

Isolation, characterization, molecular cloning and molecular modelling of two lectins of different specificities from bluebell (*Scilla campanulata*) bulbs

Lisa M. WRIGHT*¹, Els J. M. VAN DAMME†², Annick BARRE‡, Anthony K. ALLEN§, Fred VAN LEUVEN||, Colin D. REYNOLDS*, Pierre ROUGE‡ and Willy J. PEUMANS†

*School of Biomolecular Sciences, Liverpool John Moores University, Liverpool L3 3AF, U.K., †Laboratory for Phytopathology and Plant Protection, Katholieke Universiteit Leuven, B-3001 Leuven, Belgium, ‡Institut de Pharmacologie et Biologie Structurale, UPR CNRS 9062, 31077 Toulouse Cedex, France, §Molecular Pathology Section, Division of Biomedical Sciences, Imperial College School of Medicine, London SW7 2AZ, U.K., and ||Center for Human Genetics, Katholieke Universiteit Leuven, 3001 Leuven, Belgium

Two lectins have been isolated from bluebell (*Scilla campanulata*) bulbs. From their isolation by affinity chromatography, they are characterized as a mannose-binding lectin (SCAman) and a fetuin-binding lectin (SCAfet). SCAman preferentially binds oligosaccharides with $\alpha(1,3)$ - and $\alpha(1,6)$ -linked manno-pyranosides. It is a tetramer of four identical protomers of approx. 13 kDa containing 119 amino acid residues; it is not glycosylated. The fetuin-binding lectin (SCAfet), which is not inhibited by any simple sugars, is also unglycosylated. It is a tetramer of four identical subunits of approx. 28 kDa containing

244 residues. Each 28 kDa subunit is composed of two 14 kDa domains. Both lectins have been cloned from a cDNA library and sequenced. X-ray crystallographic analysis and molecular modelling studies have demonstrated close relationships in sequence and structure between these lectins and other monocot mannose-binding lectins. A refined model of the molecular evolution of the monocot mannose-binding lectins is proposed.

Key words: carbohydrate binding, evolution, mannose.

INTRODUCTION

Plant lectins are a heterogeneous group of carbohydrate-binding proteins comprising at least seven distinct families of structurally and evolutionarily related proteins [1]. Four of these families, namely the legume lectins, the type 2 ribosome-inactivating proteins (RIPs), the chitin-binding lectins containing hevein domains and the monocot mannose-binding lectins are considered to be 'large' families. The amarantins, the Cucurbitaceae phloem lectins and the jacalin-related lectins comprise at present only a small number of individual lectins and accordingly are considered 'small' families. Amarantins are T-antigen-specific lectins that have been found exclusively in a few *Amaranthus* species. Similarly, the Cucurbitaceae phloem lectins are a small group of chitin-binding lectins confined to the phloem sap of a few genera of the family Cucurbitaceae. Jacalin-related lectins, which are named after jacalin, the T-antigen-specific lectin from jack fruit or *Artocarpus integrifolia*, occur in several species of the family Moraceae and in a few unrelated species such as the Jerusalem artichoke (*Helianthus tuberosus*) and hedge bindweed (*Calystegia sepium*). In contrast with these 'small' lectin families, the occurrence and distribution of the larger lectin groups has been studied in more detail. Legume lectins occur exclusively within the plant family Leguminosae. Over 100 legume lectins have been characterized in detail (e.g. concanavalin A and phytohaemagglutinin). Although all legume lectins are built up of protomers with high sequence similarities and strikingly similar three-dimensional structures, they differ from each other strongly with respect to their sugar-binding specificity. Type 2 RIPs have

been found in plants from different families. Well-known examples of this family are ricin and abrin. In spite of their different taxonomic origins, all type 2 RIPs consist of protomers with high sequence similarities and very similar three-dimensional structures. Most type 2 RIPs have similar though not identical specificities, usually directed against galactose- or *N*-acetyl-galactosamine-containing glycans. Chitin-binding lectins composed of hevein domains are widespread in the plant kingdom. Examples are wheatgerm agglutinin and pokeweed mitogen. All members of this lectin family consist of protomers built up of one, two, three, four or seven so-called hevein domains. Both the amino acid sequences and the three-dimensional structures of the hevein domains are markedly conserved, which explains why all chitin-binding lectins have similar carbohydrate-binding specificities. In contrast with the other large lectin families, which have been studied intensively for several decades, the first monocot mannose-binding lectin was reported only in 1987 when a lectin with an exclusive specificity towards mannose was isolated from snowdrop (*Galanthus nivalis*) bulbs [2]. Since then, related lectins have been found in various tissues of the monocot families Alliaceae, Amaryllidaceae, Araceae, Bromeliaceae, Iridaceae, Liliaceae and Orchidaceae [1]. Biochemical analyses and molecular cloning clearly indicated that all these lectins belong to a single superfamily of mannose-binding proteins, which in accordance with their origin and specificity have been named monocot mannose-binding lectins [3,4]. At present, the monocot mannose-binding lectins are still being studied intensively because of their interesting biological properties, for example as potent inhibitors of retroviruses [5,6] and possible

Abbreviations used: AMA, tuber lectin from *Arum maculatum*; DOM, domain; GNA, *Galanthus nivalis* agglutinin; HCA, hydrophobic cluster analysis; LECSCA, cDNA encoding the lectin from *Scilla campanulata*; RIPs, ribosome-inactivating proteins; SCAman, mannose-binding *Scilla campanulata* lectin; SCAfet, fetuin-binding *Scilla campanulata* lectin.

¹ Present address: Department of Chemistry, University of York, Heslington, York YO1 5DD, U.K.

² To whom correspondence should be addressed (e-mail Els.VanDamme@agr.kuleuven.ac.be).

The nucleotide sequence data reported will appear in DDBJ, EMBL and GenBank Nucleotide Sequence Databases under the accession numbers U97751 and U97752.

applications in crop protection against insects and nematodes [1]. In addition, there is also a great interest in the structural analysis of monocot mannose-binding lectins. X-ray crystallographic studies of the snowdrop and amaryllis lectins revealed a new class of protein fold that consists of three anti-parallel four-stranded β -sheets arranged as a 12-stranded β -barrel [7,8]. Moreover, the monocot mannose-binding lectins exhibit a marked structural diversity, which is reflected in both the number and the overall structure of the protomers. Most monocot mannose-binding lectins are built up of one, two or four protomers consisting of a single domain of approx. 12 kDa. Others, however, are composed of one, two or four protomers consisting of two similar [9] or dissimilar [10] domains of approx. 12 kDa. The degree of similarity between the individual domains varies but can result in two domains that recognize structurally different sugars as is exemplified by the tulip lectin TxLC-I, the protomers of which consist of an N-terminal mannose-binding domain tandemly arrayed to an independently acting GalNAc-binding C-terminal domain [11]. For these reasons the monocot mannose-binding lectins represent a unique system for the study of the molecular evolution of a large family of carbohydrate-binding proteins in terms of both sequence and structure similarities. Because the lectins built up of two-domain protomers are especially important in the further unravelling of the evolution and phylogeny of the monocot mannose-binding lectins, the search for these lectins continues.

This paper reports the isolation, partial characterization, molecular cloning and molecular modelling of two different lectins from bluebell (*Scilla campanulata*), both of which exhibit sequence similarity to the monocot mannose-binding lectins. One of these lectins, called *S. campanulata* mannose-binding lectin (SCAman), has been crystallized, both with and without bound saccharides, and from X-ray diffraction studies seems to have a similar structure to that of snowdrop lectin [12–15]. The second lectin, called *S. campanulata* fetuin-binding lectin (SCAfet), is built up of protomers consisting of two dissimilar tandemly arrayed domains. Sequence analysis of the bluebell lectins allowed us to refine the molecular evolution of the monocot mannose-binding lectins.

EXPERIMENTAL

Reagents

Monosaccharides, fetuin-agarose, glycoproteins and other chemicals were obtained from Sigma Chemical Co. (St. Louis, MO, U.S.A.). Mannose oligosaccharides were from Dextra Laboratories (Reading, U.K.). Sepharose 4B and Superose 12 were purchased from Pharmacia (Uppsala, Sweden).

Fetuin and mannose were coupled to Sepharose 4B by activation with divinylsulphone (1 ml per 10 ml gel) in 0.5 M sodium carbonate, pH 11, for 3 h at 25 °C. After activation, the gel was washed extensively with water. The coupling to fetuin (10 mg/ml) or mannose (100 mg/ml) was for 15 h at 37 °C in 0.5 M sodium carbonate, pH 10. After coupling, the gel was washed thoroughly with water and the remaining activated groups were blocked by incubation in 0.2 M Tris/HCl, pH 8.5, for 3 h at 25 °C.

Plant material

Bluebell (*Scilla campanulata* Ait.) bulbs were obtained from a local garden centre in Leuven, Belgium. For the isolation of the lectins, resting bulbs were washed and then frozen at –20 °C until required. For the isolation of RNA the shoots present in

resting bulbs were dissected, frozen in liquid N₂ and stored at –80 °C until use.

Isolation of the lectins

Because SCAman and SCAfet exhibited different saccharide specificities, the lectins could be isolated from the same extract by successive chromatography steps on mannose–Sepharose 4B and fetuin–Sepharose 4B affinity columns respectively.

Bulbs (200 g) were thawed, diced and homogenized in a Waring blender in 1 litre of distilled water containing 0.2% ascorbic acid, pH 6.5. The homogenate was filtered through cheesecloth and the filtrate was centrifuged (3000 g for 5 min). To the clarified supernatant was added CaCl₂ (2 g/l) and the pH was increased to 9.0 with 0.5 M NaOH. After being left for 12 h at 4 °C the solid material was removed by centrifugation (3000 g for 10 min) and the supernatant was filtered through glass wool. The pH was lowered to 3.0 with 1 M HCl, and the extract was applied to an S Fast Flow column (2.6 cm × 5 cm; 25 ml bed volume; Pharmacia). After the column had been washed with water, the protein was eluted with 1 M NaCl until the A_{280} decreased below 0.01. The pH of the eluate was adjusted to 7.0 and then centrifuged (3000 g for 5 min). Solid (NH₄)₂SO₄ (132 g/l) was added to the decanted supernatant and the solution was loaded on a mannose–Sepharose 4B affinity column (2.6 cm × 5 cm; 25 ml bed volume) that had been equilibrated with 1 M (NH₄)₂SO₄. After the column had been washed with 1 M (NH₄)₂SO₄ until the A_{280} fell below 0.01, the bound lectin (SCAman) was desorbed with 20 mM unbuffered 2,3-diaminopropane. The lectin fractions were pooled, then neutralized to pH 7.0 and stored at –20 °C until required.

To purify SCAfet, the extract that had previously been passed through the mannose–Sepharose 4B column was loaded on an affinity column (2.6 cm × 5 cm; 25 ml bed volume) of fetuin–Sepharose 4B. After the column had been washed with 1 M (NH₄)₂SO₄ until the A_{280} decreased below 0.01, the bound lectin (SCAfet) was desorbed with 20 mM unbuffered diaminopropane, neutralized to pH 7.0 and kept at –20 °C until needed.

Analyses

Samples of the lectin were hydrolysed in 6 M HCl for 24 h and 120 h at 110 °C, followed by evaporation and redissolving in a pH 2.2 buffer. Hydrolysates were analysed on an LKB Alpha Plus Analyzer with sodium buffers and ninhydrin detection in accordance with the manufacturer's instructions. The tryptophan content was determined spectrophotometrically. For cysteine determination the performic acid oxidation method was used followed by hydrolysis in HCl. Glucosamine and labile amino acids such as methionine and tyrosine were analysed after hydrolysis in 3 M toluene-*p*-sulphonic acid [16]. Neutral sugars were analysed after methanolysis and trimethylsilylation by GLC [17].

Lectin activity was assayed by haemagglutination with a 4% (v/v) suspension of rabbit erythrocytes and a serial dilution method (with 2-fold increments) was used for determining the inhibitory effects of sugars and glycoproteins [18].

The methodology of the anti-HIV assays has been described previously [19,20]. In brief, CEM cells (4×10^5 cells/ml) were suspended in fresh culture medium and infected with HIV-1 (IIB) and HIV-2 (ROD) at 100 CCID₅₀ per ml cell suspension (1 CCID₅₀ being the cell culture dose infective for 50% of the cell cultures). Then 100 μ l of the infected cell suspension was transferred to microplate wells, mixed with 100 μ l of the appropriate dilutions of the test compounds, and further incubated at 37 °C. After 4 days, syncytium formation was examined in the HIV-

infected cell cultures. Antiviral activity was expressed as EC_{50} (the compound concentration required to inhibit HIV-induced syncytium formation by 50%).

Proteins were analysed by SDS/PAGE [12.5–25% (w/v) gradient gels] as described by Laemmli [21].

FPLC was used for gel filtration of the purified *S. campanulata* lectins on a Pharmacia Superose 12 column previously equilibrated with PBS containing 0.1 M mannose and 0.1 M galactose. Various well-characterized proteins were used as molecular mass reference markers.

Protein sequencing was conducted on an Applied Biosystems (Foster City, CA, U.S.A.) Model 477A protein sequencer interfaced with an Applied Biosystems model 120A on-line analyser.

RNA isolation, and construction and screening of cDNA library

Total cellular RNA was prepared from very young shoots of *S. campanulata* (still contained in the bulbs) essentially as described by Van Damme and Peumans [22]. A cDNA library was constructed from total RNA by using the cDNA synthesis kit from Pharmacia. cDNA fragments were inserted into the *EcoRI* site of PUC18 (Pharmacia). The library was propagated in *Escherichia coli* XL1 Blue (Stratagene, La Jolla, CA, U.S.A.).

Recombinant lectin clones were screened with the use of synthetic oligonucleotides derived from the N-terminal amino acid sequences of the *S. campanulata* lectin polypeptides (QPDDNH, 3'-TGA/G TTA/G TCA/G TCN GGC/T TG-5'; DNHPQL, 3'-ATC/T TGN GGA/G TGA/G TTA/G TC-5') or the random-primer-labelled cDNA clones encoding the tulip lectins [11] as probes. In a later experiment cDNA clones encoding the *S. campanulata* lectins were used as probes. Hybridization was performed overnight as reported previously [11]. Colonies that produced positive signals were selected and rescreened at low density under the same conditions. Plasmids were isolated from purified single colonies on a miniprep scale by using the alkaline lysis method described by Mierendorf and Pfeffer [23] and sequenced by the dideoxy method [24]. DNA sequences were analysed with the use of programs from PC Gene (Intelligenetics, Mountain View, CA, U.S.A.) and Genepro (Riverside Scientific, Seattle, WA, U.S.A.).

Northern blotting

RNA electrophoresis was performed by the method of Maniatis et al. [25]. Approx. 50 μ g of total RNA was denatured in glyoxal and DMSO and separated in a 1.2% (w/v) agarose gel. After electrophoresis the RNA was transferred to Immobilon N membranes (Millipore, Bedford, MA, U.S.A.) and the blot was hybridized with a random-primer-labelled lectin cDNA insert. Hybridization was performed as reported by Van Damme et al. [9]. An RNA ladder (0.16–1.77 kb) was used as a marker.

Molecular modelling of the *Scilla* sequences

The amino acid sequence alignments were performed on a MicroVAX 3100 (Digital, Evry, France) with the IALIGN program of PIR/NBRF (Washington, DC, U.S.A.). The program SEQVU (Gardner, J., 1995, The Garvan Institute of Medical Research, Sydney, Australia) running on a Macintosh LC 630 was used to compare the amino acid sequences of the lectins. MACCLADE [26] was run on a Macintosh LC 630 to build a parsimony phylogenetic tree of the monocot mannose-binding lectins.

A hydrophobic cluster analysis (HCA) [27,28] was performed to delineate the structurally conserved β -sheets along the amino acid sequences of the *S. campanulata* lectins and GNA, the

mannose-specific *Galanthus nivalis* agglutinin that was used as a model. HCA plots were generated on a Macintosh LC with the program HCA-PLOT2 (Doriane, Paris, France).

Molecular modelling of the two domains composing SCAfet was performed on a Silicon Graphics iris 4D25G workstation, with the programs INSIGHII, HOMOLOGY and DISCOVER (Biosym Technologies, San Diego, CA, U.S.A.). The atomic coordinates of SCAMAN [28a] and GNA (Brookhaven Protein Data Bank code 1msa) [7] were used to build the three-dimensional models of the two lectin domains composing SCAfet. Energy minimization and relaxation of the loop regions was performed by several cycles of steepest descent and conjugate gradient by using the cvff forcefield of DISCOVER. The program TURBOFRDO (Bio-Graphics, Marseille, France) was run on a Silicon Graphics Indigo R3000 workstation to perform the superposition of the models and the docking of mannose into the binding sites of the lectins. The lowest apparent binding energy [E_{bind} , expressed in kcal/mol (1 kcal \approx 4.2 kJ)] compatible with the four hydrogen bonds (considering van der Waals interactions and strong [$2.5 \text{ \AA} < \text{dist}(\text{D}-\text{A}) < 3.1 \text{ \AA}$ and $120^\circ < \text{ang}(\text{D}-\text{H}-\text{A})$] and weak [$2.5 \text{ \AA} < \text{dist}(\text{D}-\text{A}) < 3.5 \text{ \AA}$ and $105^\circ < \text{ang}(\text{D}-\text{H}-\text{A}) < 120^\circ$] hydrogen bonds, in which D is donor, A acceptor and H hydrogen) found in the GNA-mannose complex [7] was calculated with the cvff forcefield and used to anchor the pyranose ring of mannose into the binding sites of SCAfet. Cartoons were generated by using MOLSCRIPT [29].

RESULTS

Bluebell bulbs contain two structurally different lectins belonging to the superfamily of monocot mannose-binding lectins

Two different lectins, SCAMAN and SCAfet, were isolated from a partly purified extract of bluebell bulbs by successive affinity chromatography on mannose-Sepharose 4B and fetuin-Sepharose 4B respectively. The overall yields of SCAMAN and SCAfet were approx. 1.4 and 1.1 mg/g fresh tissue, indicating that both lectins are present at comparable and reasonably high levels.

To determine the molecular structure, affinity-purified lectins were analysed by SDS/PAGE and gel filtration. In addition, the lectins were subjected to carbohydrate analyses and N-terminal sequencing. Analysis of SCAMAN by SDS/PAGE gave bands of an apparent molecular mass of approx. 13 kDa for both reduced and unreduced samples. On gel filtration on a Pharmacia Superose 12 column, SCAMAN was eluted with an apparent molecular mass of 58 kDa (results not shown), which suggests that SCAMAN is a tetramer composed of four identical non-covalently linked protomers of approx. 13 kDa. N-terminal sequencing yielded the single sequence NNIIFSKQPDDNHP-QILHAT. No sugars were detected by GLC and no glucosamine was detected on the amino acid analyser trace, indicating that SCAMAN is not glycosylated.

SDS/PAGE of SCAfet under reducing as well as non-reducing conditions gave one band with an apparent molecular mass of 28 kDa. On gel filtration on the Superose 12 column the native SCAfet was eluted with an apparent molecular mass of 110 kDa (results not shown). These results indicate that SCAfet is organized as a tetramer containing four identical subunits of 28 kDa. As will be shown below, the 28 kDa polypeptide of SCAfet is composed of two different lectin domains. N-terminal sequencing of the 28 kDa SCAfet polypeptide yielded the single sequence NNILFGLSHEGSHPQTL. Determination of the total carbohydrate content of SCAfet by GLC and the amino acid analyses for glucosamine yielded similar results to those of SCAMAN, indicating that this lectin is also unglycosylated.

Table 1 Inhibition of SCAMan by saccharides

Abbreviation: n.d., not determined.

Saccharide	IC ₅₀ (mM)
D-Man	10
α-Me Man	10
β-Me Man	40
D-Lyxose	40
2-Deoxy-D-Man	n.d.*
D-ManNAc	n.d.†
D-Glc	n.d.*
Man-α(1,2)-Man	20
Man-α(1,3)-Man	1
Man-α(1,6)-Man	2
Man-α(1,3:1,6)-mannotriose	0.25

* No inhibition at 100 mM.

† No inhibition at 50 mM.

Carbohydrate-binding specificity and biological activities of the bluebell lectins

SCAMan and SCAFet readily agglutinate trypsin-treated rabbit erythrocytes, the minimum concentrations required for agglutination being 15 and 25 µg/ml respectively. In the same test GNA showed a specific agglutination activity of 1 µg/ml. Both *S. campanulata* lectins were inactive towards untreated and trypsin-treated human erythrocytes (irrespective of the blood group) even at concentrations as high as 5 mg/ml. GNA was also unreactive towards human erythrocytes.

The carbohydrate-binding specificities of SCAMan and SCAFet were determined in some detail by using hapten inhibition assays of the agglutination of trypsin-treated rabbit erythrocytes. As shown in Table 1, only mannose derivatives inhibited the agglutinating action of SCAMan with a preference for α- over β-anomers. The axial OH at C-2 is essential for binding, because glucose (with an equatorial OH) was not an inhibitor and 2-deoxymannose (with an H in position 2) and *N*-acetylmannosamine (with an acetamido group in place of the OH group) were not inhibitory. The 6-OH does not seem to be essential for binding, because lyxose (which is the pentose with an equivalent structure to mannose, except for the lack of the 6-hydroxymethyl group) was an inhibitor of the lectin. The inhibition assays showed that SCAMan exhibits a strong affinity for disaccharides or trisaccharides containing α(1,3)- or α(1,6)-linked mannosyl residues, but only a weak affinity for the α(1,2)-linked disaccharide.

Similar hapten inhibition assays with SCAFet showed that none of the monosaccharides or oligosaccharides tested had an inhibitory effect on the agglutination of trypsin-treated rabbit erythrocytes. A combination of mannose and GalNAc also failed to prevent the agglutination of rabbit erythrocytes by SCAFet. Therefore the apparent complex specificity of SCAFet cannot be ascribed to the simultaneous occurrence of both mannose- and GalNAc-binding domains as was demonstrated for a related fetuin-binding lectin from tulip bulbs [11]. Assays with some animal glycoproteins revealed that the agglutination activity of SCAFet can be inhibited by thyroglobulin, asialofetuin, fetuin and ovomucoid, the concentrations required for 50% inhibition being 60, 16, 250 and 125 µg/ml respectively.

Samples of both bluebell lectins were also tested for their inhibitory effect against HIV-1- and HIV-2-induced cytopathicity in CEM cells. SCAMan inhibited the infection of the target cells

A LECSCA1

```

CILSLIVIPGIL/LQSLICNA NNILFESKOPDDNHPQLHATESLEILYGTHVYRPI 35
MQTDCNLVLYDNNPIWATNTGGGLGNGCRAVLQPDGVLVITNENVTVWQSPVAGKA 92
GHYVLVLQPRDRNVVIYGDALWATQTVRGGSGVMPLLRKTRQIE 134

```

B LECSCA2

```

STLSLILIFGIFSQLSLICYG NNILFGLSHGSHQPTLHAAQSLRLSSFRFTMQSD 35
CNLVLFPDSDRVWASNTAGATGCRAVLQSDGLLVILTAQNTIRWSGRTKSGIGNYVL 92
VLQPDRTVTVIYGPGLWDSGTSNKGSVVAVANNNSILYSTQGNMHPQLHATQSLQL 149
SPYRLSMETDCNLVLPDRDRVWSTNTAGKGTGCRAVLQPNGRMDVLTNQNIAVWTS 206
GNSRSAGRYVFLQPRDLNLAITYGGALWTTGSTAKVAAA 244

```

Figure 1 Deduced amino acid sequences of the cDNA clones LECSCA1 and LECSCA2 encoding SCAMan (A) and SCAFet (B) respectively

The N-terminal sequences of the proteins are underlined. The arrowhead indicates the cleavage site of the signal peptide.

by HIV at EC₅₀ values of 4.6 and 8 µg/ml for HIV-1 and HIV-2 respectively. However, the inhibitory potency of SCAMan was much lower than that of, for example, the orchid lectin from *Listera ovata*, which has an EC₅₀ of 0.5 µg/ml and is known as a potent antiviral protein [5,6]. SCAFet did not inhibit the infection of the target cells at a concentration of 40 µg/ml.

Isolation and characterization of cDNA clones encoding the *Scilla* lectins

Screening of a cDNA library constructed from total RNA isolated from young shoots of *S. campanulata* resulted in the isolation of two classes of cDNA clones (Figure 1). Sequencing of the clones revealed that the first group of clones of approx. 600 bp (called LECSCA1) encodes SCAMan because their deduced amino acid sequence comprised the N-terminal 20 residues of the SCAMan polypeptide. The second group of cDNA clones of approx. 1 kb (called LECSCA2) encodes SCAFet because the deduced amino acid sequence of this clone matched the N-terminal sequence (17 residues) of the 28 kDa SCAFet polypeptide. Northern blot analysis further demonstrated that SCAMan is translated from an mRNA of approx. 800 nt, whereas SCAFet is encoded by an mRNA with an estimated length of approx. 1100 nt (results not shown).

LECSCA1 contains an open reading frame of 465 bp encoding a 155-residue precursor. By using the program PSIGNAL from the software package PCGENE, a putative signal sequence was identified in the deduced amino acid sequence of LECSCA1. However, the signal sequence was incomplete in that it lacked a translation initiation codon. Cleavage of the signal sequence between residues 21 and 22 conformed to the rules for protein processing of Von Heijne [30] and yielded a lectin polypeptide of 134 residues (14783 Da) with an N-terminal sequence identical with that obtained by sequencing the SCAMan polypeptides of 13 kDa. Because according to X-ray crystallographic studies of SCAMan [31] the mature protein contains only 119 residues, the biosynthesis of the lectin involves a post-translational cleavage of a 15-residue C-terminal propeptide. The deduced amino acid sequence of LECSCA1 contains one putative N-glycosylation site at position 101 of the lectin precursor. However, according to X-ray crystallographic studies [15] and carbohydrate analysis, SCAMan is, like all other monocot mannose-binding lectins, not glycosylated.

The amino acid composition of SCAMan calculated from the sequence results is in good agreement with that of the purified

GNA	1	D	N	L	I	L	V	S	-	-	-	-	G	E	L	S	T	G	E	F	N	-	-	G	S	F	V	F	I	M	E	D	28								
SCAman	1	N	N	L	I	F	S	K	Q	P	D	N	H	P	Q	I	L	H	A	T	E	I	S	L	E	L	L	Y	G	T	H	V	Y	R	F	I	M	Q	T	D	39
SCAfet-DOM1	1	N	N	L	I	F	S	K	Q	P	D	N	H	P	Q	I	L	H	A	T	E	I	S	L	E	L	L	Y	G	T	H	V	Y	R	F	I	M	Q	T	D	35
SCAfet-DOM2	1	N	S	L	L	V	S	T	G	N	D	N	H	P	Q	I	L	H	A	T	E	I	S	L	E	L	L	Y	G	T	H	V	Y	R	F	I	M	Q	T	D	35
GNA	29	C	N	L	V	L	V	D	K	P	I	W	A	T	N	T	G	L	S	R	S	C	I	F	L	S	M	Q	T	D	C	N	L	V	V	N	P	67			
SCAman	40	C	N	L	V	L	V	D	N	N	P	I	W	A	T	N	T	G	L	N	G	C	R	A	V	L	Q	P	D	G	V	L	V	I	T	N	78				
SCAfet-DOM1	36	C	N	L	V	L	V	D	S	D	V	R	V	W	A	S	N	T	A	G	-	A	T	G	C	R	A	V	L	Q	S	D	G	L	V	I	T	N	73		
SCAfet-DOM2	36	C	N	L	V	L	V	D	R	D	R	V	W	S	T	N	T	A	G	K	G	T	G	C	R	A	V	L	Q	P	N	G	R	M	D	V	L	T	N	74	
GNA	68	S	N	K	P	I	W	A	S	N	T	G	G	O	N	G	N	Y	V	C	L	Q	K	D	R	N	V	V	I	Y	G	T	D	R	W	A	T	G	T	106	
SCAman	79	E	N	V	T	V	W	S	P	V	A	G	K	A	G	H	Y	V	L	V	L	Q	P	D	R	N	V	V	I	Y	G	D	A	L	W	A	T	G	T	117	
SCAfet-DOM1	74	Q	N	T	I	W	S	S	G	T	K	G	S	L	G	N	Y	V	L	V	L	Q	P	D	R	T	V	T	I	Y	G	P	L	W	S	G	T	G	112		
SCAfet-DOM2	75	Q	N	I	A	V	W	T	S	G	N	S	R	S	A	G	R	I	V	F	V	L	Q	P	D	R	N	L	A	L	Y	G	A	L	W	T	G	S	113		
GNA	107	H	T	G	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	109		
SCAman	118	V	R	G	G	S	S	G	V	M	P	L	L	K	T	R	Q	I	E	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	134		
SCAfet-DOM1	113	S	N	K	G	S	V	V	A	N	N	G	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	124		
SCAfet-DOM2	114	T	A	K	-	-	V	A	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	120		

Figure 2 Comparison of the amino acid sequences of GNA with those of SCAman, SCAfet-DOM1 and SCAfet-DOM2

Deletions are indicated by dashes and identical residues are boxed.

protein (results not shown), which confirms that LECSA1 encodes SCAman.

LECSA2 contains an open reading frame of 795 bp. Like LECSA1, LECSA2 contains only a partial signal sequence, which, according to the rules for protein processing of Von Heijne [30], can be cleaved between residues 21 and 22, resulting in a lectin polypeptide of 26224 Da (244 residues). A detailed analysis of the deduced amino acid sequence of the cDNA clone LECSA2 revealed that the sequence contains two very similar domains, designated SCAfet-DOM1 and SCAfet-DOM2, which show 55% sequence identity and 68% sequence similarity (Figure 2). The sequence of SCAman shows 48% and 53%

sequence identity to the first and second domains of SCAfet respectively.

The amino acid composition of SCAfet calculated from the sequence data is in good agreement with that of the purified protein (results not shown), which confirms that LECSA2 encodes SCAfet.

Molecular modelling of the *Scilla* sequences

The amino acid sequences of SCAfet-DOM1 and SCAfet-DOM2 are closely related to that of GNA except for an insertion of seven residues at the N-terminal end of the sequences (Figure 2). Such an insertion also occurs in SCAman, which exhibits an additional N-terminal insertion of four residues that are lacking in GNA and SCAfet. Accordingly, percentages of identity and similarity close to 40–50% and 65–70% respectively relate all these amino acid sequences.

In addition to these sequence similarities, structural similarities occur when the HCA plots of both domains of SCAfet are compared with those of GNA and SCAman, suggesting that all these proteins have very similar three-dimensional structures. In this respect the localization of the 12 strands of β -sheet occurring along the HCA plot of GNA are readily recognized on the HCA plots of both domains of SCAfet (Figure 3) and of SCAman. These structurally conserved regions were used to build the three-dimensional models of both domains of SCAfet from the X-ray coordinates of GNA. However, owing to the occurrence of an extra loop of seven residues at the N-terminal end of SCAfet (which is also present in SCAman), the X-ray coordinates of this

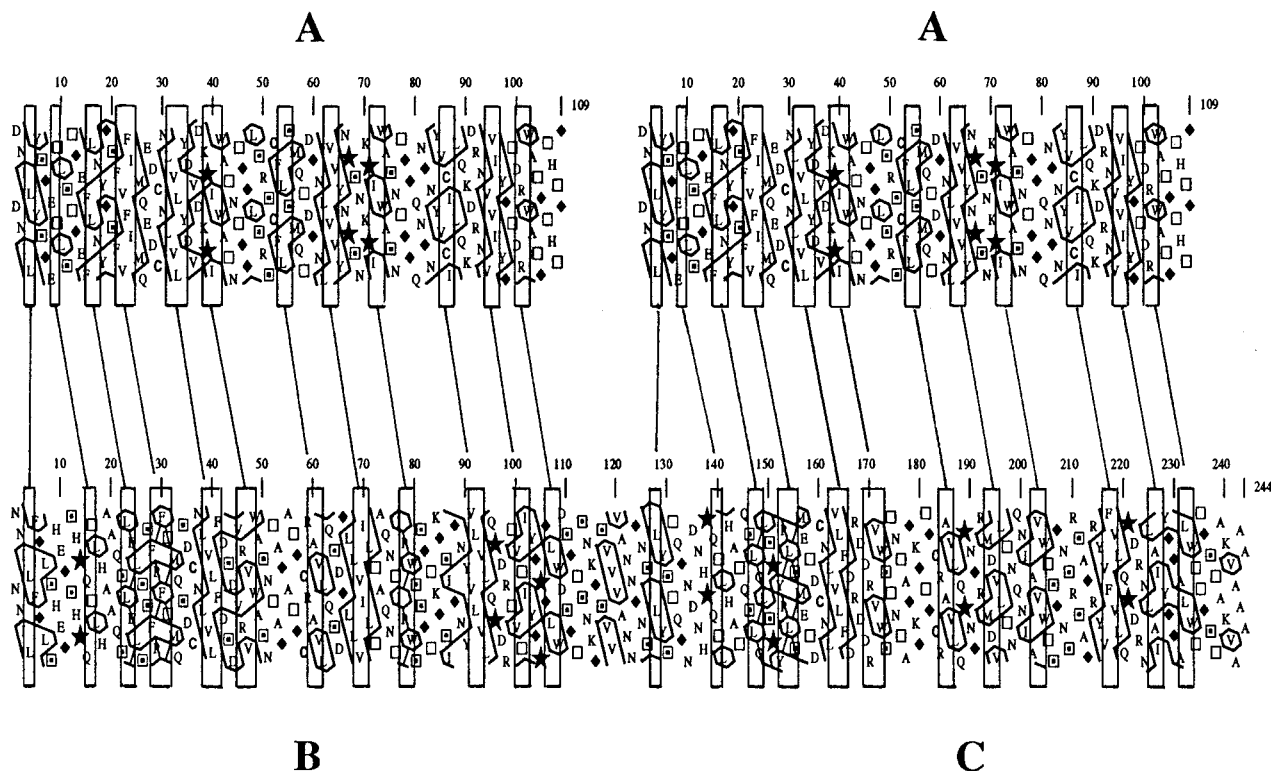


Figure 3 Comparison of the HCA plots of GNA (A) with those of SCAfet-DOM1 (B) and SCAfet-DOM2 (C)

The 12 strands of β -sheet delineated on the HCA plot of GNA are reported on the HCA plots of both domains of SCAfet. These delineations were used to recognize the structurally conserved regions between GNA and SCAfet.

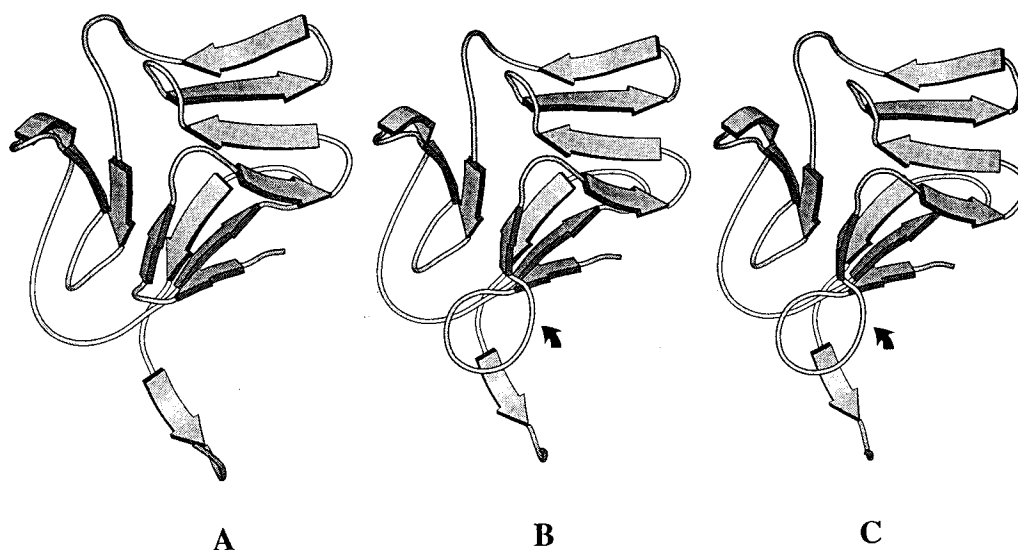


Figure 4 Diagrams generated with MOLSCRIPT of GNA (A), SCAfet-DOM1 (B) and SCAfet-DOM2 (C)

Strands of β -sheet are represented by arrows; the arrowhead indicates the extra loop occurring at the N-terminus of SCAfet-DOM1 and SCAfet-DOM2.

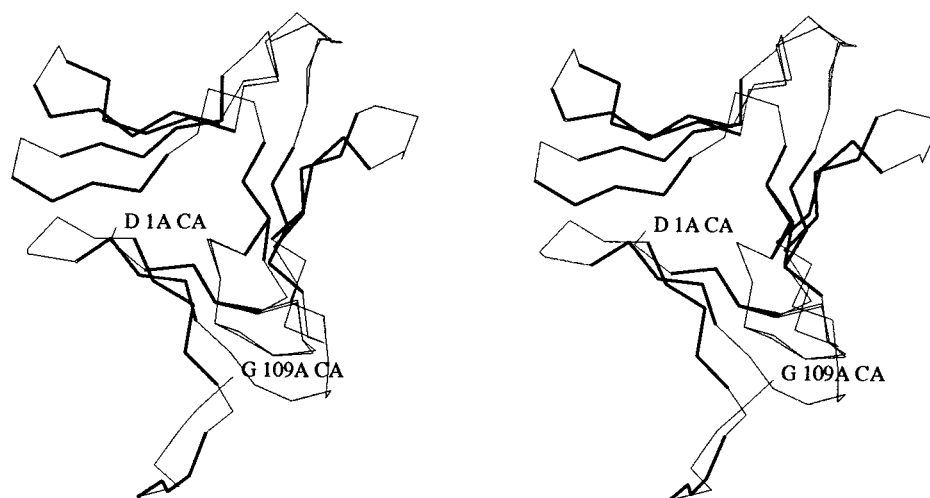


Figure 5 Stereo view showing the superposition of the α -carbon tracings of the three-dimensional models of GNA, SCAfet-DOM1 and SCAfet-DOM2

Strands of β -sheet (thick lines) are well superimposed, whereas a few conformational changes occur in loops (thin lines).

latter lectin were used for a more accurate modelling of this loop region [12].

The three-dimensional models obtained for SCAfet-DOM1 and SCAfet-DOM2 (Figure 4) from the coordinates of GNA and SCAMAN were readily superposable on those of the model lectins (Figure 5). All these proteins exhibited three bundles of anti-parallel β -sheet interconnected by loops to form a 12-stranded β -barrel. However, some changes occurred in the overall folding of the polypeptide chain, mainly located in the extra-loop region and in another region of SCAfet-DOM1 where a deletion of a single residue was shown to occur when compared with GNA or SCAMAN.

The amino acid residues forming the three mannose-binding sites in both domains of SCAfet have undergone some changes from those found in GNA. Tyr³⁴ of the binding site of subdomain

III of GNA is replaced by Phe⁴¹ in both domains of SCAfet, which suggests that this binding site is non-reactive because, as is shown by docking experiments, no hydrogen bond can occur between Phe⁴¹ and O-4 of mannose (Figure 6A). Similarly, Asn⁶¹ and Tyr⁶⁵ of the binding site of subdomain II of GNA are replaced by Leu⁶⁷ and Leu⁷¹ in SCAfet-DOM1 and Arg⁶⁸ and Leu⁷² in SCAfet-DOM2. Leu⁷¹ of SCAfet-DOM1 and Leu⁷² of SCAfet-DOM2 cannot make hydrogen bonds with O-4 of mannose (Figure 6B). In addition, the side chain of Arg⁶⁸ is too far from O-2 to create a hydrogen bond. Accordingly, this binding site is believed to be non-reactive towards mannose in both domains of SCAfet. The amino acid residues forming the binding site of subdomain I of GNA are unchanged in SCAfet-DOM2, whereas a single substitution (Asn⁹³ replaced by Thr⁹⁹) occurs in the binding site of SCAfet-DOM1. This suggests that

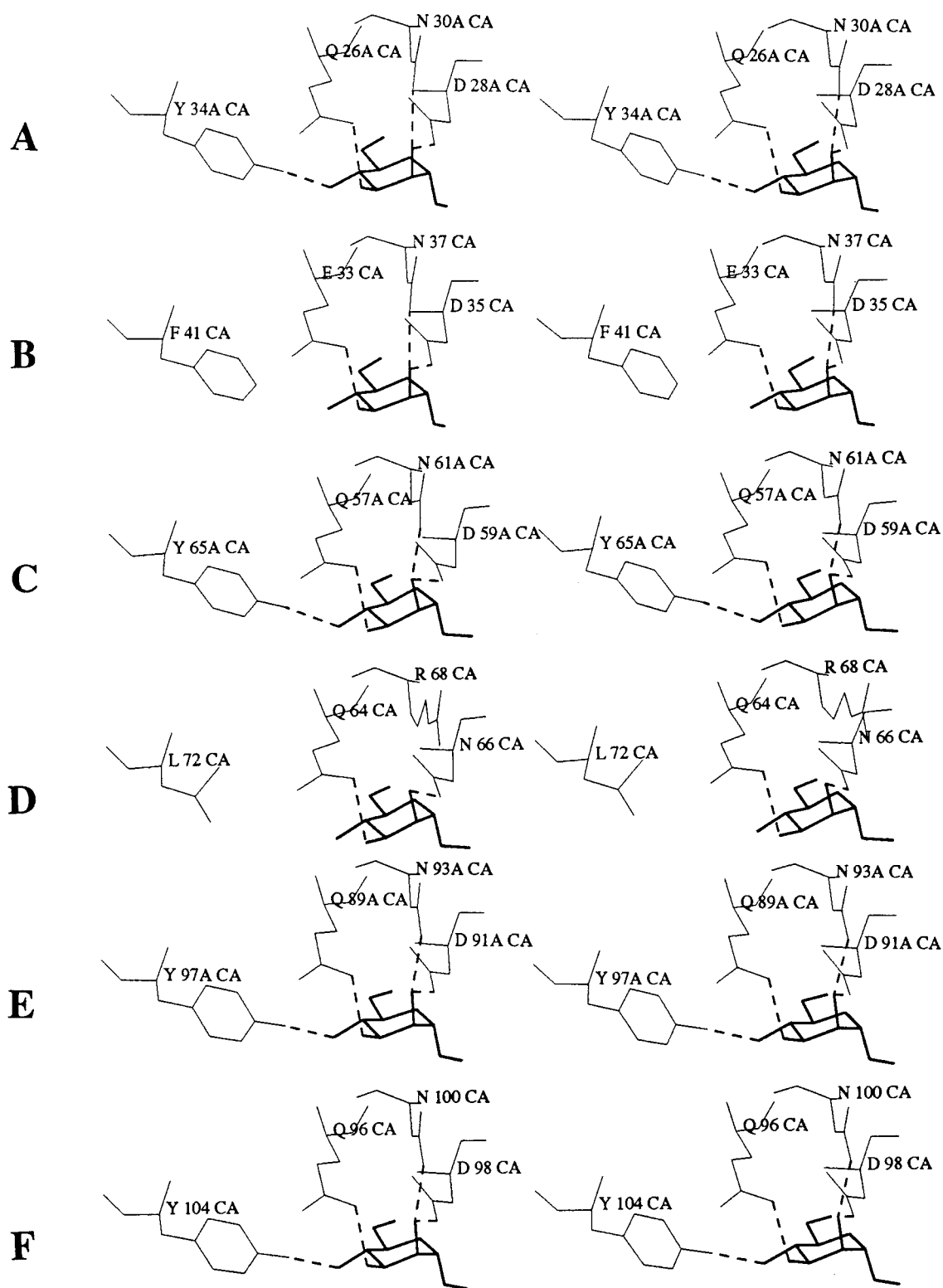


Figure 6 Stereo views showing the docking of mannose into the three mannose-binding sites of SCAfet-DOM2 (B, D, F) compared with mannose bound to the three mannose-binding sites of GNA (A, C, E)

(A, B) Site of subdomain III; (C, D) site of subdomain II; (E, F) site of subdomain I. Broken lines correspond to the hydrogen bonds connecting mannose to the amino acid residues of the binding sites.

the binding site of SCAfet-DOM2 can bind mannose (Figure 6C), whereas that occurring in SCAfet-DOM1 should be unreactive because Thr⁹⁹ is rather too distant from O-2 of mannose (3.61 Å) to make a hydrogen bond. In summary, only the C-terminal binding site of the SCAfet polypeptide presumably possesses mannose-binding activity.

Evolutionary relationships of SCAMAN and SCAfet to other monocot mannose-binding lectins

To trace the evolutionary relationships of SCAMAN and SCAfet, a phylogenetic tree based on a distance matrix was built from the sequences of the individual domains of both bluebell lectins and other monocot mannose-binding lectins. As is shown in Figure 7, both *S. campanulata* lectins are grouped in a single cluster that comprises, besides SCAMAN and SCAfet, the single-domain mannose-binding lectins from the Liliaceae species *Polygonatum multiflorum* [32] and *Aloe arborescens* [33].

DISCUSSION

This report describes a biochemical and molecular biological study of the lectins from *S. campanulata* bulbs. By using a combination of protein purification and analysis together with cDNA cloning it was demonstrated that bluebell bulbs contain two structurally different lectins that both belong to the superfamily of the monocot mannose-binding lectins. SCAMAN, which is a tetramer of four identical one-domain protomers, closely resembles the classical monocot mannose-binding lectins with respect to its molecular structure and amino acid sequence. In addition, the carbohydrate-binding specificity of SCAMAN is similar to, but not identical with, the mannose-specific lectins from snowdrop [34], daffodil and amaryllis [35]. SCAMAN exhibits a strong affinity for disaccharides or trisaccharides containing $\alpha(1,3)$ - or $\alpha(1,6)$ -linked mannosyl residues, but only a weak affinity for the $\alpha(1,2)$ -linked disaccharide; it seems to differ from the above-mentioned Amaryllidaceae lectins chiefly in a higher

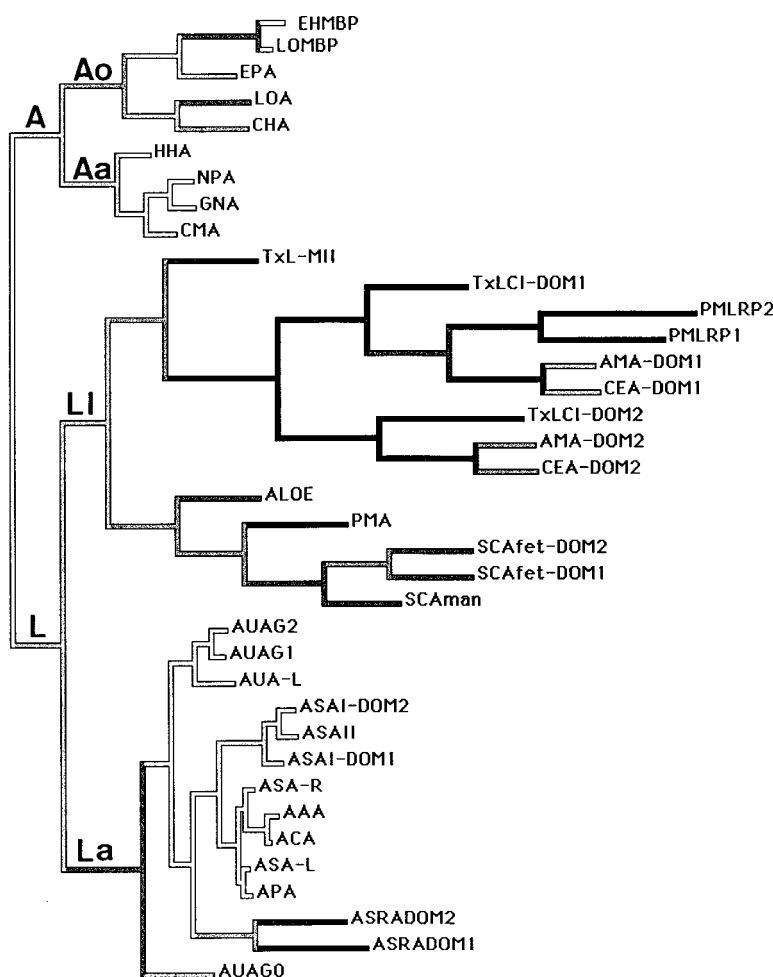


Figure 7 Phylogenetic tree built up from the amino acid sequences of mannose-binding lectins from different monocot families

AAA, *Allium ascalonicum* agglutinin; ACA, *Allium cepa* agglutinin; Aloe, *Aloe* lectin; AMA-DOM1 and AMA-DOM2, domains 1 and 2 of *Arum maculatum* agglutinin; APA, *Allium porrum* agglutinin; ASAI-DOM1 and ASAI-DOM2, domains 1 and 2 of *Allium sativum* agglutinin I; ASAI, *Allium sativum* agglutinin II; ASA-L, *Allium sativum* leaf lectin; ASA-R, *Allium sativum* root lectin; ASRADOM1 and ASRADOM2, domains 1 and 2 of *Allium sativum* lectin-related protein; AUAG0, lectin polypeptide composing *Allium ursinum* agglutinin II; AUAG1 and AUAG2, lectin polypeptides composing *Allium ursinum* agglutinin I; AUA-L, *Allium ursinum* leaf lectin; CEA-DOM1 and CEA-DOM2, domains 1 and 2 of *Colocasia esculenta* agglutinin; CHA, *Cymbidium* hybrid agglutinin; CMA, *Clivia miniata* agglutinin; EHMBP, *Epipactis helleborine* monomeric mannose-binding protein; EPA, *Epipactis helleborine* agglutinin; HHA, *Hippeastrum* hybrid agglutinin; LOA, *Listera ovata* agglutinin; LOMBP, monomeric mannose-binding protein of *Listera ovata*; NPA, *Narcissus pseudonarcissus* agglutinin; PMA, *Polygonatum multiflorum* agglutinin; PMLRP1 and PMLRP2, domains 1 and 2 of *Polygonatum multiflorum* lectin-related protein; TxL-LI-DOM1 and TxL-LI-DOM2, domains 1 and 2 of *Tulipa* lectin TxL-LI; TxL-MII, *Tulipa* sp. lectin MII. Branches of the tree are shaded according to the number of amino acid changes.

Table 2 Monocot mannose-binding lectins and lectin-related proteins composed of two-domain protomers

Abbreviation: n.d., not determined.

Source	Molecular structure	Amino acid sequence identity between the two domains (%)	Specificity of the N-terminal and C-terminal domain	
			N-terminal	C-terminal
Identified lectins				
<i>Allium sativum</i> (ASA-I)	[12 + 12.5 kDa]	85	Mannose	Mannose
<i>Arum maculatum</i> (AMA)	[12 + 12 kDa] ₂	41	n.d.	n.d.
<i>Tulipa</i> sp. (TxLC-I)	[30 kDa] ₄ and [15 + 15 kDa] ₄	20	Mannose	GalNAc
<i>Scilla campanulata</i> (SCAfet)	[30 kDa] ₄	55	n.d.	n.d.
Putative lectins or lectin-related proteins				
<i>Polygonatum multiflorum</i> (PMLRP)	Unknown	27	Unknown	Unknown
<i>Allium sativum</i> (ASRA)	Unknown	45	Unknown	Unknown

affinity for lyxose and a lower affinity for $\alpha(1,2)$ -mannobiose. In spite of its strong similarity (in terms of molecular structure and specificity) to the Amaryllidaceae and Orchidaceae lectins, SCAman also differs with respect to its biological activities. For example, SCAman has little if any antiviral activity against human and animal retroviruses, whereas both the Amaryllidaceae and Orchidaceae lectins are potent antiviral proteins in the same assays.

On the basis of its amino acid sequence, SCAfet also clearly belongs to the family of the monocot mannose-binding lectins. However, in contrast with most other lectins of this family, SCAfet is unable to bind mannose. The inability of SCAfet to interact with simple sugars clearly results from the replacement of several amino acid residues involved in the monosaccharide-binding sites by hydrophobic residues (Leu and Phe). Owing to the presence of these hydrophobic residues, the hydrogen bonds required to anchor simple sugars into the binding site can no longer be formed. Molecular docking of monosaccharides has shown that only 4 out of 24 monosaccharide-binding sites of the SCAfet tetramer (i.e. one per protomer) can accommodate mannose. However, it is uncertain whether these presumptive reactive sites are sufficiently exposed on the surface of the tetramer to interact with mannose. Taken together, these results can explain why SCAfet is unable to interact with simple saccharides.

SCAfet consists of four identical two-domain protomers. Hitherto only three other cases have been reported of monocot mannose-binding lectins composed of either intact or cleaved two-domain protomers (Table 2). The garlic bulb-specific lectin ASA-I is built up of a single completely cleaved protomer consisting of two tandemly arrayed domains that share 85% sequence identity; both domains exhibit mannose-binding activity [9]. AMA, the tuber lectin from *Arum maculatum*, also consists of completely cleaved protomers but occurs as a dimer of two such protomers. The two tandemly arrayed domains of AMA share only 41% sequence identity and most probably exhibit different specificities [10]. Native TxLC-I, a typical bulb protein of the tulip, is a tetramer of four two-domain protomers that are processed only partly, giving rise to both cleaved and uncleaved protomers. The N-terminal and C-terminal domains share only 20% sequence identity and bind to mannose and GalNAc respectively [11]. As result, TxLC-I recognizes two structurally different sugars and is therefore considered to be a so-called superlectin [1]. In addition to ASA-I, AMA and TxLC-

I, cDNA clones have been identified in roots of garlic and rhizomes of Solomon's seal (*Polygonatum multiflorum*) encoding proteins consisting of similar two-domain protomers. However, because the corresponding proteins have not yet been identified it remains to be demonstrated that the presumed *Polygonatum multiflorum* lectin-related protein [31] and *Allium sativum* lectin-related protein [36] are expressed and, if so, whether they have lectin activity. The identification of SCAfet as a fourth example of a monocot mannose-binding lectin consisting of two-domain protomers is important for several reasons: first, it indicates that two-domain protomers are probably more widespread than has been believed; secondly, SCAfet provides valuable additional information about the specificity of the two-domain monocot mannose-binding lectins because it definitely differs from ASA-I, AMA and TxLC-I with respect to its sugar specificity; and thirdly, the sequence of SCAfet gives additional clues to the molecular evolution of the superfamily of monocot mannose-binding lectins. As shown in Figure 7, the dendrogram of the currently known monocot mannose-binding lectins consists of an A and an L branch. Branch A further bifurcates into the side branches Ao and Aa, clustering the Orchidaceae and Amaryllidaceae lectins respectively. Branch L also bifurcates in an La side branch, which clusters all Alliaceae lectins, and an Ll side branch, which comprises all known Liliaceae and Araceae monocot mannose-binding lectins and lectin-related proteins. The overall topology of the Ll side branch is clearly more complex than that of the La, Ao and Aa side branches. This apparent complexity not only is due to the fact that the Ll side branch comprises two different plant families but also reflects the existing heterogeneity of the Liliaceae family. Within the Ll side branch both *S. campanulata* lectins form a small cluster together with the one-domain lectins from the Liliaceae species *Polygonatum multiflorum* and *Aloe arborescens*, which is in good agreement with the close taxonomic relationships between bluebell, Solomon's seal and aloe. The topology of the dendrogram not only demonstrates that both bluebell lectins are closely related evolutionarily but also suggests that the SCAfet gene arose from a recent (in evolutionary terms) self-duplication and in-tandem insertion of a domain similar to SCAman. It is interesting to note that the Ll side branch comprises, besides SCAfet, three other two-domain lectins or lectin-related proteins, namely TxLC-I, the Araceae lectins and *Polygonatum multiflorum* lectin-related protein. The topology of the cluster comprising these three proteins indicates that TxLC-I and the Araceae

lectins have a common two-domain ancestor because the N- and C-terminal domains of the respective lectins form two subclusters. It has been speculated that this presumed two-domain ancestor arose from the self-duplication/in-tandem insertion of an ancestral single-domain lectin in an evolutionary event that took place before the Liliaceae and Araceae families diverged [1]. According to this scheme, self-duplication/in-tandem insertion events have taken place at least three times in the Ll side branch. In side branch La also, two independent self-duplication/in-tandem insertion events are believed to have given rise to the two-domain garlic lectin ASA-I and the presumed *Allium sativum* lectin-related protein. In summary, self-duplication/in-tandem insertion of a single lectin domain had an important role in the evolution of the monocot mannose-binding lectins. Hitherto, two-domain lectins or lectin-related proteins have not yet been found in the Amaryllidaceae or Orchidaceae families. However, this does not preclude the possibility that also within the A branch as-yet undiscovered self-duplication/in-tandem insertion events have occurred.

We thank Professor Jan Balzarini (Rega Institute for Medical Research, Katholieke Universiteit Leuven, Leuven, Belgium) for testing the antiviral activity of the bluebell lectins. This work was supported in part by grants from the Katholieke Universiteit Leuven (OT/94/17 and OT/98/17), the CNRS and the Conseil Régional de Midi-Pyrénées (A.B., P.R.), and the Fund for Scientific Research Flanders (grant G.0223.97). W.P. is Research Director and E.V.D. a Postdoctoral Fellow of this fund. We acknowledge grant 7.0047.90 from the Nationaal Fonds voor Wetenschappelijk Onderzoek-Levenslijn fund. C.R. acknowledges support from the Leverhulme Trust (U.K. grant F/754/A) and the Mitzutani Foundation for Glycoscience, Japan.

REFERENCES

- Van Damme, E. J. M., Peumans, W. J., Barre, A. and Rougé, P. (1998) *Crit. Rev. Plant Sci.* **17**, 575–692
- Van Damme, E. J. M., Allen, A. K. and Peumans, W. J. (1987) *FEBS Lett.* **215**, 140–144
- Van Damme, E. J. M., Smeets, K. and Peumans, W. J. (1995) in *Lectins, Biomedical Perspectives* (Pusztai, A. and Bardocz, S., eds.), pp. 59–80, Taylor and Francis, London
- Barre, A., Van Damme, E. J. M., Peumans, W. J. and Rougé, P. (1996) *Plant Physiol.* **112**, 1531–1540
- Balzarini, J., Schols, D., Neyts, J., Van Damme, E., Peumans, W. and De Clercq, E. (1991) *Antimicrob. Agents Chemother.* **35**, 410–416
- Balzarini, J., Neyts, J., Schols, D., Hosoya, M., Van Damme, E., Peumans, W. and De Clercq, E. (1992) *Antivir. Res.* **18**, 191–207
- Hester, G., Kaku, H., Goldstein, I. J. and Wright, C. S. (1995) *Nat. Struct. Biol.* **2**, 472–479
- Chantalat, L., Wood, S. D., Rizkallah, P. J. and Reynolds, C. D. (1996) *Acta Crystallogr.* **D52**, 1146–1152
- Van Damme, E. J. M., Smeets, K., Torreken, S., Van Leuven, F., Goldstein, I. J. and Peumans, W. J. (1992) *Eur. J. Biochem.* **206**, 413–420
- Van Damme, E. J. M., Goossens, K., Smeets, K., Van Leuven, F., Verhaert, P. and Peumans, W. J. (1995) *Plant Physiol.* **107**, 1147–1158
- Van Damme, E. J. M., Briké, F., Winter, H. C., Van Leuven, F., Goldstein, I. J. and Peumans, W. J. (1996) *Eur. J. Biochem.* **236**, 419–427
- Wood, S. D., Allen, A. K., Wright, L. M. and Reynolds, C. D. (1996) in *Lectins: Biology, Biochemistry and Clinical Biochemistry*, vol. 11 (Van Driessche, E., Rougé, P., Beeckmans, S. and Bog-Hansen, T. C., eds.), pp. 86–90, Textop, Hellrup, Denmark
- Wright, L. M., Wood, S. D., Reynolds, C. D., Rizkallah, P. J., Peumans, W. J., Van Damme, E. J. M. and Allen, A. K. (1996) *Acta Crystallogr.* **D52**, 1021–1023
- Wright, L. M., Wood, S. D., Reynolds, C. D., Rizkallah, P. J. and Allen, A. K. (1997) *Protein Peptide Lett.* **4**, 343–348
- Wright, L. M. (1998) Ph.D. thesis, Liverpool John Moores University
- Allen, A. K. and Neuberger, A. (1975) *FEBS Lett.* **60**, 76–80
- Allen, A. K., Ellis, J. and Rivett, D. E. (1991) *Biochim. Biophys. Acta* **1074**, 331–334
- Allen, A. K. (1979) *Biochem. J.* **183**, 133–137
- Balzarini, J., Naesens, L., Herdewijn, P., Rosenberg, I., Holy, A., Pauwels, R., Baba, M., Johns, D. G. and De Clercq, E. (1989) *Proc. Natl. Acad. Sci. U.S.A.* **86**, 332–336
- Balzarini, J., Naesens, L., Slachmuylders, J., Niphuis, H., Rosenberg, I., Holy, A., Schellekens, H. and De Clercq, E. (1991) *AIDS* **5**, 21–28
- Laemmli, U.K. (1970) *Nature (London)* **227**, 680–685
- Van Damme, E. J. M. and Peumans, W. J. (1993) in *Lectins and Glycobiology* (Gabius, H.-J. and Gabius, S., eds.), pp. 458–468, Springer-Verlag, Berlin
- Mierendorf, R. C. and Pfeiffer, D. (1987) *Methods Enzymol.* **152**, 556–562
- Sanger, F., Nicklen, S. and Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U.S.A.* **74**, 5463–5467
- Maniatis, T., Fritsch, E. F. and Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
- Maddison, W. P. and Maddison, D. R. (1992) *MacClade: Analysis of Phylogeny and Character Evolution*, version 3.0, Sinauer Associates, Sunderland, MA
- Gaboriaud, C., Bissery, V., Benchetrit, T. and Mornon, J. P. (1987) *FEBS Lett.* **224**, 149–155
- Lemesle-Varloot, L., Henrissat, B., Gaboriaud, C., Bissery, V., Morgat, A. and Mornon, J. P. (1990) *Biochimie* **72**, 555–574
- Wood, S. D., Wright, L. M., Reynolds, C. D., Rizkallah, P. J., Allen, A. K., Peumans, W. J. and Van Damme, E. J. M. (1999) *Acta Crystallogr. Sect. D*, in the press
- Kraulis, P. J. (1991) *J. Appl. Crystallogr.* **24**, 946–950
- Von Heijne, G. (1986) *Nucleic Acids Res.* **11**, 4683–4690
- Wright, L. M., Wood, S. D., Reynolds, C. D., Rizkallah, P. J. and Allen, A. K. (1998) *Acta Crystallogr.* **D54**, 90–92
- Van Damme, E. J. M., Barre, A., Rougé, P., Van Leuven, F., Balzarini, J. and Peumans, W. J. (1996) *Plant Mol. Biol.* **31**, 657–672
- Koike, T., Titani, K., Suzuki, M., Beppu, H., Kuzuya, H., Maruta, K., Shimpo, K. and Fujita, K. (1995) *Biochem. Biophys. Res. Commun.* **214**, 163–170
- Shibuya, N., Goldstein, I. J., Van Damme, E. J. M. and Peumans, W. J. (1988) *J. Biol. Chem.* **263**, 728–734
- Kaku, H., Van Damme, E. J. M., Peumans, W. J. and Goldstein, I. J. (1990) *Arch. Biochem. Biophys.* **279**, 298–304
- Smeets, K., Van Damme, E. J. M., Verhaert, P., Barre, A., Rougé, P., Van Leuven, F. and Peumans, W. J. (1997) *Plant Mol. Biol.* **33**, 223–234

Received 9 December 1998/17 February 1999; accepted 10 March 1999