The tomato lectin consists of two homologous chitin-binding modules separated by an extensin-like linker

Willy J. PEUMANS*, Pierre ROUGÉ† and Els J. M. VAN DAMME*1

*Department of Molecular Biotechnology, Ghent University, Coupure Links 653, 9000 Gent, Belgium, and †Signaux et Messages Cellulaires chez les Végétaux, UMR CNRS-UPS 5546, Pôle de Biotechnologies Végétales, 24 chemin de Borde Rouge, B.P. 17, 31326 Castanet-Tolosan, France

A cDNA encoding a putative lectin expressed in tomato leaves was identified and analysed. The lectin consists of two homologous chitin-binding modules interconnected by a short proline-rich domain containing a single Ser[Pro]$_n$ repetitive motif. Each module comprises two in-tandem-arrayed hevein domains separated by a tetrapeptide linker. Besides the chitin-binding modules and proline-rich domain, the lectin contains two short unrelated domains located at the N- and C-termini of the protein respectively. Eventual elucidation of the molecular structure of the tomato lectin confirms the presumed chimaeric nature of the Solanaceae lectins but also indicates that all previously proposed models need to be revised.

Key words: cDNA, chimaeric protein, extensin, lectin, Solanaceae, tomato.

INTRODUCTION

Several species of the plant family Solanaceae accumulate lectins in seeds and/or vegetative tissues [1]. Detailed biochemical analyses indicated that the lectins from potato (*Solanum tuberosum*) tubers [2], tomato (*Lycopersicon esculentum*) fruit [3] and thornapple (*Datura stramonium*) seeds [4] are structurally and functionally similar. Moreover, these analyses also revealed that the so-called Solanaceae lectins distinguish themselves from all other plant lectins by their very high content (up to 30 %) of hydroxyproline and covalently bound galactose and arabinose. Since the high contents of hydroxyproline and arabinose/galactose were reminiscent of the hydroxyproline-rich glycoproteins, it had already been suggested after the first analyses that the Solanaceae lectins may share some structural features with some hydroxyproline-rich glycoproteins from the plant cell wall [2,3]. Detailed amino acid and carbohydrate analyses of native and deglycosylated lectins, and proteolytic fragments derived thereof, eventually led to the development of the concept that the Solanaceae lectins are chimaeric proteins comprising two structurally and functionally different domains, namely a cysteine-rich chitin-binding domain resembling wheat-germ agglutinin and an hydroxyproline-rich glycoprotein domain resembling the cell-wall glycoprotein extensin [5]. Although there is little doubt that the Solanaceae lectins comprise multiple CBMs (chitin-binding modules) that are structurally related to hevein, the exact molecular structure of these presumed chimaeric proteins remains enigmatic because both the number and arrangement of the hevein domains as well as the structure and location of the putative extensin domain are still unclear. In the past, several hypotheses have been put forward with respect to the overall structure of the protomer of the Solanaceae lectins. According to a first model, the potato lectin protomer comprises a cysteine-rich lectin domain comprising four hevein domains and an ELD (extensin-like domain) fused to each other in an unknown order [5]. Partial sequencing confirmed the modular structure of the potato lectin and suggested that the cysteine-rich lectin domain is located at the N-terminus of the protein [6]. However, further studies led to the concept that the potato lectin is a three-domain glycoprotein comprising an N-terminal region that is rich in proline but poor in hydroxyproline, a central domain that is extremely rich in glycosylated hydroxyproline residues and shares sequence similarity with the extensins and a C-terminal cysteine-rich domain comprising two hevein domains [7]. Similar analyses yielded a markedly different model for the tomato lectin. Naito et al. [8] concluded that the genuine tomato lectin is considered to be a closely related orthologue of the potato lectin, based on its amino acid and carbohydrate composition. It is a three-domain protein comprising an N-terminal extensin-like domain, a central cysteine-rich lectin domain resembling wheat-germ agglutinin and a C-terminal glutamine-rich domain that shares sequence similarity with the large subunit of the tomato seed 2 S albumin. According to this model, the tomato lectin not only lacks the first domain of the potato lectin but also contains a C-terminal domain that is apparently absent from the potato lectin. In addition to the genuine lectin, Naito et al. [8] also identified a 42 kDa protein that is structurally related to the genuine lectin but lacks the N-terminal extensin-like domain. At present, no reasonable explanation can be given for the presumed differences in the overall structures of the potato and tomato lectins and the nature of the structural composition of their domains, because the proposed models are based on partial sequences. Therefore the problem of the molecular structure of tomato lectin(s) can be resolved only by determining the complete amino sequence of the proteins or cloning the corresponding genes.

The present paper deals with the cDNA cloning of a putative tomato leaf lectin. It appears that this lectin consists of two similar CBMs, each comprising two contiguous hevein domains, interspersed by a short proline-rich domain containing a single Ser[Pro]$_n$ repetitive motif. The elucidation of this structure confirms the presumed chimaeric nature of the Solanaceae lectins but indicates that the previously proposed model of the molecular structure of the tomato lectin needs to be revised.

Abbreviations used: CBM, chitin-binding module; ELD, extensin-like domain; EST, expressed sequence tag; HCA, hydrophobis cluster analysis; Lycesca, *Lycopersicon esculentum* agglutinin; UDA, *Urtica dioica* agglutinin.

[1] To whom correspondence should be addressed (e-mail ElsJM.VanDamme@UGent.be).

EXPERIMENTAL

Tomato clone

Tomato clone cLET21N20 (GenBank® accession nos. AW092734; EST285914; GI:6058329) was purchased from Clemson University Genomics Institute (Clemson, SC, U.S.A.). The cDNA library cLET was preprared from leaves of 4–6-week-old tomato plants (Cultivar Rio Grande PtoR), inoculated with a variety of disease response elicitors [i.e. plants were exposed to 2,6-dichloroisonicotinic acid, benzo(1,2,3)thiadiazole-7-carbothioic acid *S*-methyl ester, jasmonic acid, ethylene, fenthion, ethylene-inducing xylanase, okadaic acid or systemin before tissue harvest].

Isolation of the tomato leaf and fruit lectin

Lectins were isolated from leaves and fruits of the tomato cultivar 'Moneymaker'. Juice was collected from fully mature fruits (5 kg), filtered through glass wool, adjusted to pH 3.0 with 1 M HCl and centrifuged (8000 ***g***, 10 min). The cleared juice was filtered through Whatman 3 MM filter paper, diluted with 2 vol. of water (adjusted to pH 3.0) and loaded on to a Sepharose Fast Flow anion-exchange column (Amersham Biosciences; 5 cm × 5 cm; approx. 100 ml bed volume), equilibrated with 20 mM acetic acid. After washing the column with 20 mM acetic acid, the bound protein was eluted with 0.1 M Tris (pH 8.7) containing 0.5 M NaCl. The eluate was adjusted to pH 7.0 and 0.2 M NaCl before loading on to a column of immobilized fetuin. After washing the column, the lectin was desorbed using 20 mM acetic acid. The total yield was approx. 2 mg of lectin/kg of fruit.

Detached tomato leaves (2 kg) were treated with jasmonic acid by floating on a 50 μM solution of jasmonate methyl ester for 4 days, washed with distilled water and homogenized in distilled water containing 1 $g \cdot l^{-1}$ ascorbic acid. After centrifugation of the extract at 3000 ***g*** for 10 min, 1 $g \cdot l^{-1}$ calcium chloride was added to the supernatant and the pH adjusted to 9.0. The solution was then centrifuged at 8000 ***g*** for 10 min and the supernatant filtered through Whatman 3 MM filter paper. The filtrate was brought to pH 3.0 with 1 M HCl and subjected to the same procedure as that followed for the isolation of the fruit lectin. The total yield was approx. 0.5 mg of lectin/kg of leaves.

Sequencing of clone cLET21N20

The complete sequence of tomato clone cLET21N20 was analysed on a capillary sequencer (Applied Biosystems 3730 DNA Analyzer) by the VIB Genetic Service Facility Team (Antwerp, Belgium).

Molecular modelling of the hevein domains of Lycesca (*L. esculentum* agglutinin)

Molecular modelling of Lycesca was performed on a Silicon Graphics O2 10 000 workstation, using the programs InsightII, Homology and Discover (Accelrys, San Diego, CA, U.S.A.). The atomic co-ordinates of hevein (code 1HEV) [9] were taken from the RCSB Protein Data Bank and used to build the three-dimensional model of the four hevein-like domains of the lectin. Amino acid sequence alignment was performed with CLUSTAL-X [10] and HCA (hydrophobic cluster analysis) was performed to recognize the structurally conserved regions common to hevein and the hevein domains of the tomato lectin [11]. HCA plots were generated with the program *drawhca* of L. Canard (http://www.lmcp.jussieu.fr/~soyer/www-hca/hca-form.html). Steric conflicts resulting from the replacement or the deletion of some residues in the modelled domains were corrected during the model-building procedure using the rotamer library [12] and the search algorithm implemented in the Homology program [13] to maintain proper side-chain orientation. Energy minimization and relaxation of the loop regions were performed by several cycles of steepest descent and conjugate gradient using Discover 3. After correction of the geometry of the loops using the minimize option program TurboFrodo (Bio-Graphics, Marseille, France), a final energy minimization step was performed by 25 cycles of conjugate gradient using Discover 3, keeping the amino acid residues forming the carbohydrate-binding site of the domains constrained. Cysteine residues involved in disulphide bridges were covalently bonded before energy minimization. The program TurboFrodo was used to draw the Ramachandran plots and perform the superimposition of the models. PROCHECK [14] was used to check the stereochemical quality of the three-dimensional models. Cartoons were drawn with Molscript [15] and rendered with Bobscript [16] and Raster3D [17].

RESULTS AND DISCUSSION

Isolation and purification of tomato fruit and leaf lectins

Until now, all results reported on the tomato lectin were obtained from analyses of lectin preparations isolated from fruits. Since the putative lectin cDNA we retrieved from the EST (expressed sequence tag) database was identified in a cDNA library prepared from leaves of plants inoculated with a variety of disease–response elicitors, it was a prerequisite to check whether the leaf tissue of these plants contains lectin and if so whether the leaf lectin is identical or at least comparable with the fruit lectin. Therefore, detached tomato leaves were treated with jasmonic acid, homogenized and the extract subjected to the same procedure as that followed for the isolation of the fruit lectin. This approach eventually yielded a leaf protein fraction with a high agglutination activity. SDS/PAGE indicated that this leaf lectin yielded a major polypeptide of approx. 90 kDa and a minor band of approx. 45 kDa (results not shown). The 90 kDa polypeptide apparently corresponds to the subunits of the previously described tomato lectin [3,8,18], whereas the 45 kDa polypeptide most probably represents a leaf homologue of the 42 kDa lectin-related protein described by Naito et al. [8]. N-terminal sequencing of the electroblotted polypeptides yielded no signal, indicating that they are both N-terminally blocked. This is in contrast with the findings of Nachbar et al. [3], who reported that the tomato lectin yielded a methionine residue after Edman degradation and Naito et al. [8], who reported an N-terminal sequence of 9 and 29 residues for the genuine lectin and the 42 kDa protein respectively.

Identification of a putative EST encoding the tomato lectin

In a first attempt to retrieve putative EST sequences encoding the tomato lectin, a tBLASTn search was performed using the NCBI database (limited to tomato ESTs) with the available amino acid sequences of the protein as a query. Since no positive hits were found an alternative approach was followed, which was based on the assumption that the tomato lectin polypeptide comprises multiple hevein domains. To this end, BLASTn and tBLASTn searches were performed using the nucleotide and deduced amino acid sequence respectively of the lectin domain (comprising two hevein repeats) of the UDA (*Urtica dioica* agglutinin) as a query. This approach yielded a large number of positive hits but a closer examination of the sequences indicated that they all corresponded to proteins with a single

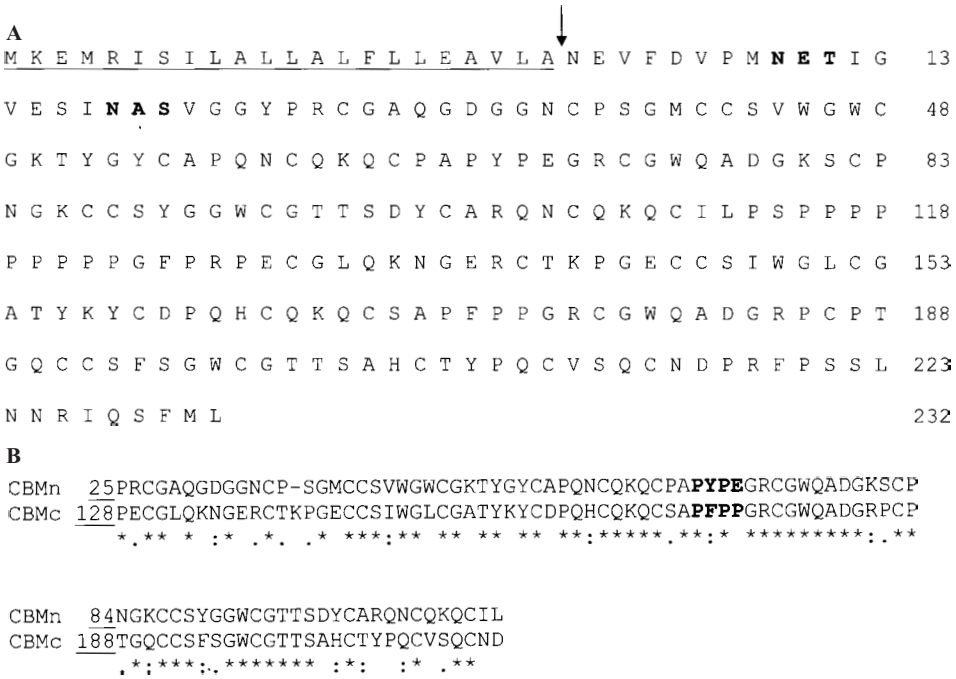

Figure 1 Deduced amino acid sequence of Lycesca

(**A**) Complete deduced amino acid sequence of tomato clone cLET21N20. The primary translation product contains a signal peptide (underlined). The two putative N-glycosylation sites are indicated by boldface letters. (**B**) Sequence alignment of CBMn and CBMc. The linker tetrapeptide between both hevein domains is shown in boldface. Key to symbols: *, identity; :, conservative replacement; ·, non-conservative replacement.

hevein domain (most probably class I chitinases). However, when the expected value was increased, a single EST (tomato clone cLET21N20) was found in which a complete hevein domain could be distinguished that was followed by the N-terminal part of a second hevein domain. A BLAST search of the TIGR Tomato Gene Index database using this clone domain as a query yielded no additional clones. However, a closer examination of the raw sequencing data retrieved from the TIGR database (tomato AW092734) provided additional evidence for the occurrence of two in-tandem-arrayed hevein domains in this EST and also indicated that the second hevein domain was followed by a proline-rich sequence reminiscent to the presumed extensin domain of the Solanaceae lectins. For these reasons, clone cLET21N20 was considered a good candidate lectin cDNA and subjected to a detailed analysis.

Complete sequence of tomato clone cLET21N20 and analysis of the deduced amino acid sequence of the putative lectin polypeptide

Sequence analysis of cLET21N20 revealed an open reading frame of 762 bp. Translation starting with the first initiation codon yields a 254-amino-acid polypeptide (Figure 1A). Analysis of the deduced sequence indicated that the primary translation product contains a signal peptide (residues −22 to −1). Co-translational removal of this signal peptide yields a 232-amino-acid polypeptide. This putative lectin polypeptide will hereafter be referred to as Lycesca. Calculations based on the deduced sequence of Lycesca yielded a molecular mass of 24.8971 kDa, a theoretical pI of 8.02 and an absorption A_{280} 2.062 for a 1 mg · ml^{-1} solution (i.e. for the non-glycosylated protein). The amino acid composition calculated from the deduced sequence is summarized in Table 1.

A closer examination of the deduced sequence reveals that Lycesca exhibits a modular overall structure. The most conspicuous structural units are the two CBMs spanning residues 25–112 [N-terminal CBM (CBMn)] and residues 128–216 [C-terminal CBM (CBMc)] respectively. Both modules were highly similar and consist of two adjacent hevein domains separated by a linker tetrapeptide (PYPE and PFPP in CBMn and CBMc respectively) (Figure 1B). The first and second hevein domains of the CBMn will further be referred to as Hev1 and Hev2, and those of the CBMc as Hev3 and Hev4 respectively. A CLUSTAL multiple alignment of the sequences of CBMn and CBMc revealed that they share 61 % identity (Figure 1B).

A next readily identifiable structural element is the central proline-rich domain (spanning residues 113–127) that connects the two CBMs. In this domain, a single extensin repetitive Ser[Pro]$_n$ motif can be distinguished. In addition to these readily identifiable domains, Lycesca contains also a short (residues 1–24) N-terminal domain preceding the CBMn and a short C-terminal domain (residues 217–232) following the CBMc (Figure 2). Neither of these domains exhibits an obvious similarity to any known protein. Interestingly, the two putative N-glycosylation sites found in the Lycesca polypeptide are located in the N-terminal peptide preceding the first block of hevein domains (Asn-9 and Asn-18 respectively).

Molecular modelling of the four hevein domains of the tomato leaf lectin

To check whether the hevein domains in the protein encoded by cLET21N20 are functional, their overall fold and three-dimensional structure were determined by molecular modelling using the atomic co-ordinates of hevein as a model. HCA indicated

Table 1 Amino acid composition of the cloned tomato leaf lectin and the previously described tomato fruit lectins

	Number of residues per polypeptide chain			
Amino acid	Lycesca*	Fruit lectin†	Cherry tomato fruit lectin‡	Cherry tomato 42 kDa protein§
Ala	10	11 (+1)	8 (−2)	16 (+6)
Arg	9	12 (+3)	4 (−5)	8 (−1)
Asx	18	23 (+5)	20 (+2)	24 (+6)
Cys	32	32 (0)	32 (0)	32 (0)
Glx	24	38 (+14)	40 (+3)	56 (+32)
Gly	30	46 (+16)	64 (+34)	52 (+22)
His	2	6 (+4)	8 (+6)	16 (+14)
Ile	5	3 (−2)	8 (+3)	16 (+11)
Leu	5	0 (−5)	4 (−1)	16 (+11)
Lys	9	21 (+12)	12 (+3)	8 (−1)
Met	3	12 (+9)	8 (+5)	8 (+5)
Phe	6	Trace (−6)	4 (+2)	16 (+10)
Pro + Hyp	30	134 (+104)	96 (+66)	40 (+10)
Ser	17	64 (+47)	76 (+59)	64 (+47)
Thr	10	24 (+14)	24 (+14)	24 (+14)
Trp	7	14 (+7)	ND	ND
Tyr	9	11 (+2)	8 (−1)	8 (−1)
Val	6	0 (+6)	4 (−2)	8 (−2)
Hyp		81	44	8
Total	232	451 (+219)	420 (+188)	464 (+232)
Molecular mass of the polypeptide (kDa)				
Calculated	24.897	46.965	41.301	45.820
SDS/PAGE			70	
CHO content (%)	Maximally 25	50	52	10
Molecular mass of native protein (kDa)				
Calculated‖	33.397	93.930	86.044	
SDS/PAGE			110	50.402
Ultracentrifugation		71		42

* Calculated from the deduced sequence of the protein encoded by cLET21N20 (after removal of the signal peptide).

† Calculated from the data reported by Nachbar et al. [3] assuming that the lectin contains 32 cysteine residues. Values in parentheses indicate the differences between the number of residues calculated from the amino acid composition and the number calculated from the deduced sequence of the leaf lectin.

‡ Calculated from the data reported by Naito et al. [8] assuming that the lectin contains 32 cysteine residues. Values in parentheses indicate the differences between the number of residues calculated from the amino acid composition and the number calculated from the deduced sequence of the leaf lectin.

§ Calculated from the data reported by Naito et al. [8] assuming that the 'WGA-like domain' of the 42 kDa lectin-related protein comprises four hevein domains with a total number of 32 cysteine residues. Figures between parentheses indicate the differences between the number of residues calculated from the amino acid composition and the number calculated from the deduced sequence of Lycesca.

‖ Taking into account the carbohydrate content.

that all four hevein domains of Lycesca exhibit an HCA plot very similar to that of hevein [9] and the hevein domains of the stinging nettle lectin UDA [19,20] (Figure 3), and accordingly have a very similar overall structure. This typical structural motif consists of three antiparallel strands of β-sheet (β1–β3) associated with two short α-helical segments (α1, α2) (Figure 3A). Eight extremely conserved cysteine residues participate in four disulphide bridges that tightly maintain the backbone fold of all these domains. Detailed structural analyses demonstrated that the interaction of GlcNAc residues of chito-oligosaccharides (chitotriose or chitotetraose) with the saccharide-binding site of hevein or the UDA domains relies on four conserved residues (Ser-19, Trp-21, Trp-23 and Tyr-30 of hevein and the first UDA domain; Ser-65, His-67, Trp-69 and Tyr-76 of the second UDA domain) [19,20], which create two hydrogen bonds with O3 (Oγ of Ser-19 or Ser-65) and O7 (Oη of Tyr-30 or Tyr-76) and a stacking interaction between the pyranose ring of the sugars and the aromatic ring of Trp-23/Trp-69. Additional hydrogen bonds are formed between O and N of Cys-24 and O6 of GlcNAc residues [19,20]. After binding to the UDA hevein domains, the chito-oligosaccharide chain adopts an extended linear conformation. Accordingly, the side chains of the four residues forming the binding site of hevein or UDA must be properly orientated to form the hydrogen bonding and the stacking interactions with the GlcNAc residues. Hevein domain 1 of Lycesca contains a strictly conserved chito-oligosaccharide-binding site with similarly orientated serine (Ser-19) and aromatic residues (Trp-21, Trp-23 and Tyr-30) (Figure 3B). In contrast, hevein domain 2 of Lycesca definitely differs from hevein by the replacement of Trp-21 by a Gly-21 residue. However, since Tyr-20 residue of this second hevein domain of Lycesca is similarly orientated as Trp-21 of hevein, stacking of the pyranose ring can still take place and accordingly one can reasonably expect that the binding site is fully active (Figure 3C). A very similar situation is encountered in hevein domain 4 of Lycesca. Here Ser-21 replaces Trp-21 of hevein but stacking of the pyranose ring to the aromatic residue Phe-20 can still take place, which has the same orientation as Trp-21 in hevein (Figure 3E). It should also be mentioned that in hevein domain 4, His-30 replaces Tyr-30. The same replacement has been found in the second hevein domain of UDA. Domain 3 of Lycesca strikingly differs from the three other domains by the replacement of Trp-23 of hevein by the Leu-24 residue (Figure 3D). Although all other key residues are conserved, such a replacement abolishes the stacking interaction with the chito-oligosaccharide chain and accordingly reduces the

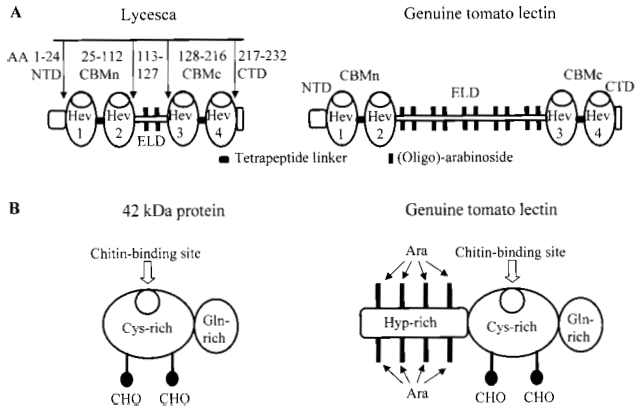

Figure 2 Modular organization of Lysesca and the genuine tomato lectin

(**A**) Schematic representation of the modular organization of Lysesca and the genuine tomato lectin. The two CBMs (CBMn and CBMc) are interconnected by a central ELD. Both hevein domains of each module are connected by linker tetrapeptide: NTD, CTD, the N-terminal and C-terminal domain respectively. (**B**) Previous model of the structure of the tomato lectin and the 42 kDa lectin-related protein (as described by Naito et al. [8] with slight modifications). Ara, arabinoside; CHO, carbohydrate.

binding capacity of hevein domain 3. Summarizing, it can be concluded from the modelling studies that hevein domains 1, 2 and 4 are most probably fully active, whereas hevein domain 3 has no or a strongly reduced binding activity towards extended chitotriose or chitotetraose chains. This implies that the Lycesca protomer is a trivalent rather than a tetravalent lectin.

Comparison of the overall structure and sequence of the protein encoded by tomato clone cLET21N20 and previously described tomato lectins

Alignment of Lycesca with previously reported protein sequences

In a first approach to confirm the identity of Lycesca and corroborate its relationship with the previously described tomato lectins, the N-terminal and internal sequences reported by Naito et al. [8] were aligned with the deduced sequence of the tomato clone. As shown in Figure 4, two of the V8 proteolytic fragments of the lectin (corresponding to peptides A2 and B4) can reasonably well be aligned with two different regions of Lycesca (located in Hev1 and Hev3 respectively). This striking sequence similarity taken together with the fact that matches are found in both CBMn and CBMc indicates that cLET21N20 encodes a protein that is related to the genuine cherry tomato fruit lectin. No match could be found between the reported N-terminal sequence [8] and the N-terminus of the deduced sequence of cLET21N20. However, the reported N-terminal sequence definitely exhibits similarity with the N-terminal part of the proline-rich linker between the two CBMs. At present, no explanation can be given for this discrepancy. Since the lectin preparations we isolated yielded no signal after Edman degradation, the question has to be addressed whether the reported N-terminal sequence really corresponds to the N-terminus of the lectin. Possibly, the reported sequence was derived from a fragment of the lectin generated during the purification procedure (by the action of endogenous proteases). Although not explicitly mentioned, the sequences reported by Naito et al. [8] were obtained by sequencing the lectin/peptides from a solution. If so, the presumed N-terminal sequence of the lectin may be derived from a fragment of the lectin polypeptide generated by a proteolytic cleavage at the end of the N-terminal block of two hevein domains (thus exposing the linker between the two CBMs).

Amino acid composition

In a second approach, the amino acid composition of the protein, encoded by cLET21N20, was compared with that of the previously described tomato lectins (Table 1). Since the molecular mass of the lectins we isolated and the previously described tomato lectins are comparable, one can assume reasonably that they all have a similar overall structure. Therefore it seemed preferable to make a comparison based on the number of residues of amino acids (rather than percentages). Accordingly, the data published by Nachbar et al. [3] and Naito et al. [8] were recalculated and expressed as residues per molecule. Thereby, all calculations were based on the assumption that all these lectins comprise the same number of cysteine residues (i.e. eight per hevein domain or 32 in total).

When expressed in this way, two striking differences become apparent. First, assuming an equal number of cysteine residues, the lectins described by Nachbar et al. [3] and Naito et al. [8] contain 219 and 188 extra residues respectively (when compared with Lycesca). Secondly, the four amino acids, glycine, proline, serine and threonine, account for the bulk of these extra residues (181 out of 219 and 173 out of 188 respectively). This implies that according to the data shown in Table 1, the tomato lectins described by Nachbar et al. [3] and Naito et al. [8] contain an extra sequence of approx. 200 residues comprising mainly proline, serine, glycine and threonine residues (in a 6:3:1:1 and 4:4:2:1 ratio respectively).

Irrespective of the obvious differences in amino acid composition, the deduced sequence of Lycesca poses a problem because the total number of amino acids and the calculated molecular mass of the polypeptide chain are apparently too low to account for the molecular mass of the tomato lectin. At present, exact values for the molecular mass of deglycosylated tomato lectin subunits have not been reported, but one can reasonably assume that this value is comparable with that of deglycosylated potato lectin [31.431 kDa as determined by MALDI–TOF (matrix-assisted laser-desorption ionization–time-of-flight)] [21]. Taking into account the calculated mass from the deduced sequence of cLET21N20, this implies that the Lycesca polypeptide is approx. 6.5 kDa (or approx. 60 residues) smaller than the naked potato lectin polypeptide chain. Although the glycosylation status of Lycesca cannot be determined from the deduced amino acid sequence, the maximal carbohydrate content can be calculated. Assuming that the serine residue is O-galactosylated and all proline residues are hydroxylated and O-glycosylated (with a tetra-arabinose chain), the carbohydrate moiety of the ELD can maximally account for approx. 6 kDa. Even if both putative N-glycosylation sites are occupied with a typical plant heptasaccharide (of approx. 1.250 kDa each), the maximal carbohydrate content of Lycesca cannot exceed the equivalent of 8.5 kDa, which implies that the maximal molecular mass of the native protomer is approx. 33.4 kDa. Since this value is considerably lower than that of the native subunit of the potato lectin (55.01 kDa as determined by MALDI–TOF) [21] and the closely related tomato fruit lectin (approx. 70 kDa as estimated by SDS/PAGE and ultracentrifugation) (Table 1), it seems unlikely that Lycesca corresponds to the previously described genuine tomato fruit lectins. However, the calculated molecular mass of Lycesca closely resembles that of the 42 kDa lectin-related protein described by Naito et al. [8] (which, on taking into account a carbohydrate content of 10 %, consists of a naked polypeptide of approximately the same size as Lycesca). This taken together with the fairly good

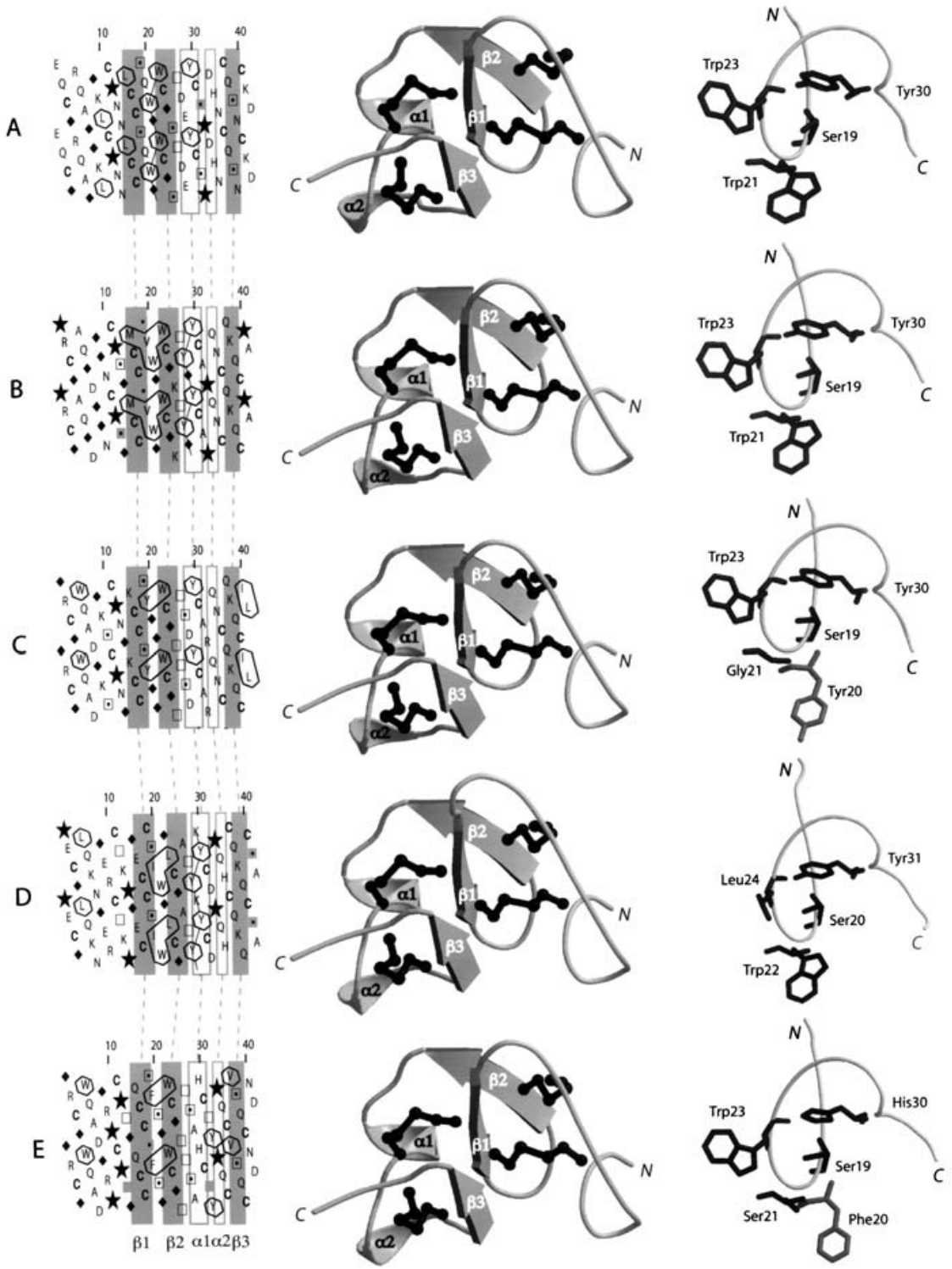

Figure 3 Structural similarity between Lycesca and hevein

Comparison of the HCA plots (left-hand side), three-dimensional models (centre) and carbohydrate-binding region (right-hand side) of hevein (**A**), and Hev1 (**B**), Hev2 (**C**), Hev3 (**D**) and Hev4 (**E**) of Lycesca. The strands of β-sheet and α-helices occurring along the HCA plots are shown by grey and white boxes respectively. They are numbered ($\beta1$–$\beta3$ and $\alpha1$, $\alpha2$) on the ribbon models and the cysteine residues involved in the four intrachain disulphide bridges are shown by black ball-and-sticks. The amino acid residues forming the saccharide-binding site are shown by black balls-and-sticks diagrams and are labelled.

matches between the partial sequences of the 42 kDa protein and Lycesca suggest that cLET21N20 encodes a leaf homologue of the fruit lectin-related protein.

The deduced sequence leaves no doubt that Lycesca is considerably smaller than the previously described tomato lectins. According to the differences in amino acid composition, the

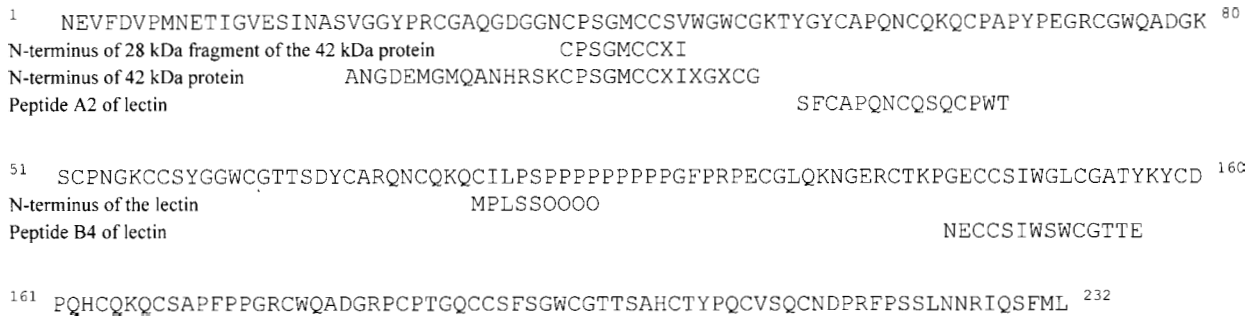

Figure 4 Sequence alignment of Lycesca and reported sequences

Alignment of previously reported sequences of the tomato lectin and lectin-related protein with the deduced amino acid sequence of tomato clone cLET21N20.

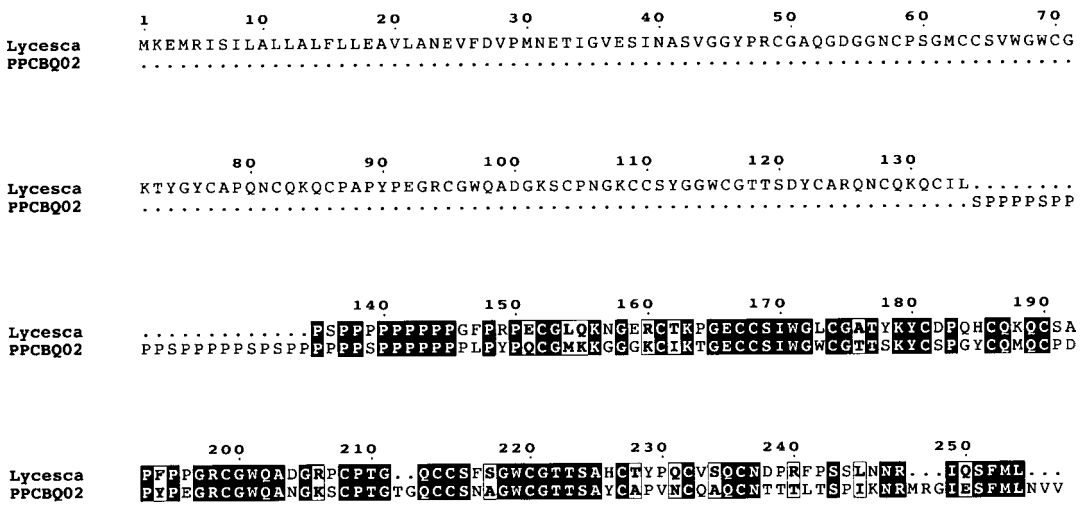

Figure 5 Sequence alignment of Lycesca and a potato EST

CLUSTAL-X (1.81) aligment of the deduced sequence of Lycesca and *Solanum tuberosum* cDNA clone PPCBQ02 (BI434398; EST537159; GI:15259088). The clone PPCBQ02 corresponds most probably to the CBMc of a putative potato lectin.

apparent difference in size is due to the higher number of proline, serine, threonine and glycine residues of the previously described tomato lectins. This may indicate that the ELD of the latter lectins is considerably longer than the relatively short proline-rich linker found in Lycesca. Indirect evidence for the occurrence of a longer ELD in the tomato lectin is provided by the analysis of a potato EST encoding chimaeric proteins, comprising an ELD of approx. 30 amino acid residues linked to a module resembling the CBMc of Lycesca (Figure 5). Interestingly, the C-terminus of the lectin encoded by the potato EST shares a reasonable sequence identity/similarity with the C-terminal domain of Lycesca.

A novel model of the molecular structure of the tomato lectin

The identification and sequencing of cLET21N20 provides for the first time unambiguous evidence for the structure of the protomer of a typical representative of the Solanaceae lectins. As already mentioned above, Lycesca consists of two homologous blocks of twin hevein domains, which are separated from each other by a relatively short proline-rich linker sequence. In addition, the protomer contains two short unrelated peptides located at the N- and C-termini respectively. Although this novel structure confirms the 'canonical' chimaeric nature of the tomato lectin there are three important differences. First, the previously proposed N-terminal location of the extensin domain does not correspond to the central location of the proline-rich domain of Lycesca. Secondly, the cysteine-rich chitin-binding domain does not correspond to a homologue of WGA (consisting of four tandemly arrayed hevein repeats) but to two blocks of twin hevein domains interspersed by a short ELD. Thirdly, Lycesca definitely lacks the glutamine-rich domain (resembling the large subunit of the tomato seed 2 S albumin) that according to the model of Naito et al. [8] is located at the C-terminus of the cherry tomato fruit lectin. This observation raises the question whether the C-terminal glutamine-rich domain is an integral part of the lectin or probably associates with the lectin through non-covalent interactions. Taking into consideration the amphiphilic character of the 2 S albumins and their tendency to interact with other proteins by non-specific interactions, this possibility can certainly not be excluded. It is also worthwhile to refer here to the work by Kilpatrick [23], who found that extracts from tomato seeds (containing 2 S albumin) did not exhibit any cross-reaction with an antiserum against the tomato fruit lectin, indicating that their lectin preparation lacked a 2 S albumin-like domain.

The determination of the complete (deduced) sequence has important consequences for our understanding of the molecular structure of the tomato lectin and its apparent anomalous behaviour after electrophoresis and centrifugation. It is evident, indeed, that the central position of the ELD profoundly affects the shape of the tomato lectin protomer because blocks of (hydroxy) proline contribute in terms of conformation to molecular inflexibility [22]. Since the (rigid) central extensin domain behaves as a rod-like structure to which, at both ends, a block of two in-tandem hevein domain is attached, the tomato lectin protomer cannot adapt a globular structure but rather exhibits an elongated shape. This conclusion is in good accordance with the results of analytical ultracentrifugation (including both sedimentation velocity and sedimentation equilibrium experiments), which indicated that the tomato lectin behaves as a prolate ellipsoid with an axial ratio of approx. 1:20 [3]. The presence of the central ELD also explains why the tomato lectin, like all other Solanaceae lectins, migrates with a much higher apparent molecular mass than can be expected on the basis of the length of the polypeptide. Denaturation (with SDS) and reduction (with dithiothreitol or 2-mercaptoethanol) eventually yield a completely unfolded polypeptide chain (with a more or less even distribution of bound SDS molecules) that behaves as a rod-like structure after SDS/PAGE. However, the densely glycosylated ELD (carrying multiple clustered O-linked oligoarabinosides on a rigid and inflexible polypeptide backbone) cannot be unfolded and hence retains its typical rod-like structure. Since this structure has a much larger diameter than that of a denatured SDS-bound polypeptide, the central extensin domain acts as a sort of knot in a string and reduces the mobility of the tomato lectin after SDS/PAGE (especially when the pore size is small). The striking differences between the reported molecular mass values of the tomato lectin most probably rely on differences in the pore size of the gels used for SDS/PAGE. Unfortunately, the effect of the pore size cannot be estimated because most papers give only the percentage of acrylamide and not the percentage of bis-acrylamide.

Conclusions

The protein encoded by cLET21N20 consists of two modules of twin hevein domains interspersed by a short proline-rich spacer containing a single extensin repetitive motif. Since, according to the modelling studies, at least three hevein domains are fully functional, one can reasonably assume that the protein behaves as a multivalent chitin-binding lectin. Most probably, Lycesca does not correspond to the genuine high molecular-mass tomato lectin but rather to a leaf homologue of the 42 kDa lectin-related protein. Taking into account the well-documented close structural similarity, one can reasonably expect that the genuine tomato lectin contains an ELD that is considerably longer than that of Lycesca and comprises > 30 residues (as can be inferred from an analysis of potato ESTs encoding chimaeric proteins comprising a twin hevein domain linked to an ELD of approx. 30 amino acid residues).

This work was partially supported by grants from the Fund for Scientific Research-Flanders (FWO grant no. G.0113.01).

Received 17 July 2003; accepted 22 September 2003
Published as BJ Immediate Publication 22 September 2003, DOI 10.1042/BJ20031069

REFERENCES

1 Van Damme, E. J. M., Peumans, W. J., Barre, A. and Rougé, P. (1998) Plant lectins: a composite of several distinct families of structurally and evolutionary related proteins with diverse biological roles. Crit. Rev. Plant Sci. **17**, 575–692
2 Allen, A. K., Desai, N. N., Neuberger, A. and Creeth, J. M. (1978) Properties of potato lectin and the nature of its glycoprotein linkages. Biochem. J. **171**, 665–674
3 Nachbar, M. S., Oppenheim, J. D. and Thomas, J. O. (1980) Lectins in the U.S. diet. Isolation and characterization of a lectin from the tomato (*Lycopersicon esculentum*). J. Biol. Chem. **255**, 2056–2061
4 Ashford, D., Desai, N. N., Allen, A. K., Neuberger, A., O'Neill, M. A. and Selvendran, R. R. (1982) Structural studies of the carbohydrate moieties of lectins from potato (*Solanum tuberosum*) tubers and thorn-apple (*Datura stramonium*) seeds. Biochem. J. **201**, 199–208
5 Allen, A. K. (1983) Potato lectin – a glycoprotein with two domains. Prog. Clin. Biol. Res. **138**, 71–85
6 Kieliszewski, M. J., Showalter, A. M. and Leykam, J. F. (1994) Potato lectin: a modular protein sharing sequence similarities with the extensin family, the hevein lectin family, and snake venom disintegrins (platelet aggregation inhibitors). Plant J. **5**, 849–861
7 Allen, A. K., Bolwell, G. P., Brown, D. S., Sidebottom, C. and Slabas, A. R. (1996) Potato lectin: a three-domain glycoprotein with novel hydroxyproline-containing sequences and sequence similarities to wheat-germ agglutinin. Int. J. Biochem. Cell Biol. **28**, 1285–1291
8 Naito, Y., Minamihara, T., Ando, A., Marutani, T., Oguri, S. and Nagata, Y. (2001) Domain construction of cherry-tomato lectin: relation to newly found 42-kDa protein. Biosci. Biotechnol. Biochem. **65**, 86–93
9 Anderson, N. H., Cao, B., Rodriguez-Romero, A. and Arreguin, B. (1993) Hevein: NMR assignment and assessment of solution-state folding for the agglutinin-toxin motif. Biochemistry **32**, 1407–1422
10 Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F. and Higgins, D. G (1997) The CLUSTAL-X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tool. Nucleic Acids Res. **15**, 4876–4882
11 Gaboriaud, C., Bissery, V., Benchetrit, T. and Mornon, J. P. (1987) Hydrophobic cluster analysis: an efficient new way to compare and analyse amino acid sequences. FEBS Lett. **224**, 149–155
12 Ponder, J. W. and Richards, F. M. (1987) Tertiary templates for proteins. Use of packing criteria in the enumeration of allowed sequences for different structural classes. J. Mol. Biol. **193**, 775–791
13 Mas, M. T., Smith, K. C., Yarmush, D. L., Aisaka, K. and Fine, R. M. (1992) Modeling the anti-CEA antibody combining site by homology and conformational search. Proteins Struct. Funct. Genet. **14**, 483–498
14 Laskowski, R. A., MacArthur, M. W., Moss, D. S. and Thornton, J. M. (1993) PROCHECK: a program to check the stereochemistry of protein structures. J. Appl. Crystallogr. **26**, 283–291
15 Kraulis, P. J. (1991) Molscript: a program to produce both detailed and schematic plots of protein structures. J. Appl. Crystallogr. **24**, 946–950
16 Esnouf, R. M. (1997) An extensively modified version of Molscript that includes greatly enhanced coloring capabilities. J. Mol. Graphics **15**, 132–134
17 Merritt, E. A. and Bacon, D. J. (1997) Raster3D photorealistic molecular graphics. Methods Enzymol. **277**, 505–524
18 Kilpatrick, D. C., Weston, J. and Urbaniak, S. J. (1983) Purification and separation of tomato isolectins by chromatofocusing. Anal. Biochem. **134**, 205–209
19 Saul, F. A., Rovira, P., Boulot, G., Van Damme, E. J. M., Peumans, W. J., Truffa-Bachi, P. and Bentley, G. A. (2000) Crystal structure of *Urtica dioica* agglutinin, a superantigen presented by MHC molecules of class I and class II. Structure Fold. Des. **8**, 593–603
20 Harata, K. and Muraki, M. (2000) Crystal structures of *Urtica dioica* agglutinin and its complex with tri-*N*-acetylchitotriose. J. Mol. Biol. **297**, 673–681
21 Kieliszewski, M. J. and Orlando, R. (1997) Matrix assisted laser desorption/ionization time-of-flight mass spectrometry of tomato extensin monomers and potato lectin. Phytochemistry **45**, 9–14
22 Kieliszewski, M. J. and Lamport, D. T. A. (1994) Extensin: repetitive motifs, functional sites, post-translational codes, and phylogeny. Plant J. **5**, 157–172
23 Kilpatrick, D. C. (1983) Tomato (*Lycopersicon esculentum*) lectin and serologically related molecules. Prog. Clin. Biol. Res. **138**, 63–70