Cloning and characterization of the lectin cDNA clones from onion, shallot and leek

Els J.M. Van Damme[1], Koen Smeets[1], Iris Engelborghs[1], Helen Aelbers[1], Jan Balzarini[2], Arpad Pusztai[3], Fred van Leuven[4], Irwin J. Goldstein[5] and Willy J. Peumans[1]
[1] *Laboratory for Phytopathology and Plant Protection, Katholieke Universiteit Leuven, Willem de Croylaan 42, 3000 Leuven;* [2] *Laboratory of Experimental Chemotherapy, Rega Institute for Medical Research, Katholieke Universiteit Leuven, Belgium;* [3] *The Rowette Research Institute, Aberdeen, UK;* [4] *Centre for Human Genetics, Katholieke Universiteit Leuven, Belgium;* [5] *Department of Biological Chemistry, University of Michigan, Ann Arbor, MI, USA*

Abstract

Characterization of the lectins from onion (*Allium cepa*), shallot (*A. ascalonicum*) and leek (*A. porrum*) has shown that these lectins differ from previously isolated Alliaceae lectins not only in their molecular structure but also in their ability to inhibit retrovirus infection of target cells.

cDNA libraries constructed from poly(A)-rich RNA isolated from young shoots of onion, shallot and leek were screened for lectin cDNA clones using colony hybridization. Sequence analysis of the lectin cDNA clones from these three species revealed a high degree of sequence similarity both at the nucleotide and at the amino acid level.

Apparently the onion, shallot and leek lectins are translated from mRNAs of ca. 800 nucleotides. The primary translation products are preproteins (ca. 19 kDa) which are converted into the mature lectin polypeptides (12.5–13 kDa) after post-translational modifications.

Southern blot analysis of genomic DNA has shown that the lectins are most probably encoded by a family of closely related genes which is in good agreement with the sequence heterogeneity found between different lectin cDNA clones of one species.

Introduction

Plant lectins are a very heterogenous group of (glyco)proteins classified together on the basis of a single common property, namely their ability to recognize and bind specifically to carbohydrate ligands [10]. Since the discovery of the first phytohaemagglutinen by Stillmark in 1888, several hundred plant lectins have been isolated and characterized in some detail with respect to their

---

molecular structure, biochemical properties and carbohydrate-binding specificity. Although in the past most interest in plant lectins has focused on lectins in dry seeds, especially from leguminous species, evidence has accumulated now that lectins also occur in vegetative tissues and are widespread in a large number of plant families belonging to all major taxonomical groupings.

A survey of lectin research in monocotyledonous species shows that studies in this field have long been restricted to Gramineae species. However, during the past five years it has been established clearly that monocot species outside the grass family also are a rich source of new plant lectins with some interesting properties. At present plant lectins have been reported in the families Gramineae [for review see 24], Liliaceae [5, 16, 17], Orchidaceae [25], Amaryllidaceae [26], Alliaceae [27] and Araceae [20] of the class Liliatae. The observation that agglutinins also occur in numerous plant families of the Monocotyledonae acknowledges once more that lectins are widespread indeed within the plant kingdom. Furthermore, characterization of some of these lectins has demonstrated that they differ from dicotyledonous lectins and exhibit several unique properties especially with respect to their carbohydrate-binding specificity. For instance, Amaryllidaceae and Alliaceae species contain reasonable amounts of lectins with an exclusive specificity towards mannose, which makes these lectins quite distinct from the mannose/glucose-binding lectins isolated from leguminous species [10]. Indeed, unlike concanavalin A, pea or lentil lectin, the Ameryllidaceae and Alliaceae lectins do not recognize glucose or N-acetylglucosamine. Hence their unique carbohydrate-binding properties makes the Amaryllidaceae and Alliaceae lectins very useful tools in biomedical and glycoconjugate research [11, 22].

Recently we described the characterization and molecular cloning of two different mannose-binding lectins from young garlic (*Allium sativum*) bulbs [29]. The first lectin, referred to as ASAI, is a dimer composed of 11.5 and 12.5 kDa subunits. It is apparently translated from an mRNA of 1400 nucleotides that encodes a polypeptide of 306 amino acids with two very homologous domains. N-terminal sequencing of the two polypeptides of the lectin confirmed that both subunits are derived from the same precursor and that each of them corresponds to one of the two domains of the sequence. In contrast to ASAI, the second garlic lectin, referred to as ASAII, is composed of two identical 12 kDa subunits. It is translated from an mRNA of 800 nucleotides encoding a polypeptide of 154 amino acids which, in contrast to ASAI, contains only one domain. Preliminary studies indicate that the post-translational modifications of the primary ASAII translation products are very similar to those described for the *Galanthus nivalis* lectin [28].

This paper deals with the molecular cloning of the mannose-specific lectins from onion (*Allium cepa*), shallot (*A. ascalonicum*) and leek (*A. porrum*) species from the family Alliaceae. In contrast to the previously described Alliaceae lectins from garlic (*A. sativum*) cloves [29], which contain two very abundant mannose-binding lectins, the bulbs and shoots of onion, shallot and leek contain only one type of lectin which is present at a very low level. A detailed study of the anti-retroviral activities and molecular structure, and cDNA cloning of these lectins has shown that the onion, shallot and leek lectins clearly differ from the garlic lectins and that different classes of lectins can be distinguished within the Alliaceae species.

Materials and methods

Plant material

Bulbs of onion (*Allium cepa* L.) and shallot (*A. ascalonicum* L.), and leek (*A. porrum* L.) plants were purchased from a local store.

Viruses

Human immunodeficiency virus type 1 (HIV-1) (strain HTLV-III$_B$) was originally obtained from the culture supernatant of the persistently HIV-infected H9 cell line (H9/HTLV-III$_B$) [18] and was kindly provided by Dr R.C. Gallo (National

Cancer Institute, National Institutes of Health, Bethesda, MD). HIV-2 (strain LAV-2) [6] was a gift from Dr L. Montagnier (Pasteur Institute, Paris, France).

Antiretrovirus assays

The methodology of the anti-HIV assays has been described previously [1, 2]. Briefly, MT4 cells $(4.5 \times 10^5$ cells/ml) were suspended in fresh culture medium and infected with HIV-1 and HIV-2 at 100 $CCID_{50}$ (cell culture infection dose 50%) per ml cell suspension (1 $CCID_{50}$ being the dose infective for 50% of the cell cultures). Then, 100 μl of the infected cell suspension was transferred to microplate wells, mixed with 100 μl of the appropriate dilutions of the test compounds, and further incubated at 37 °C. After 5 days, the number of viable cells was determined in a blood cell-counting chamber by Trypan blue staining. Antiviral activity was expressed as EC_{50} being the compound concentration required to inhibit HIV-induced cytopathicity by 50%. Cytotoxic activity was expressed as CC_{50}, being the compound concentration required to reduce MT-4 cell viability by 50%.

Extraction and isolation of the lectins.

The onion, shallot and leek lectins were isolated from extracts of bulbs and/or leaves by affinity chromatography on immobilized mannose as reported previously [27]. To ensure complete purity of the proteins the lectin was subsequently loaded on an anion exchange column of Q Fast Flow (Pharmacia, Uppsala, Sweden) and eluted using a salt gradient.

Sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE)

Lectin preparations were analysed by SDS-PAGE using 12.5–25% (w/v) acrylamide gradient gels as described by Laemmli [13].

Gel filtration

Gel filtration of purified Alliaceae lectins was done on a Pharmacia Superose 12 column equilibrated with phosphate-buffered saline containing 0.2 M mannose to avoid binding of the lectin to the column. About 250 μg of the lectin was loaded on the column. The flow rate was 20 ml/h. Molecular mass reference markers were: bovine serum albumin (68 kDa), ovalbumin (45 kDa), chymotrypsinogen (25 kDa) and cytochrome c (12.5 kDa).

Ultracentrifugation

Equilibrium sedimentation experiments were performed as described by Pusztai and Stewart [19].

Amino acid sequence analysis

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

RNA isolation

Total cellular RNA was prepared from plant material stored at −80 °C essentially as described by Finkelstein and Crouch [9]. Poly(A)-rich RNA was enriched by chromatography on oligodeoxythymidine cellulose (Sigma, St Louis, MO) as described by Siflow et al. [23] except that poly(A)-rich RNA was eluted at room temperature.

Construction and screening of cDNA library

cDNA libraries were constructed from poly(A)-rich mRNA isolated from young shoots of onion, shallot and leek using the cDNA synthesis kit from Pharmacia (Uppsala, Sweden). cDNA fragments were inserted into the Eco RI site of the

multifunctional phagemid pT_7T_3 18U (Pharmacia, Uppsala, Sweden). The library was propagated in *Escherichia coli* XL1 Blue (Stratagene, La Jolla, CA).

Recombinant lectin clones were screened using either the random-primer-labelled garlic lectin cDNA clone [28] or a ^{32}P-end-labelled degenerate oligonucleotide probe (17-mer, 5'-AAT/CGTNGTNATT/C/ATAT/CGG-3') derived from residues 94–99 of a partially known amino acid sequence for the shallot lectin (I.J. Goldstein, unpublished results) as a probe. Hybridization using the synthetic oligonucleotide was carried out overnight at 37 °C as described by Van Damme *et al.* [28] whereas hybridization with the garlic lectin cDNA clone was done overnight at 50 °C [29]. Colonies that produced positive signals were selected and rescreened at low density using the same conditions. Plasmids were isolated from purified single colonies on a miniprep scale using the alkaline lysis method as described by Mierendorf and Pfeffer [15] and sequenced by the dideoxy method [21].

Northern blot

RNA electrophoresis was performed according to Maniatis *et al.* [14]. ca. 3 μg of poly(A)-rich RNA was denatured in glyoxal and dimethylsulphoxide and separated in a 1.2% (w/v) agarose gel. Following electrophoresis the RNA was transferred to Immobilon N membranes (Millipore, Bedford, USA) and the blot hybridized using a random-primer-labelled lectin cDNA insert. Hybridization was performed as reported by Van Damme *et al.* [29]. An RNA ladder (0.16–1.77 kb) was used as a marker.

Genomic DNA analysis

Total DNA from garlic was isolated according to the procedures described by Dellaporta *et al.* [8] and de Kochko and Hamon [7]. The DNA preparation was treated with RNAse in order to remove any contaminating RNA. Ca. 30 μg of

DNA was digested with restriction endonucleases and subjected to electrophoresis in a 0.8% agarose gel. DNA was transferred to Immobilon N membranes (Millipore, Bedford, USA) and hybridized using the ^{32}p-labelled cDNA insert encoding the lectin.

Computer analyses

DNA sequences were analysed using programs from PC Gene (Intelligenetics, Mountain View, CA) and Genepro (Riverside Scientific, Seattle, USA).

Results

Molecular structure of the onion, leek and shallot lectins

Lectins were isolated from bulbs of onion (*Allium cepa*) and shallot (*A. ascalonicum*), and leaves of leek (*A. porrum*), and characterized in some detail with respect to their most important biochemical and physicochemical properties. Unlike garlic (*A. sativum*) [29], onion, shallot and leek contain only one type of lectin. Moreover, although onion, shallot and leek lectins also exhibit a carbohydrate-binding specificity towards mannose they definitely differ from the garlic lectins in their mo-

Table 1. Molecular structure of different Alliaceae lectins.

Lectin[1]	Molecular weight (kDa) determined by			
	SDS-PAGE	Gel filtration	Ultracentrifugation	
			M_r^2	$S_{20}w$
AAA	12.5	25	23.6	2.37
ACA	12.5	50	ND[3]	ND
APA	13.0	50	34.4	3.10
ASAI	11.5 + 12.5	25	24.9	2.47
ASAII	12.0	25	25.7	2.53

1 AAA, *A. ascalonicaum* agglutinin; ACA, *A. cepa* agglutinin; APA, *A. porrum* agglutinin; ASA, *A. sativum* agglutinin.
2 Calculated molecular mass.
3 Not determined.

lecular structure. As shown in Table 1, SDS-PAGE of the onion, shallot and leek lectins yielded lectin polypeptides of 12.5–13 kDa. Judging from gel filtration experiments on a Superose 12 column these lectin polypeptides associate into dimers (*A. ascalonicum*) or tetramers (*A. cepa* and *A. porrum*). However, since non-specific interactions with the gel filtration matrix often occur and cannot be abolished completely by the addition of the specific sugar to the running buffer the molecular mass of the different lectins was also determined by ultracentrifugation. As shown in Table 1, these experiments have shown that the leek lectin is a trimer rather than a tetramer.

Anti-retroviral activity of the onion, shallot and leek lectins

Purified lectins from different Alliaceae species were evaluated for their inhibitory effect against HIV-1- and HIV-2-induced cytopathicity in MT4 cells. As shown in Table 2 the lectins from shallot (*A. ascalonicum*, AAA) and leek (*A. porrum*, APA) strongly inhibit the infection of target cells by HIV at concentrations ranging from 0.01 to 0.91 μg/ml whereas the *A. sativum* (ASA) lectins show poor if any antiviral activity. The leek lectin proved to be the most effective inhibitor of HIV infection since it was about 10-fold more antivirally active than the shallot lectin.

Table 2. Anti-retrovirus activity of Alliaceae lectins.

Lectin[1]	EC_{50}[2] (μg/ml)		CC_{50}[3] (μg/ml)
	HIV-1	HIV-2	
APA	0.01 ± 0.005	0.08 ± 0.05	57.4 ± 48.0
AAA	0.28 ± 0.02	0.91 ± 0.33	82.7 ± 9.2
ASAI	47	> 100	> 100
ASAII	> 100	> 100	> 100

[1] AAA, *A. ascalonicaum* agglutinin; APA, *A. porrum* agglutinin; ASA, *A. sativum* agglutinin.
[2] Concentration required to inhibit HIV-1 and HIV-2-induced cytopathicity in MT-4 cells by 50%.
[3] Concentration required to inhibit MT-4 cell viability by 50%.

Molecular cloning of Alliaceae lectins from onion, shallot and leek

In addition to bulbs, low levels of lectins were also detected in the shoots of *A. cepa* (onion), *A. ascalonicum* (shallot) and *A. porrum* (leek). Therefore cDNA libraries were constructed from poly(A)-rich RNA isolated from young developing shoots of the aforementioned species for the reason that these tissues are much easier to work with than the fleshy bulbs of the Alliaceae species which contain a great deal of polysaccharides.

Since N-terminal sequencing of the purified leek lectin and partial sequencing of tryptic peptides of the shallot lectin (I.J. Goldstein, unpublished results) revealed sequence homology to the amino acid sequence of the previously cloned garlic lectin, initial screenings of the different cDNA libraries were done using either a degenerate oligonucleotide probe derived from the amino acid sequence of the shallot lectin or a previously isolated cDNA insert encoding the garlic lectin [29]. It was not surprising that these screenings revealed only a very small number of positive colonies for onion, shallot and leek lectins since the lectin concentration in these species is very low, being only a few mg/kg fresh weight material. Preliminary results from northern blot analyses indicated that the lectins are encoded by mRNAs of ca. 800 nucleotides. Therefore the cDNA libraries were enriched for cDNA fragments of this length, resulting in a 50-fold increase of positive colonies after screening of the libraries.

Multiple lectin cDNA clones were isolated for each species and their sequences analysed. The lectin cDNA sequences all contain an open reading frame encoding a preprolectin which besides the coding sequence of the mature lectin also contains the information for both a signal peptide and a C-terminal peptide which are removed co- and post-translationally, respectively. The signal peptide (24 amino acids) was determined both by application of the rules of von Heijne [31] and by comparison of the deduced amino acid sequence of the lectin cDNA clone to the N-terminal lectin sequence. Since the calculated molecular mass of the lectin polypeptides encoded by the differ-

ent cDNA clones exceeds the molecular mass of the lectin subunits determined by SDS-PAGE by at least 3000 Da, we assume that the onion, shallot and leek lectin cDNA clones like the previously cloned Amaryllidaceae and Alliaceae lectins also contain the information for a C-terminal peptide [28, 29, 30].

As shown in Figs. 1 and 2, the lectin cDNA clone LECACA encoding the *A. cepa* lectin encodes a preproprotein with a calculated molecular mass of 17 536 Da which only contains part of the signal peptide. Co-translational removal of this peptide will result in a 16 046 Da proprotein which contains two putative glycosylation sites in its C-terminal sequence. Using the *A. cepa* lectin cDNA clone as a probe highly homologous lectin cDNA clones were isolated from *A. ascalonicum* (LECAAA) and *A. porrum* (LECAPA). The cDNA clones encoding the shallot and the leek lectins yield 18 883 Da and 19 232 Da precursor polypeptides, respectively, which like the onion lectin precursor contain two potential glycosylation sites in their C-terminal sequence at identical positions. Co-translational removal of the signal peptide of the shallot and leek lectin precursors will result in lectin polypeptides with a molecular mass of 16 243 and 16 157 Da, respectively.

Characterization of cDNA clones encoding different isolectins from leek

A detailed study of different lectin cDNA sequences from leek revealed some striking differences in the deduced amino acid sequences of the different clones (Fig. 3). Although the different leek lectin clones show between 82 and 91% sequence homology for the total coding region they clearly differ from each other at certain positions within their nucleotide and deduced amino acid sequences. Moreover, since some differences in amino acids create different charges along the lectin polypeptide these lectin cDNA clones encode lectin polypeptides with different isoelectric points which can explain the occurrence of multiple lectin isoforms at the molecular level [27].

The deduced amino acid sequences of three lectin cDNA clones from leek are depicted in Fig. 3. The leek clones contain an open reading frame encoding 177 to 185 amino acids with possible initiation codons at positions 2 and 13 of the sequence LECAPA 1. Translation starting with the first methionine which according to the criteria of Kozak [12] is most likely to be used, generates lectin polypeptides of 179 and 180 amino acids encoding preprolectins with a calculated molecular mass of approximately 19.2 kDa. Compared to the lectin clones LECAPA 2 and LECAPA 3, LECAPA 1 contains an additional amino acid at position 11, resulting in a signal peptide of 30 amino acids in stead of 29 amino acids. Although the differences in the amino acid sequences are most prominent in the C-terminal peptide the position of the two potential glycosylation sites has been conserved in all lectin clones.

Sequence homology between lectin cDNA clones from Alliaceae and Amaryllidaceae species

A comparison of the deduced amino acid sequences of the lectin cDNA clones from onion, shallot and leek reveals that the onion and shallot lectin show ca. 95% sequence homology whereas the onion and the leek lectin, and the shallot and the leek lectin show between 83 and 96% and between 80 and 99% sequence identity, respectively, for the total coding sequence.

Sequence alignment of the onion, shallot and leek lectin sequences with the garlic lectins ASAI and ASAII revealed 65 to 75% sequence homology at the amino acid level for the total coding region, whereas only 36 to 47% homology was found with the Amaryllidaceae lectins. However, it is clear from Fig. 2 that some parts of the lectin sequences have been conserved between Amaryllidaceae and Alliaceae species. Unlike in garlic lectin sequences the sequence Thr-Glu, which is the processing site for the C-terminal peptide in Amaryllidaceae lectins [28, 30], is also present in the deduced amino acid sequences of the onion, shallot and leek lectin cDNA clones.

```
LECAPA : AAGAATGGGCCGTACTACTCCATCTCCTAAACTAATTATGAGCATTACCA  50
LECAAA :                C-GC------------------...----G-GG---  35
LECACA :                                                 --  2


LECAPA : CTGTAGCCGCTATTCTAACCATACTGGCATCCACATGCATGGCCAGGAAC  100
LECAAA : --------A-C--C--------TT-------T-------------A---   85
LECACA : --------A-C--C--------TT-------T-------------A---   52


LECAPA : CTACTGACCAACGGTGAAGGACTGTACGCAGGCCAATCCCTGGATGTAGA  150
LECAAA : G--T--GTG---AAC-------------------------A-TC-AG--  135
LECACA : G--T--GTG---AAC-------------------------A-TC-----  102


LECAPA : ACAGTACAAGTTTATAATGCAGGATGACTGCAACCTCGTGTTGTACGAAT  200
LECAAA : --------CT----------------------T--AC----T----  185
LECACA : --------CT----------------------T--AC---------  152


LECAPA : ACAGCACCCCCATCTGGGCCTCAAACACCGGTGTCACCGGCAAAAACGGG  250
LECAAA : ----------------------------G--CA-------T-----T---  235
LECACA : ----------------------------G--C-----------------  202


LECAPA : TGCAGGGCTGTCATGCAGAAGGATGGCAACTTTGTGGTCTACGATGTTAA  300
LECAAA : --------C--G------CCT--------------C------A-------  285
LECACA : --------C--G------GCT--------------C-------------  252


LECAPA : CGGGCGTCCCGTTTGGGCCACTAATAGTGTAAGAGGGAATGGGAACTATA  350
LECAAA : G------G----C-------G---C--CAG--------C----------  385
LECACA : G------G----C-------G---C--CAG--------C----------  302


LECAPA : TCCTGGTGCTTCAGCAGGACAGAAACGTTGTCATTTACGGCTCTGATATT  400
LECAAA : --------------A---------------T--------A---------  385
LECACA : --------------A-------G--------T--------A---------  352


LECAPA : TGGTCTACTGGTACCTACAGAAGAAGTGCTGGTGGACCTGTTGTTATGGC  450
LECAAA : ------------C--A---C-G-A--AA-TG------A-------------  435
LECACA : --------------G-----G-A--AA-TG------A-------------  402


LECAPA : TATGAATGGCACCGTTAATGGAGGCTCTGTGGTCGGACCCGTAATCGTGA  500
LECAAA : ---------T--G--CG----------C-C-A-T-----G----CA----  485
LECACA : ---------T--T--CG---------------T-----G----CA----  452

                                                          *
LECAPA : ATCAGAATGTCACTGCCATCCGAAAGGTTGGTAC..CAG.TGCTGCTTGA  547
LECAAA : ------------------G-----------C-G-AG-T-C----------  535
LECACA : ------------------G-----------C-G-.....------A---  496


LECAPA : GTTTTGAGATGAGTTCCATGCATGTTTAAGTTGAATAATGAATTGTGTGC  597
LECAAA : A---G-----------T---..-T--G-GTA----C---A--------T  583
LECACA : A---G--------------..-T--G-GTA----C---A---------  544


LECAPA : GTGATGGTGATGCACTGTACGCATGTAACTGTTATGCTGTATTATGTATA  647
LECAAA : -AC-GT-ATGCACTCTAT..GTACGTTATTGTCATCTTGTATTATGT    628
LECACA : -ACGGCGATGCACTCTATATGTATGTTATTGTTATCTTCTATTATGTATT  594


LECAPA : TTTGGCACTGTTTGCT                                    663
LECAAA :
LECACA : TTTGCACTGCATGCTTGCTGGTTACATCTGTAATATTCATGTATGTGGCC  644


LECAPA :
LECAAA :
LECACA : TTAAGCGCTTAAGGTTATGATAAGATGACCACCAGCATTCTGGCTC(A)9  699


LECAPA :
LECAAA :
LECACA :
```

Fig. 1. Comparison of DNA sequences of cDNA clones encoding the lectin from leek (LECAPA), shallot (LECAAA) and onion (LECACA). The sequences were aligned to maximize homology. Dashes denote sequence similarity to LECAPA. Dots represent gaps introduced for maximal alignment. Putative start codons and polyadenylation signals are underlined. The stop codon is indicated by an asterisk.

```
A:        MAKTSFLILATIFLGVITLPSCLSDNILYSGDTLSTGQFLSYG    43
B:         -I---V----------.-------------E---P-ES-N--
C:         ---IL-----------.-----N------E---A-E--NQ-

D:        SP-VMSIATVAAI-TILA.ST-MAR---MNDEG-YA--S-DVE
E:        S--LMSIAAVAAI-TILA.ST-MAR-L-TN-EG-YA--S-NVE
F:                           R---RNDEG-YA--S-DVE

G:        STPSP-LMSVATVATI-TILA.ST-MAR-V-VNNEG-YA--S-VEE
H:                   TVATI-TILA.ST-MAR-V-VNNEG-YA--S-VVE
I:  MGRTTPSPKLIMSITTVAAILTILA.ST-MAR-L-TNGEG-YA--S-DVE
```

```
A:  SYVFIMQEDCNLVLYDVDKPIWATNTGGLSS.DCHLSMQTDGNLVVYSPQ    92
B:  P-T----Q---------------S-----AR.----N--S-------A-R
C:  N-----------------L-E-------R.R-Y-N----------N-S

D:  P-HL------------HSTAV-TS--DIPGKKG-KAVL-S---F---DAE
E:  P-H------------HSTA--S---DIPGKKE-RAVL-S---F---DAD
F:  P-H---H--------I-HNTAV--S--DIPGKKG-KAVL-S---F---DAE

G:  Q-T----D-------EYST----S---ITGKNG-RAV--P---F---NVK
H:  Q-T----D-------EYST----S---VTGKNG-RAV--A---F---DVK
I:  Q-K----D-------EYST----S---VTGKNG-RAV--K---F---DVN
```

```
A:  NKAIWASNTDGENGHFVCILQKDRNVVIYGTDRWATGTYT...GAVGIPE   139
B:  -QV------Q----NY---------------A-------NIHGA-I--V-G
C:  --P-------G----NY-------G-IA---PAI-----NIHGA-I--VLG

D:  GRSL---HSVRG--NY-LV--E-G------S-I-S-N--K...-SARG-V
E:  GRPL---HSVRG--NY-LV--A-G------S-I-S----V...K-ARG-V
F:  GRSL---HSVRG--NY-LV--E-G------S-I-S-D--R...KSARG-V

G:  GR-V----SRRG--NYILV----------S-I-S----RKKV-GTVVMA
H:  GR-V----SRRG--NYILV----------S-I-S----RKKV-GTVVMA
I:  GRPV----SVRG--NYILV--Q--------S-I-S----RRSA-GPVVMA
```

```
A:  SPASEKYPTSGKITPTSEKYPTTGKIKLVTGK            171
B:  -APQNT..-AEM-KLVRKYLI-K
C:  -APQNS..-AEM-KLVRKYLITK

D:  IRKVGTLAGAA
E:
F:  IRKVGTLAGAA

G:  MNGTVDGGSAIGPVTVNQNVTAVR-VAAAAAA
H:  MNGTVDGGSVVGPVTVNQNVTAVR-VAAAA
I:  MNGTVNGGSVVGPVIVNQNVTAIR-VGTSAA
```

Fig. 2. Deduced amino acid sequences of lectin cDNA clones from different Amaryllidaceae and Alliaceae species: *Narcissus* cv. Fortune (A), *Galanthus nivalis* (B), *Hippeastrum* hybr. (C), *Allium sativum* lectin II (D), *A. sativum* lectin I domain 1 (E), *A. sativum* lectin I domain 2 (F), *A. ascalonicum* (G), *A. cepa* (H) and *A. porrum* (I). In the case of garlic only the sequence of the two domains of the lectin precursor is shown. The arrowheads indicate the possible processing sites for the cleavage of the signal peptide and the C-terminal peptide, respectively. Dashes denote sequence identity with A. Dots represent gaps introduced for maximal alignment. Putative glycosylation sites are underlined.

```
LECAPA 1:      rMGRTTPSPKLIMSITTVAAILTILASTCMA▼RNLLTNGEGLYAGQS
LECAPA 2: aaakr--G------P.-I---------------- --A-----------
LECAPA 3:      S------.--VA---T----------- --V-V-N--------


LECAPA 1: LDVEQYKFIMQDDCNLVLYEYSTPIWASNTGVTGKNGCRAVMQKDGNFVVY
LECAPA 2: ---------I----------------------------------------
LECAPA 3: -VE---T----------------------I-----------P-------


LECAPA 1: DVNGRPVWATNSVRGNGNYILVLQQDRNVVIYGSDIWSTGTYRRSAGGPVV
LECAPA 2: ---------S---------------E----------------------I-
LECAPA 3: --K--A---S--R-----------K-----------------KKV--T--


LECAPA 1: MAMNGTVNGGSVVGPVIVNQNVTAIRKVGTSAA
LECAPA 2: -------D----I-S-T------V------G--
LECAPA 3: -------D---AI---T-----------AAA--A
```

Fig. 3. Deduced amino acid sequences of different lectin cDNA clones from leek. The arrowhead indicates the possible processing sites for the cleavage of the signal peptide. Dashes denote sequence identity with LECAPA 1. Dots represent gaps introduced for maximal alignment. Putative glycosylation sites are underlined. Since the first ATG codon is probably used as the translation initiation site, the deduced amino acids preceding this methionine are shown in lower-case letters.

Northern blot analysis

To establish the total length of the RNA encoding the onion, shallot and leek lectins a blot containing poly(A)-rich RNA from different Alliaceae species was hybridized with the labelled lectin clone LECACA encoding the onion lectin. As shown in Fig. 4, hybridization of the blot revealed one band of lectin mRNA for onion, shallot and leek shoots. Using RNA markers the length of the RNA encoding the lectins was estimated to be ca. 800 nucleotides. Identical results were obtained when the blot was hybridized using the lectin cDNA clones from shallot and leek.

Southern blot analyses

Since sequence analysis of different lectin cDNA clones of one species revealed minor differences in their sequences Southern blot analysis was performed to establish whether the lectins are encoded by a gene family. As shown in Fig. 5, hybridization of a blot containing fragments of genomic DNA from onion with a labelled lectin cDNA clone reveals a complex pattern of bands

indicating that the lectins are most probably encoded by a small gene family. Similar results were obtained when Southern blots of genomic DNA from shallot and leek were hybridized with the lectin cDNA probe.
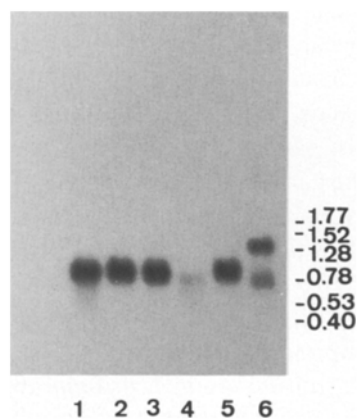

Fig. 4. Northern blot of poly(A)-rich RNA isolated from different Alliaceae species. Poly(A)-rich RNA was isolated from shoots of leek (lane 1), onion (lane 2), shallot (lane 3), ramsons (lane 4) and garlic (lane 5) and young (lane 6) garlic bulbs. The blot was hybridized using the ^{32}P-labelled cDNA insert from LECACA. Numbers on the right show RNA size (kb).
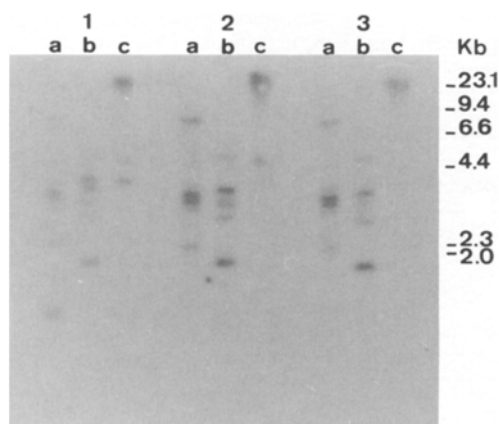
374

Fig. 5. Southern blot of genomic DNA isolated from different onion varieties. DNA was digested with *Eco* RI (a), *Hind* III (b) and *Pst* I (c), and hybridized with the ^{32}P-labelled cDNA insert from LECACA.

Discussion

Although all lectins isolated thus far from Alliaceae species exhibit an exclusive specificity towards mannose they clearly differ from each other with regard to the size and the number of their subunits. Moreover, some species like for example garlic (*Allium sativum*) [29] and ramsons (*A. ursinum*) (W. Peumans, unpublished results) were found to contain two mannose-binding lectins with different molecular structures. In addition, a study of the antiviral activity of the Alliaceae lectins has revealed another striking difference since the lectins from shallot and especially leek like the Amaryllidaceae lectins [3, 4] strongly inhibit the infection of target cells by retroviruses *in vitro* whereas the garlic and the ramsons [4] lectins are completely inactive in this test system. It appears therefore that the Alliaceae lectins can be divided in at least two groups with respect to both their molecular structure and antiviral properties. One group comprises the agglutinins from the garlic species *A. sativum* and *A. ursinum* which both contain two very abundant mannose binding lectins that have no marked antiviral activity when tested against HIV-1 and HIV-2. The second group comprises the agglutinins from species related to onion and leek, namely *A. cepa*, *A. ascalonicum* and *A. porrum*. Plants of these three spe-

cies contain only one lectin which is present in low concentrations but, unlike the garlic lectins, has a pronounced antiviral activity. The apparent differences between the Alliaceae lectins which could be demonstrated by biochemical and functional analysis of the purified proteins could further be evidenced by molecular cloning of their genes.

cDNA libraries constructed from poly(A)-rich RNA isolated from young developing shoots of onion, shallot and leek were screened for lectin cDNA clones and their sequences analysed. Characterization of the lectin cDNA clones revealed that, similarly to the previously cloned mannose-binding lectins from Amaryllidaceae species, the onion, shallot and leek lectins are apparently translated from mRNAs of ca. 800 nucleotides. Furthermore, the lectins are also synthesized as preproproteins which, upon post-translational modifications, are converted into the mature lectin polypeptides. Unlike garlic lectin sequences but similar to the Amaryllidaceae lectins, the onion, shallot and leek lectin sequences contain the sequence Thr-Gly which in Amaryllidaceae species is the cleavage site for the processing of the C-terminal peptide [28, 30].

Sequence comparison for the total coding region of Amaryllidaceae and Alliaceae lectins, shows only 40% overall sequence homology at the amino acid level. However, a detailed study of the sequence similarity between the different lectins reveals that some parts of the sequence are highly conserved between the different species. Furthermore, the position of the two cysteines in Amaryllidaceae lectins which are involved in the internal disulphide bond (amino acids 29 and 52 of the mature lectin polypeptide) is identical in Alliaceae lectins whereas the third (free) cysteine in Amaryllidaceae lectins does not occur in the Alliaceae lectins.

Although Amaryllidaceae and Alliaceae lectins both show a unique carbohydrate-binding specificity towards mannose and at least some of the lectins have similar biological activities, the overall sequence homology in the total coding region of the lectins is rather low, indicating that the lectin-binding site probably does not reside in one

stretch of amino acids buth rather is a three-dimensional cavity built up of amino acid scattered over the lectin polypeptide. Whereas immunodiffusion assays indicated that both onion and leek lectins were more closely related serologically to the Amaryllidaceae lectins than the other Alliaceae lectins isolated [27], a comparison of the nucleotide and amino acid sequences of Amaryllidaceae and Alliaceae lectins does not permit confirmation of this statement since the sequence homology of onion and leek lectins to Amaryllidaceae lectins is not considerably higher than the sequence homology of other Alliaceae lectins with Amaryllidaceae lectin sequences.

Acknowledgements

References

1. Balzarini J, Naessens L, Herdewijn P, Rosenberg I, Holy A, Pauwels R, Baba M, Johns DG, De Clercq E: Marked in vivo anti-retrovirus activity of PMEA [9-(2-phosphonylmethoxyethyl)adenine], a new selective anti-HIV agent. Proc Nat Acad Sci USA 86: 332–336 (1989).
2. Balzarini J, Naesens L, Slachmuylders J, Niphuis H, Rosenberg I, Holy A, Schellekens H, De Clercq E: 9-(2-phosphonylmethoxyethyl)adenine (PMEA) effectively inhibits retrovirus replication in vitro and simian immunodeficiency virus infection in Rhesus monkeys. AIDS 5: 21–28 (1991).
3. Balzarini J, Schols D, Neyts J, Van Damme E, Peumans W, De Clercq E: α-(1-3)- and α-(1-6)-D-mannose-specific plant lectins are markedly inhibitory to human immunodeficiency virus and cytomegalovirus infections in vitro. Antimicrob Agents Chemother 35: 410–416 (1991).
4. Balzarini J, Neyts J, Schols D, Hosoya M, Van Damme E, Peumans W, De Clercq E: The mannose-specific plant lectins from Cymbidium hybrid and Epipactis helleborine and the (N-acetylglucosamine)ₙ-specific plant lectin from Urtica dioica are potent and selective inhibitors of human immunodeficiency virus and cytomegalovirus replication in vitro. Antiviral Res 18: 191–207 (1992).
5. Cammue BPA, Peeters B, Peumans WJ: A new lectin from tulip (Tulipa) bulbs. Planta 169: 583–588 (1986).
6. Clavel F, Guétard D, Brun-Vézinet F, Chamaret S, Rey M-A, Santos-Ferreira MO, Laurent AG, Dauget C, Katlama C, Rouzioux C, Klatzmann D, Champalimaud JL, Montagnier L: Isolation of a new retrovirus from west-African patients with AIDS. Science 233: 343–346 (1986).
7. de Kochko A, Hamon S: A rapid and efficient method for the isolation of restrictable total DNA from plants of the genus Abelmoschus. Plant Mol Biol Rep 8: 3–7 (1990).
8. Dellaporta SL, Wood J, Hicks JB: A plant DNA minipreparation: version II. Plant Mol Biol Rep 1: 19–21 (1983).
9. Finkelstein RR, Crouch ML: Rapeseed embryo development in culture on high osmoticum is similar to that in seeds. Plant Physiol 81: 907–912 (1986).
10. Goldstein IJ, Poretz RD: Isolation, physicochemical characterization and carbohydrate-binding specificity of lectins. In: Liener IE, Sharon N, Goldstein IJ (eds) The Lectins: Properties, Functions and Applications in Biology and Medicine, pp. 33–248. Academic Press, New York (1986).
11. Haselbeck A, Schickaneder E, von der Eltz H, Hösel W: Structural characterization of glycoprotein carbohydrate chains by using digoxigenin-labelled lectins in blots. Anal Biochem 191: 25–30 (1990).
12. Kozak M: Possible role of flanking nucleotides in recognition of the AUG initiator codon by eukaryotic ribosomes. Nucl Acids Res 9: 5233–5252 (1981).
13. Laemmli UK: Cleavage of structural proteins during the assembly of the head of bacteriophage T₄. Nature 227: 680–685 (1970).
14. Maniatis T, Fritsch EF, Sambrook J: Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, (1982).
15. Mierendorf RC, Pfeffer D: Direct sequencing of denatured plasmid DNA. Meth Enzymol 152: 556–562 (1987).
16. Oda Y, Minami K: Isolation and characterization of a lectin from tulip bulbs, Tulipa gesneriana. Eur J Biochem 159: 239–245 (1986).

17. Peumans WJ, Allen AK, Cammue BPA: A new lectin from meadow saffron (Colchicum autumnale) Plant Physiol 82: 1036–1039 (1986).

18. Popovic M, Sarngadharan MG, Read E, Gallo RC: Detection, isolation, and continuous production of cytopathic retroviruses (HTLV-III) from patients with AIDS and pre-AIDS. Science 224: 497–500 (1984).

19. Pusztai A, Stewart JC: Molecular size, subunit structure and microheterogeneity of glycoprotein II from the seeds of kidney bean (Phaseolus vulgaris L.). Biochim Biophys Acta 623: 418–428 (1980).

20. Sandhu RD, Arora JS, Chopra SK, Pelia SS, Kamboy SS, Naidu YC, Nath I: New sources of lectins from Araceous indian plants. In: Kocourek J, Freed DLJ (eds) Lectins: Biology, Biochemistry, Clinical Biochemistry, Vol. 7, pp. 19–25. Sigma Chemical Co., St. Louis, MD (1990).

21. Sanger F, Nicklen S, Coulson AR: DNA sequencing with chain terminating inhibitors. Proc Natl Acad Sci USA 74: 5463–5467 (1977).

22. Shibuya N, Berry JE, Goldstein IJ: One-step purification of murine IgM and human α_2-macroglobulin by affinity chromatography on immobilized snowdrop bulb lectin. Arch Biochem Biophys 267: 676–680 (1988).

23. Siflow CD, Hammett JR, Key JL: Sequence complexity of polyadenylated ribonucleic acid from soybean suspension culture cells. Biochemistry 189: 2725–2731 (1979).

24. Stinissen HM, Peumans WJ: Recent advances in biochemistry, cell biology, physiology, biosynthesis and genetics of Gramineae lectins. Biochem Physiol Pflanz 180: 85–106 (1985).

25. Van Damme EJM, Allen AK, Peumans WJ: Leaves of the orchid twayblade (Listera ovata) contain a mannose-specific lectin. Plant Physiol. 85: 566–569 (1987).

26. Van Damme EJM, Allen AK, Peumans WJ: Related mannose-specific lectins from different species of the family Amarylidaceae. Physiol Plant 73: 52–57 (1988).

27. Van Damme EJM, Goldstein IJ, Peumans WJ: Comparative study of related mannose-binding lectins from Amaryllidaceae and Alliaceae species. Phytochemistry 30: 509–514 (1991).

28. Van Damme EJM, Kaku H, Perini F, Goldstein IJ, Peeters B, Yagi F, Decock B, Peumans WJ: Biosynthesis, primary structure and molecular cloning of snowdrop (Galanthus nivalis L.) lectin. Eur J Biochem 202: 23–30 (1991).

29. Van Damme EJM, Smeets K, Torrekens S, Van Leuven F, Goldstein IJ, Peumans WJ: The closely related homomeric and heterodimeric mannose-binding lectins from garlic are encoded by one-domain and two-domain lectin genes, respectively. Eur J Biochem 206: 413–420 (1992).

30. Van Damme EJM, Goldstein IJ, Vercammen G, Vuylsteke J, Peumans WJ: Lectins of members of the Amaryllidaceae are encoded by multigene families which show extensive homology. Physiol Plant 86: 245–252 (1992).

31. von Heijne G: A method for predicting signal sequence cleavage sites. Nucl Acids Res 11: 4683–4690 (1986).