.8                | 2.7                    | 2.6                    | 2.6                    |
| 04–Gln29NE2            | 3.3/3.1                | 3.3/3.1                | 3.2                    | 3.2                    | 3.6                    |
| 04–Asn19N              | 2.8/3.1                | 2.8/3.1                | 2.8                    | 3.0                    | 2.9                    |
| 06–Gln29NE2            | 3.0/3.0                | 3.0/3.0                | 3.0                    | 3.0                    | 2.6                    |
| 06–Asn19N              | 2.5/3.3                | 2.7/3.4                | 2.2                    | 3.1                    | 2.6                    |
| 06–Asn190D1            | 3.0/3.3                | 2.9/3.3                | 2.7                    | 3.3                    | 2.6                    |
| 06–Wat                 | 3.0/2.6                | 2.9/2.6                | 2.8                    | 2.7                    | 2.6                    |
| 07–Ser231OG            | 3.0/2.5                |                        |                        |                        |                        |

| Glc03-Asn190D1         | 2.8                    |                        |                        |                        |                        |

*The two distances correspond to epimers α and β, respectively.

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.7 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 IA, 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 Asp227, and with the main-chain nitrogen atom of Ala230. In the saccharide-binding site of subdomain IA, additional interactions involve the O6 atom of the sugar and both the main-chain nitrogen atom of Asn19, the side-chain amide group of Gln29 and with the main-chain nitrogen atom of 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)].
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 Ia, an acetate ion being found in the saccharide-binding site of subdomain Iγ (Ser231OG-O7) and in the case of Lac in the saccharide-binding site of subdomain Ia (Asn19OD1-GlcO3). The Tn antigen is only bound in the saccharide-binding site of the first β-trefoil domain. The GalNAc part of

![Figure 5](image.png)

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 indentified 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.
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 interac-
tions were also observed in other structures of lectin
complexed to the Tn antigen.63,64

In the hexagonal native crystal form, the saccharide-
binding site of subdomain Iα is occupied by a β-D-xylo-
pyranose unit of the glycan chain attached to Asn63 of a
symmetry-related molecule [Fig. 4(b)].65 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.66 This
saccharide-binding site occurs in a β-trefoil structure
highly similar to the β-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 sac-
charide-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 pre-
dominantly as a dimeric entity of 51–60 kDa in solution,
as evidenced by gel permeation chromatography.7,9
However, crystal contacts observed in the two crystal
forms reported here suggest that in addition to mono-
meric 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 chromatog-
raphy 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 concentra-
tion 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 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).

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.
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 ($\chi^2 = 2.4$). 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$^7,9$ 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 cocryostallization 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.

### Table III

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

<table>
<thead>
<tr>
<th>Model/parameter</th>
<th>$R_g$ (nm)</th>
<th>$D_{max}$ (nm)</th>
<th>MM (kDa)</th>
<th>$\chi^2$</th>
<th>$\chi^2_b$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experiment</td>
<td>2.5 ± 0.1</td>
<td>9.5 ± 0.1</td>
<td>40 ± 5</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Monomer</td>
<td>1.9</td>
<td>7.0</td>
<td>31</td>
<td>2.4</td>
<td>—</td>
</tr>
<tr>
<td>Dimer A–C, tetragonal</td>
<td>3.0</td>
<td>10.0</td>
<td>63</td>
<td>4.4</td>
<td>1.6</td>
</tr>
<tr>
<td>Dimer A–C, hexagonal</td>
<td>2.9</td>
<td>9.8</td>
<td>63</td>
<td>5.3</td>
<td>0.9</td>
</tr>
<tr>
<td>Dimer A–C from the tetramer</td>
<td>2.6</td>
<td>8.8</td>
<td>63</td>
<td>6.7</td>
<td>1.1</td>
</tr>
<tr>
<td>Dimer A–G from the tetramer</td>
<td>2.7</td>
<td>9.1</td>
<td>63</td>
<td>7.0</td>
<td>1.0</td>
</tr>
<tr>
<td>Tetramer</td>
<td>3.2</td>
<td>9.1</td>
<td>126</td>
<td>19.0</td>
<td>1.2</td>
</tr>
</tbody>
</table>

$^a\chi$ value indicative of the quality of fit between the experimental diffusion curve and the theoretical curve computed from the corresponding structure alone.

$^b\chi$ 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.
REFERENCES


