Glycoconj J (2009) 26:161–172
DOI 10.1007/s10719-008-9173-1

Complete primary structure of a newly characterized galactose-specific lectin from the seeds of *Dolichos lablab*

Nagender Rao Rameshwaram · Narasimha Kumar Karanam · Christian Scharf · Uwe Völker · Siva Kumar Nadimpalli

Received: 30 May 2008 / Revised: 24 July 2008 / Accepted: 28 July 2008 / Published online: 12 September 2008

Abstract A new unique lectin (galactose-specific) purified from the seeds of *Dolichos lablab*, designated as DLL-II is a heterodimer composed of closely related subunits α and β. These were separated by SDS-PAGE and isolated by electroelution. By ESI-MS analysis their molecular masses were found to be 30.746 kDa (α) and 28.815 kDa (β) respectively. Both subunits were glycosylated and displayed similar amino acid composition. Using advanced mass spectrometry in combination with *de novo* sequencing and database searches for the peptides derived by enzymatic and chemical cleavage of these subunits, the primary sequence was deduced. This revealed DLL-II to be made of two polypeptide chains of 281(α) and 263(β) amino acids respectively. The β subunit differed from the α subunit by the absence of some amino acids at the carboxy terminal end. This structural difference suggests that possibly, the β subunit is derived from the α subunit by posttranslational proteolytic modification at the COOH-terminus. Comparison of the DLL-II sequence to other leguminous seed lectins indicates a high degree of structural conservation.

Keywords *Dolichos lablab* · Leguminosae · Agglutinin · α and β subunits · Matrix-assisted laser desorption/ionisation mass spectrometry · Galactose-specific seed lectin (DLL-II) · Amino acid sequence

Abbreviations

DLL	*Dolichos lablab* plant
DLL- I	Glucose/mannose-specific lectin from the seeds of *Dolichos lablab*
DLL- II	Galactose-specific lectin from the seeds of *Dolichos lablab*
DLL- VL	Galactose-specific lectin from the vegetative tissues of *Dolichos lablab*
MALDI	Matrix assisted laser desorption/ionization
ESI	Electro spray ionisation
MS	Mass Spectrometry

Introduction

Boyd and Shapleigh [1] in 1954 coined the term "lectin" for a class of structurally diverse proteins that bind carbohydrates. Each lectin molecule typically contains two or more binding sites, that is, they are divalent or polyvalent. Therefore, when they react with cells, they will not only combine with sugars on their surfaces but will also cause cross-linking of the cells and their subsequent precipitation, a phenomenon referred to as cell agglutination or hemagglutination [2]. A large number of legume lectins characterized have been shown to be oligomers composed of two types of

Electronic Supplementary Material The online version of this article (doi:10.1007/s10719-008-9173-1) contains supplementary material, which is available to authorized users.

N. R. Rameshwaram · S. K. Nadimpalli (✉)
Department of Biochemistry, School of Life sciences,
University of Hyderabad,
Hyderabad 500 046, India
e-mail: drnsk7@yahoo.co.in

N. K. Karanam · C. Scharf · U. Völker
Interfaculty Institute for Genetics and Functional Genomics,
Ernst-Moritz-Arndt-University Greifswald,
Friedrich-Ludwig-Jahn-Str. 15A,
17487 Greifswald, Germany

C. Scharf
Department of Otorhinolaryngology Head and Neck Surgery,
Ernst-Moritz-Arndt-University Greifswald,
Walther-Rathenau-Str. 43–45,
17475 Greifswald, Germany

subunits [3] that display different functional properties [4]. The subunits can originate from different genes, as in the case of *Phaseolus vulgaris* lectin [5], or by proteolytic splitting of a single gene product as has been found in the biosynthesis of *Dolichos biflorus* seed lectin, DB 58 lectin [6], favin [7], the pea [8] and the lentil [9] lectins.

Plant lectins are mostly found in the vegetative and storage parts of the plant where they presumably play an important physiological role which has not been clearly established. *Dolichos lablab* (Indian lablab beans) is a legume plant that is widely grown in south India and is used as a vegetable in diets. The seeds contain two types of lectins specific for selected carbohydrates: glucose/mannose hitherto referred as DLL-I and galactose hitherto referred as DLL-II. DLL-I is a tetramer with a native molecular mass of 60 kDa that is made of two types of subunits 15k Da and 12 kDa respectively. It was completely characterized and is virtually identical to the lectin from field beans whose primary structure had been deduced by conventional protein sequencing methodology [10]. We successfully affinity purified the DLL-II which is a glycoprotein with an apparent molecular mass of 120 kDa and is made of two types of subunits with molecular masses of 31 kDa (α subunit) and 29 kDa (β subunit), respectively suggesting the lectin to be tetrameric in nature ($\alpha_2\beta_2$ type). An antibody raised against the 31 kDa subunit recognized both subunits of DLL-II [11, 12]. Furthermore, DLL-II also cross-reacts with a DLL-I antibody suggesting antigenic similarities among the two distinct sugar specific lectins from the same seeds. The stems and leaves of this plant were also found to contain the galactose-specific lectin that cross-reacts with the same antibody, suggesting that these two lectins may be related immunologically [13]. In literature there are only a few legume plant seeds that have been shown to contain two distinct types of lectins. Since we extensively characterized the newly identified DLL-II, and the available evidence suggested that it may be related to the DLL-I it became important to obtain the primary sequence of the DLL-II for a variety of reasons. First by obtaining the complete sequence information of a new lectin, we can understand its structure, and its relatedness to the DLL-I as well as with other legume lectins. Second, the identification of specific conserved regions among the different lectins would further substantiate the notion that the legume lectins are conserved proteins. Third, the specific amino acids that may be important in the lectin activity can be identified. Fourth, the sequence information would also provide a lead to obtain the three dimensional structure of this protein that will eventually establish the structure–function relationship of the newly characterized DLL-II lectin. To achieve these objectives, in the present study we used a mass spectrometry centered proteomics approach to deduce its complete primary sequence. Since the early 1980's, when mass spectrometry was first employed for the measurement of the molecular weight of short peptides, great advances have been made in the uses of MS for protein and peptide identification and MS is now a routine method for rapid identification and characterization of proteins. It is now commonly used in areas as diverse as genotyping for molecular medicine and cancer research to food quality evaluation, as well as in the analysis of plant proteomes and discovery of novel plant viruses [14]. Current methodologies generally combine two stages: fractionation of proteins or peptides followed by ionization of the sample using either MALDI or ESI. The first of these, the separation step, is necessary to reduce complexity allowing comprehensive coverage of even complex mixture by the subsequent mass spectrometric analysis. Separation protocols include: protein separation using 1- or 2-DE followed by digestion of the products using a cleavage agent, most often sequence specific proteases such as trypsin. Typically the experimental data generated by MS are then compared with theoretical peptide masses or fragment ion mass values, derived by in-silico digestion of entire sequence databases with the same enzymes/chemicals. Corresponding matches are scored with probability based methods in a way that allows the peptide or protein that best matches the data to be identified. However, if the protein of interest is not part of the existing sequence databases, the aim would be to identify those entries that exhibit closest homology, often-equivalent to related species [15]. Alternatively, the sequence of the protein could be determined by *de-novo* sequencing. The primary sequence of DLL-II deduced here by mass spectrometry provided valuable information on the relatedness of DLL-II with DLL-I and other legume lectins.

Materials and methods

The seeds of the *Dolichos lablab* were purchased from the local market. Reagents used for sequencing and other chemicals used in the study were all of high purity and procured from standard firms.

Purification of DLL-II

The *Dolichos lablab* seeds were obtained from the local market. The lectin was purified from the saline extracts of defatted lablab bean flour by fractional precipitation with ammonium sulphate, followed by affinity chromatography as described earlier [11].

Separation and confirmation of subunits

We extensively characterized DLL-II [11,12]. It is a glycoprotein with an apparent molecular mass of 120 kDa

and dissociates into 31 kDa and 29 kDa on denaturing electrophoresis indicating probably that it is tetrameric in nature. The subunits were separated on SDS-PAGE and electroeluted from the gel. The electroeluted purified subunits were collected manually, stored at 4°C and subunit total masses were determined by mass spectrometry using an API-QSTAR PULSAR instrument.

Preparation of proteins for analysis by mass spectrometry

The subunits were separated on a 12% SDS-polyacrylamide gel and visualized by colloidal coomassie blue staining. α and β subunits were excised from the gel, destained with 25 mM ammonium bicarbonate in 50% acetonitrile, dehydrated with 100% acetonitrile and vacuum dried in a speedvac. Dried gel slices were in gel-digested separately with trypsin, chymotrypsin, Glu-C and aspartase N overnight at 37°C. In gel digestion with different proteases were done according to manufacturer's description. Peptides were extracted twice with 50% acetonitrile containing 0.1% acetic acid in a water bath sonicator for 15 min each. The digests of each enzyme were pooled, vacuum dried in a speedvac and redissolved in water containing 0.1% trifluoroacetic acid (TFA) for desalting using C18 ziptips (Millipore, Bedford, MA, USA). Finally, peptides were dissolved in α- 4 hydroxy cinnamic acid (CHCA) for MALDI-MS measurements (see below).

For limited acid hydrolysis the purified subunits were desalted using C18 ziptips (Millipore, Bedford, MA, USA) (10 µl), hydrolyzed with 3 and 6 N HCl (1:10) separately and incubated for 2 and 5 min, respectively, at 100°C. Then, the samples were transferred on ice to stop the reaction and vacuum dried. The peptide containing pellet was reconstituted in 4–6 µl of CHCA.

For formic acid treatment, excised vacuum dried gel slices were swollen with 0.1 ml of 75% formic acid, which was completely absorbed, and were then incubated in tightly closed Eppendorf tubes at 37°C for 18 to 24 h. Digestion was stopped by removing the formic acid in a speedvac. Finally, peptides were dissolved in CHCA for MALDI MS measurements (see below).

2-Iodosobenzoic acid (IBA) cleavage was performed on SDS-PAGE purified subunits as described for formic acid cleavage, except that SDS was removed from the gel slices before treatment by washing in 10% (vol/vol) acetic acid-25% (vol/vol) isopropanol or by colloidal coomassie blue staining. This was necessary because solubilization of IBA requires 4 M guanidine, guanidine and SDS tend to form an insoluble gel. After removal of SDS, protein containing gel slices were dried, and 0.1 ml of 10 mg per ml IBA dissolved in 80% acetic acid – 4 M guanidine was added. The gel slices were then incubated in tightly sealed tubes at room temperature for 24 h. Guanidine was removed by washing with acetic acid; the gel slices were vacuum dried in a speedvac. Peptides were dissolved in CHCA for MALDI-MS measurements (see below).

Hydroxylamine treatment was carried out as described by Saris *et al* [16]. The vacuum dried digests were reconstituted in 4–6 µl of CHCA matrix and spotted directly onto MALDI targets. The peptide mixtures generated were analyzed with positive-ion matrix-assisted laser desorption/ionisation-mass spectrometry on a 4800 MALDI TOF/TOF Analyzer (Applied Biosystems, Foster City, CA, USA).

Carboxy-terminal sequencing

The purified α and β subunits were dissolved in 10 µl of sodium citrate buffer (50 mmol/l, pH 6.0). Carboxypeptidase Y was added to give an enzyme: protein ratio of 1:100 by weight. After 10 min incubation the same amount of carboxypeptidase P was added and incubated for 60 min. Aliquots of 1.5 µl digests were mixed with the CHCA matrix solution and spotted directly on MALDI target and analyzed with positive-ion matrix-assisted laser desorption/ionisation-mass spectrometry 4800 MALDI TOF/TOF Analyzer (Applied Biosystems, Foster City, CA, USA) following the protocol of Thiede *et al* [17].

Mass spectrometric analysis of the peptides

A 1/10 fraction of the peptides dissolved in 0.1% TFA and mixed with equal volumes of the MALDI matrix, a-cyano-4-hydroxycinnamic acid (CHCA), was analyzed with positive-ion matrix-assisted laser desorption/ionisation-mass spectrometry using a 4800 MALDI TOF/TOF Analyzer (Applied Biosystems, Foster City, CA, USA). The spectra were recorded in reflector mode in a mass range from 800 to 3700 Da with a focus mass of 2000 Da. For one main spectrum 25 sub-spectra with 100 shots per sub-spectrum were accumulated using a random search pattern. If the autolytical fragments of trypsin with the mono-isotopic $(M + H)^+$ m/z at 1045.5 and/or at 2211.1 reached a signal to noise ratio (S/N) of at least 10, an internal calibration was automatically performed using at least one peak for one- or both peaks for a two-point-calibration. Calibration was performed manually for the less than 1% samples for which the automatic calibration failed.

Additionally, MALDI-MS/MS analyses were performed for the five strongest peaks of the TOF-spectrum after subtraction of peaks corresponding to background, keratin and trypsin fragments. For one main spectrum 20 sub-spectra with 125 shots per sub-spectrum were accumulated using a random search pattern. The internal calibration was automatically performed as one-point-calibration if the mono-isotopic arginine (M + H) + m/z at 175.119 or lysine (M + H) + m/z at 147.107 reached a signal to noise ratio

(S/N) of at least 5. After calibration a combined database search of MS and MS/MS measurements was performed using the GPS Explorer software (Applied Biosystems, Foster City, CA, USA) with the following settings: (i) MS peak filtering: mass range from 800 to 3700 Da; minimum S/N filter of 10; peak density of 50 peaks per range of 200 Da and maximal 200 peaks per protein spot; mass exclusion list contained background peaks and trypsin fragments with an exclusion tolerance of 100 ppm (ii) MS/MS peak filtering: mass range from 60 Da to a mass that was 20 Da lower than the precursor mass; peak density of 50 peaks per 200 Da and maximal 65 peaks per MS/MS; minimum S/N filter of 10 (iii) database search: precursor tolerance 35 ppm and MS/MS fragment tolerance 0.65 Da. The peptide search tolerance was 35 ppm but the actual RMS value was between 5 and 15 ppm. Peak lists were compared with a SwissProt database (R46) using the Mascot search engine (Matrix Science Ltd, London, UK). Peptide mixtures that yielded a Mowse score of at least 49 (p-value = 0.05) were regarded as positive identifications.

Peptide sequencing by MS/MS

De novo sequencing was carried out on a 4800 MALDI TOF/TOF Analyzer (Applied Biosystems, Foster City, CA, USA). Nitrogen was used as collision gas for MS/MS analysis. All samples were dissolved in 50% acetonitrile containing 0.1% v/v TFA and analyzed in the positive-ion mode. Using the *De novo* Explorer™ software (Applied Biosystems, Foster City, CA, USA) candidate amino acid sequences were generated from MS/MS data through a *de novo* sequencing algorithm. Subsequently, the candidate amino acid sequences were submitted for MS Blast database searching for protein identification. Additionally, all MS/MS spectra were analyzed manually using the sequencer tool of the Data Explorer software (Applied Biosystems, Foster City, CA, USA). Resulting sequence tags were also submitted for MS Blast database searching, charting the proteomes of organisms with unsequenced genomes by MALDI-quadrupole time-of-flight mass spectrometry and BLAST homology searching [18].

Results

Characterization of α and β subunits

Our earlier studies on DLL-II have shown that the protein has a native molecular mass of 120 ± 5 kDa and is composed of two subunits (α and β subunits) [11]. Here, the subunits were separated by denaturing electrophoresis, electro eluted from the gel and their molecular masses were determined by ESI-MS. The ions generated by ESI of larger peptides/proteins usually carry multiple protons and thus the ESI mass spectra of the two subunits displayed multiple charged species, these signals appeared in different parts of the mass spectrum, so all signals were used to calculate the molecular weight of the protein, and this resulted in improved mass accuracy of the subunits as shown in Fig. 1 (a,b). Deconvolution of the spectrum was done using the vendor specific Analyst QS software. The Bayesian protein reconstruct tool was used to calculate the molecular masses of the α and β subunits to be 30.746 kDa and 28.815 kDa, respectively, which is in good agreement to the calculated molecular weight (31,000 Da and 29,000 Da) based on SDS-PAGE [11].

The primary structure of the DLL-II

Mass spectrometry is the method of choice for the rapid analysis of complex peptide/protein mixtures in proteomics. These modern techniques of proteomics that have been developed during the last decade allow the fast and reliable identification and characterization of proteins even from complex mixtures. Currently, proteomics technologies are primarily applied for the identification and characterization of proteins from organisms with sequenced genomes but with the constant introduction of new bioinformatics tools in the future will also be applicable as a routine procedure for organisms, the sequence of which has not been determined yet. Here we applied a combination of traditional biochemical, mass spectrometry and bioinformatics tools to reveal and analyze the primary sequence of both the subunits of the affinity purified galactose-specific lectin.

After separation by SDS-PAGE a fraction of the peptides obtained by enzymatic and chemical digestion of the lectin subunits were analyzed by MALDI-TOF-MS to obtain complementary peptide mass fingerprints. Experimentally obtained masses were compared to peptide mass sets obtained from the in silico theoretical digests of lectin proteins in databases. Using this approach, however, only a very small fraction of the peptides could be assigned which was due to the fact that the sequence was novel and thus only sequences absolutely conserved among the lectins from different legume species could be assigned. Therefore, the next step was *de novo* sequencing of peptides by low-energy CID on a 4800 MALDI TOF/TOF Analyzer (Applied Biosystems, Foster City, CA, USA). Using this strategy we obtained clear-cut—manually interpretable—peptide fragmentation spectra displaying the expected consecutive runs of b^- and y^- ions. In this study the MASCOT, MS-BLAST 2, and PROTEIN PROSPECTOR programs were used for the analysis of the MS data allowing us to utilize the complete set of peptide sequences of unknown order as determined by MS/MS from one protein in the same query. Particularly, we made use of the complementary sequence information obtained from the

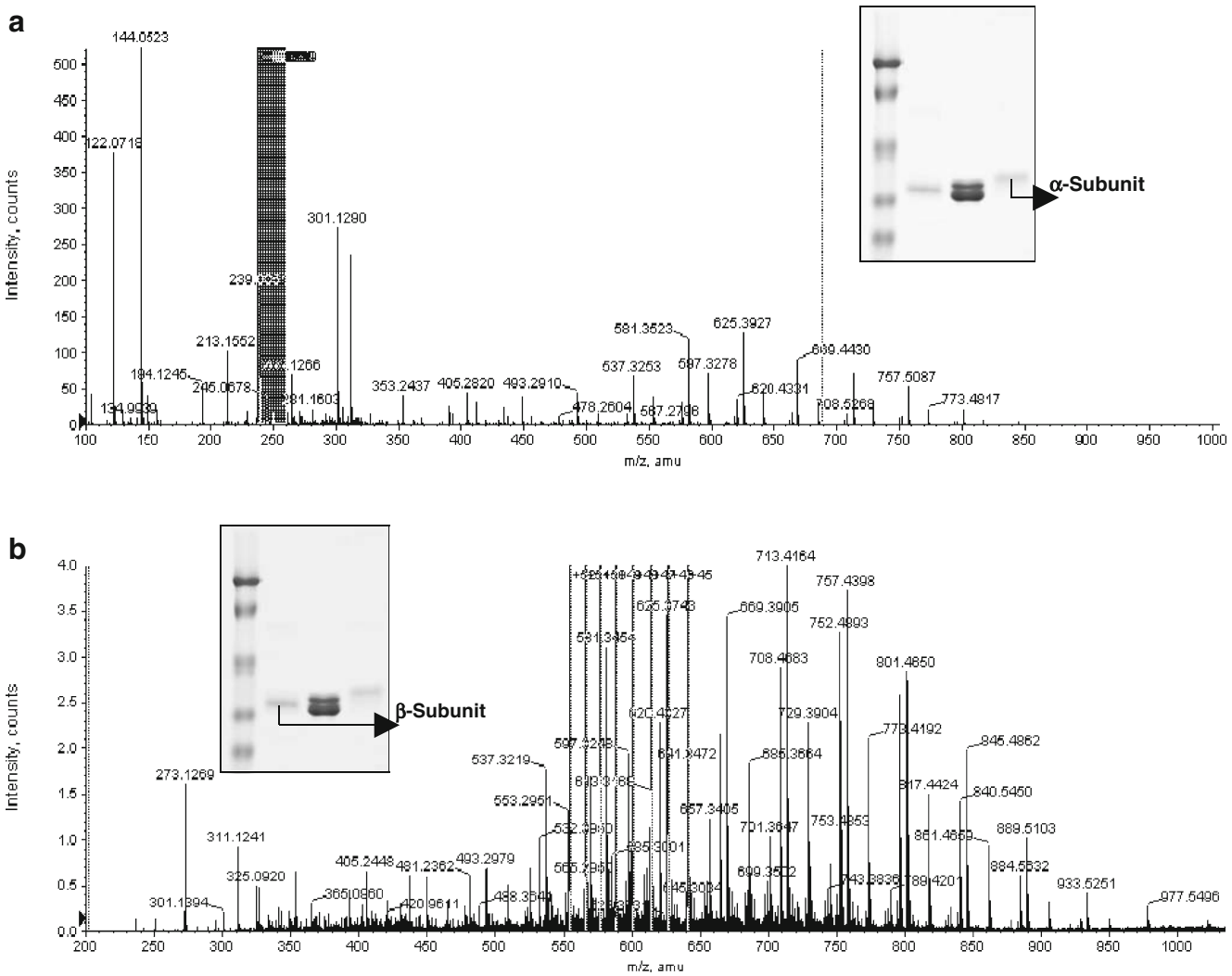

Fig. 1 Determination of the molecular mass of the lectin subunits. The molecular mass of the individual subunits was determined by ESI-MS with a API-QSTAR PULSAR triple quadrupole mass spectrometer. Mass spectra of the purified α subunit [31 kDa] (**a**) and β subunit [29 kDa] (**b**) are displayed. Inset shows migration of the separated and purified α subunit [31 kDa] (1**a**) and β subunit [29 kDa] (**b**) on a SDS-PAGE; also seen is the intact lectin with both subunits

different enzymatic and chemical digests. This information was partially overlapping. In order to distinguish between the two subunits the MS-aided determination of the lectin sequence was performed for the affinity purified lectin with isolated, separated intact subunits.

In Table of the Supplemental Data the details of the various peptide sequences that were derived from the proteolytic digestion of the subunits with different proteases and chemicals have been summarized. Digestion of the α subunit with trypsin and chymotrypsin yielded 22 peptides (T^1–T^{22}) and 28 peptides (C^1–C^{28}) respectively, from which the backbone of the sequence was established. Additional overlap in sequences for the tryptic and chymotryptic peptides was then provided with 13 peptides that resulted from cleavage at the Asp–NH_2 terminal peptide bond (A^1–A^{13}) by proteolytic digestion with the Asp-N endopeptidase. To close the remaining gaps and to support the sequence even further, 25 peptides generated by proteolytic digestion of the α subunit with Glu-C that specifically cleaves at Glu–COOH terminal (G^1–G^{25}) were added to the analysis. Chemical cleavage with 70% formic acid yielded thirteen fragments resulting from hydrolysis of the peptide bond between Asp-Pro (Fig. 2 a and b). Cleavage between Asn^{60}-Gly^{61}, Asn^{152}-Gly^{153} and Asn^{181}-Gly^{182} peptide bonds by treatment of hydroxylamine generated four large fragments. 'IBA' treatment hydrolyzed the tryptophanyl peptide bonds at positions (I^{80}, I^{90}, I^{155}, I^{177}, I^{226}, I^{250}), which resulted in generation of seven large fragments. Using the peptides generated by digestion with trypsin, chymotrypsin, Asp-N and Glu-C as well as the data generated by chemical degradation with formic acid or IBA the sequence of the α subunit was assembled (Fig. 2 a).

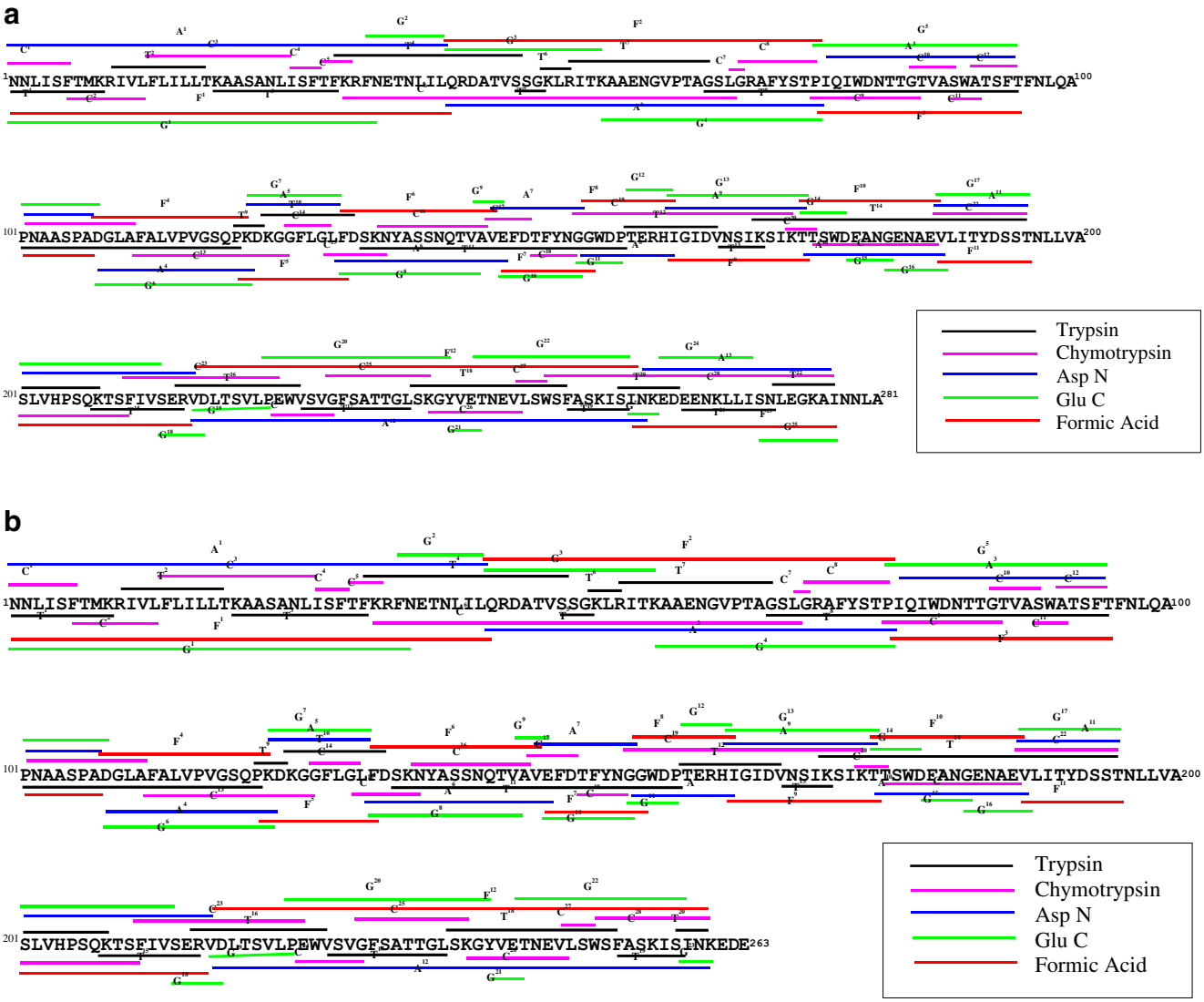

Fig. 2 Assembly of the α and β subunits of *Dolichos lablab* galactose-specific lectin primary structure from the various digestions. MALDI-MS/MS analysis of the different enzymatic and chemical digestion of the two subunits of lectin yielded overlapping data sets that were then used to assemble the primary sequence of both subunits. The figure displays the primary sequence obtained as well as the contribution of the different complementing digests. (**a**) Peptides supporting the α subunit are indicated in the amino acid sequence and those contributing to the elucidation of the β subunit are indicated in (**b**) with the amino acid sequence. The sequences obtained by MALDI-MS/MS analysis of α and β subunits is identical, differing only at the carboxy termini, with the α subunit possessing an 19 amino acid extension that is missing in the β subunit. Peptides obtained by trypsin, chymotrypsin, Glu C, aspartase N endopeptidase and 70% formic acid are indicated with different colors. Peptides generated with 2-iodosobenzoic acid, and hydroxylamines are not displayed

In a separate effort, the isolated β subunit was also subjected to the proteolytic and chemical treatments as was done for the α subunit. MS-analysis and interpretation of the MS/MS-fragmentation data of these different digests revealed 20 tryptic peptides, 28 chymotryptic peptides, 12 Asp-N peptides, 22 Glu C peptides. Peptides generated from chemical treatments with formic acid, hydroxylamine and 2-iodosobenzoic acid were also aligned as done for the α subunit. Performing the same analysis as described for the α subunit, the peptides of the β subunit were assembled to yield the complete primary structure (Fig. 2 a and b). Using the Data Explorer software—all MS spectra of both subunits was analyzed: Most of the spectra showed no differences in the peak distribution, allowing us to conclude that both subunits were highly similar. Additionally, we analyzed manually all differences between the corresponding spectra to reveal any differences in the sequence.

Since both subunits displayed identical primary sequence but yet differed in size the determination of the N- and C-termini was of particular importance. From the previous studies we learnt from the N terminal edman sequencing that at the N terminus the first 10 amino acids of the α and β

subunits were identical [11]. Thus, the determination of the carboxy-terminus was of special interest and was performed here by carboxypeptidase treatment followed by MALDI TOF MS/MS. Carboxypeptidase treatment released 18 carboxy terminal amino acids from the α subunit. Initially, we determined the mass of the α subunit by ESI-MS to be 30,746 Da which fits well to the initial rough estimation that was first based on SDS-PAGE (31 kDa). The molecular mass calculated on the basis of sequence analysis (30,420Da) was slightly smaller (326 Da) than that measured by ESI-MS likely reflecting a contribution to the total mass of post translational modifications that have not been identified yet. This observed mass difference probably does not originate from glycosylation because treatment with PNGase did not influence the mass determined by ESI-MS.

By identifying by MS peptides covering the individual N-termini of each of the subunits isolated from the SDS-PAGE we have reestablished that each of the subunits has an identical amino-terminus. Furthermore, the primary structure of the β subunit obtained from the mass spectrometry analysis of the proteolytic and chemical digests revealed that the sequence of the β subunit is identical in 263 amino acids and differs from the α subunit only in the absence of some amino acids at the carboxy-terminus described below.

Comparison of the carboxy termini of α and β subunits shows that the α subunit contains at its C-terminus four leucines, four aspargines, three isoleucines, two lysines, two alanines, two glutamic acid and an extra serine and glycine. An analysis of carboxy-terminus of the β subunit by carboxypeptidase treatment indicated EDE to be the COOH terminal sequence, which is also that of the C-terminal tryptic peptide. This assignment of the C-terminus of the β subunit was also supported by the molecular mass that was calculated on the basis of the derived sequence. This calculated value of 28,485 Da compares well with that estimated by SDS-PAGE (29,000 Da) and was again 330 Da smaller than the total mass of the subunit determined by ESI-MS (28,815 kDa) reinforcing the idea of a yet unidentified modification that was present in both the α and β subunits.

Comparison of the DLL-II sequence with other legume lectins

The sequence obtained from the *de novo* sequencing and manual sequencing using GPS explorer from Applied biosystems was compared with the glucose/mannose lectin (DLL-I) from the *Dolichos lablab* as well as with the other legume lectins. Structural comparison of the DLL-I sequences with other legume lectins, *e.g.* from *Dolichos biflorus*, Phytohemagglutinin, and Soybean agglutinin etc, suggests high degree of conservation/ identity. On the other hand the DLL-II sequenced in the present study showed only limited homology to the DLL-I but showed a higher degree of homology with other legume lectins described above, suggesting evolutionary relatedness between the DLL-I and II isolated from the *Dolichos lablab* seeds.

The sequence of the DLL-II deduced, was aligned with 14 other legume lectins of varying sugar specificity using the CLUSTAL W 1.82 (http://www.expasy.org) program. Both subunits of the DLL-II were identical except at the carboxy terminal end. Though the DLL-I and DLL-II distinctly differ in their sugar specificity, both are tetrameric glycoproteins that are made of two non-identical subunits which differ in their molecular masses. Direct sequence comparison of the DLL-I with the related DLL-II from *Dolichos lablab* revealed that both lectins showed only 36.4% identity, proving that the lectin analyzed in this study is novel and displays a different subunit heterogeneity. Comparison of the sequence with other legume lectins showed higher sequence conservation for the *Dolichos biflorus* seed lectin (59.3% identity), *Dolichos biflorus* stem and leaf lectin (62.6%) and *Phaseolus vulgaris* (55.3%). The extensive sequence conservation exhibited by both subunits of the galactose specific lectin encouraged us to investigate the similarity with other legume lectins. Indeed it is known in literature that the *Dolichos biflorus* lectin, favin and lentil lectins also exhibit this property [6–9]. Figure 3 shows the similarity/identity of the β subunit in its amino terminal region to other legume lectins. Figure 4 shows a diagrammatic representation of the sequences obtained for both the subunits of DLL-II. Table 1, summarizes the percentage of the DLL-II sequence identity with lectins of the same tribe *Phaseoleae* and with other legume lectins with different sugar specificity.

Discussion

The proteomics approach used in the present study illustrates a novel method for the characterization of lectins. Thus far, sequencing of many lectins has been exclusively performed using the conventional Edman degradation method [14] and cDNA cloning [19]. Unsequenced proteomes are now being well characterized using the advances of the mass spectrometry [20]. In the present study we used a MS based method to successfully deduce the complete primary sequence of the newly characterized galactose lectin (DLL-II) from the *Dolichos lablab* seeds. Applying mass spectrometry instead of conventional Edman sequencing greatly reduced the time and the effort of prior separation of individual peptides for sequencing because complex mixtures can be used with prefractionation by reverse phase liquid chromatography. Thus, the complete primary sequence of both subunits was deduced with a total of 25–35 μg of the purified DLL-II lectin.

```
    DLL-II      --------------NNLISFTMKRIVLFLILLTKAASANLISFTFKRFN------ETNLILQRDATVSS-GKLRITKAAE
LEC1_DOLBI      ---------------MASSTVSVVLSLFLLLLTQANSANIQSFSFKNFN------SPSFILQGDATVSS-GKLQLTKVKE
LEC5_DOLBI      ---------------MASSTVSVVLSLFLLLLTQAYSADIQSFSFKNFN------SSSFILQGDATVSS-SKLRLTKVKG
PHAE_PHAVU      ----------------MASSNLLSLALFLVLLTHANSASQTSFSFQRFN------ETNLILQRDATVSSKGQLRLTNVND
LEC_PEA         -------MASLQTQMISFYAIFLSILLTTILFFKVNSTETTSFLITKFS----PDQQNLIFQGDGYTT-KEKLTLTKAVK
LECS_SOPJA      MATSNSRPHLLQTHKPFSVVLAISITFFLLLLNKVNSAEILSFSFPKFA----SNQEDLLLQGDALVSSKGELQLTTVE-
LEC_LENCU       -------MASLQTQMISFYLIFLSILLTTIFFFKVNSTETTSFSITKFS----PDQKNLIFQGDGYTT-KGKLTLTKAVK
LEC1_LABAL      ------------------------------------LNELSFNFDKFV----PNQNNILFQGVASVSTTGVLQVTKVT-
LEC_VICFA       ----------------------------------TDEITSFSIPKFR----PDQPNLIFQGGGYTT-KEKLTLTKAVK
LEC_ERYCG       ---------------------------------VETISFSFSEFE----PGNDNLTLQGAALITQSGVLQLTKINQ
LECA_DOLLA      ---------------------------------AQSLSFSFTKFD----PNQEDLIFQGTATS---------KLDS
LEC_BOWMI       ---------------------------------ANSVCFTFTDFE----SGQQDLIFQGDASVGSNKALQLTKVDS
LEC_ERYCO       ----------MATYKLCSVLALSLTLFLLILNKVNSVETISFSFSEFE----PGNDNLTLQGAALITQSGVLQLTKINQ
LEC_SOYBN       -----MATSKLKTQNVVVSLSLTLTLVLVLLTSKANSAETVSFSWNKFV----PKQPNMILQGDAIVTSSGKLQLNKVDE
ARC2_PHAVU      ---------------MASSNLLTLALFLVLLTHANSSNDASFNVETFN------KTNLILQGDATVSSEGHLLLTNVKG

    DLL-II      NGVPTAGSLGRAFYSTPIQIWDN--TTGTVASWATSFTFNLQAPNAAS----PADGLAFALVPVGSQPK--DKGGFLGLF
LEC1_DOLBI      NGIPTPSSLGRAFYSSPIQIYDK--STGAVASWATSFTVKISAPSKAS----FADGIAFALVPVGSEPR--RNGGYLGVF
LEC5_DOLBI      NGLPTLSSLGRAFYSSPIQIYDK--STGAVASWATSFTANIFAPNKSS----SADGIAFALVPVGSEPK--SNSGFLGVF
PHAE_PHAVU      NGEPTLSSLGRAFYSAPIQIWDN--TTGAVAASPTSFTFNIDVPNNSG----PADGLAFVLLPVGSQPK--DKGGLLGLF
LEC_PEA         N------TVGRALYSSPIHIWDR--ETGNVANFVTSFTFVINAPNSYN----VADGFTFFIAPVDTKPQTG--GGYLGVF
LECS_SOPJA      NGVPIWNSTGRALYYAPVHIWDK--STGRVASFATSFSFVVKAPVAS----KSADGIAFFLAPPNNQIQ-GPGGGHLGLF
LEC_LENCU       S------TVGRALYSTPIHIWDR--DTGNVANFVTSFTFVIDAPSSYN----VADEFTFFIAPVDTKPQTG--GGYLGVF
LEC1_LABAL      N-----TGIKRALYAAPIHAWDDDSETGKVASFATSFSFVVKEPPIQSRKADGVDGLAFFLAPANSQIPSGSSAGMFGLF
LEC_VICFA       N------TVGRALYSLPIHIWDS--ETGNVADFTTTFIFVIDAPNGYN----VADGFTFFIAPVDTKPQTG--GGYLGVF
LEC_ERYCG       NGMPAWDSTGRTLYTKPVHMWDS--TTGTVASFETRFSFSIEQPYTRP---LPADGLVFFMGPTKSKP--AQGYGYLGVF
LECA_DOLLA      AGNPVSSSAGRVLYSAPLRLWED---SAVLTSFDPTIYIFTN---YTS----RIADGLAFIAPPDSVIS--YHGGFLGLF
LEC_BOWMI       KGNPQGGSVGRALYTAPIRLWQS---SSLVASFETTFTFSISQGSSTP----AAALTFFIASPDTKIPS--GSGGRLLGL
LEC_ERYCO       NGMPAWDSTGRTLYAKPVHIWDM--TTGTVASFETRFSFSIEQPYTRP---LPADGLVFFMGPTKSKP--AQGYGYLGIF
LEC_SOYBN       NGTPKPSSLGRALYSTPIHIWDK--ETGSVASFAASFNFTFYAPDTKR----LADGLAFFLAPIDTKPQTHAG--YLGLF
ARC2_PHAVU      NEE---DSMGRAFYSAPIQINDR--TIDNLASFSTNFTFRINAKNNEN----SAYGLAFALVPVGSRPK--LKGRYLGLF

    DLL-II      DSKN--YASSNQTVAVEFDTFYNG---GWDP-------TERHIGIDVNSIKSIKTTSWDFANGE---NAEVLITYDSSTN
LEC1_DOLBI      DSDV--YNNSAQTVAVEFDTFSNS---GWDP-------SMKHIGIDVNSIKSIATVSWDLAN---GENAEILITYNAATS
LEC5_DOLBI      DSDV--YDNSAQTVAVEFDTFSNT---DWDP-------TSRHIGIDVNSIKSIRTASWGLAN---GQNAEILITYNAATS
PHAE_PHAVU      NNYK--YDSNAHTVAVEFDTLYNV---HWDP-------KPRHIGIDVNSIKSIKTTTWDFVK---GENAEVLITYDSSTK
LEC_PEA         NSA--EYDKTTQTVAVEFDTFYN---AAWDP-----SNRDRHIGIDVNSIKSVNTKSWKLQNGE---EANVVIAFNAATN
LECS_SOPJA      HSS--GYNSSYQIIAVDFDTHI----NAWDP------NTR-HIGIDVNSINSTKTVTWGWQNGE---VANVLISYQAATE
LEC_LENCU       NSK--EYDKTSQTVAVEFDTFYN---AAWDP-----SNKERHIGIDVNSIKSVNTKSWNLQNGE---RANVVIAFNAATN
LEC1_LABAL      CSS--DYNSSNQIIAVEFDTYFGKAYNPWDP------DFK-HIGVDVNSIKSIKTVKWDWRNGD---VANVVITYRAPTK
LEC_VICFA       NGK--DYDKTAQTVAVEFDTFYN---AAWDP-----SNGKRHIGIDVNTIKSISTKSWNLQNGE---EAHVAISFNATTN
LEC_ERYCG       NNS--KQDNSYQTLAVEFDTFS----NPWDP------PQVPHIGIDVNSIRSIKTQPFQLDNGQ---VANVVIKYDAPSK
LECA_DOLLA      PNAAESGIAESNVVAVEFDTDYLN-PDYGDP-------NYIHIGIDVNSIRSKVTASWDWQNGK---IATAHISYNSVSK
LEC_BOWMI       FGSSNNAGSDNGVVAVEFDT-YPN-TDIGDP-------NYRHIGIDVNSIRSKAASKWDWQNGK---TATAHISYNSASK
LEC_ERYCO       NNS--KQDNSYQTLGVEFDTFS----NPWDP------PQVPHIGIDVNSIRSIKTQPFQLDNGQ---VANVVIKYDASSK
LEC_SOYBN       NEN----ESGDQVVAVEFDTFR----NSWDP------PNP-HIGINVNSIRSIKTTSWDLANNK---VAKVLITYDASTS
ARC2_PHAVU      NTAN--YDRDAHTVAVVFDTVSNR-----------------IEIDVNSIRPIATESCNFGHNN-GEKAEVRITYYSPKN

    DLL-II      LLVASLVHPS-----QKTSFIVSERVDLTSVLPEWVSVGFSATTGLSKGYVETNEVLSWSFASKISINKEDE----------
LEC1_DOLBI      LLVASLVHPS-----RRTSYILSERVDITNELPEYVSVGFSATTGLSEGYIETHDVLSWSFASKLPDDSTAEPLDLASYLVRNVL-
LEC5_DOLBI      LLVASLVHPS-----RRTSYIVSERVDITNELPEYVSIGFSATTGLSEGYTETHDVLSWSFASKLPDDSTTEPLDIASYLVRNVL-
PHAE_PHAVU      LLVASLVYPS-----LKTSFIVSDTVDLKSVLPEWVIVGFTATTGITKGNVETNDILSWSFASKLSDGTTSEALNLANFALNQIL-
LEC_PEA         VLTVSLTYPNSLEEENVTSYTLSDVVSLKDVVPEWVRIGFSATTG---AEYAAHEVLSWSFHSELSGTSSSKQAADA---------
LECS_SOPJA      TLTVSLTYPS----SQTSYILSAA-VDLKSILPEWVRVGFTAATGLTTQYVETHDVLSWSFTSTLETGDCGAKDDNVHLVSYAFI-
LEC_LENCU       VLTVTLTYPNSLEEENVTSYTLNEVVPLKDVVPEWVRIGFSATTG---AEFAAHEVHSWSFHSELGGTSSSKQAADA---------
LEC1_LABAL      SLTVSLSYPS----DQTSNIVTAS-VDLKAILPEWVSVGFSAGVGN-AAKFN-HDILSWYFTSNLEPNNPAVNQAQ----------
LEC_VICFA       VLSVTLLYPN------LTGYTLSEVVPLKDVVPEWVRIGFSATTG---AEYATHEVLSWTFLSELTGPSN----------------
LEC_ERYCG       ILHVVLVYPS-----SGAIYTIAEIVDVKQVLPDWVDVGLSGATGAQRDAAETHDVYSWSFQASLPE------------------
LECA_DOLLA      RLSVTTYYPG------RGKPATSYDIELHTVLPEWVRVGLSASTG---QNIERNTVHSWSFTSSLWTNVAKVGVASISG-------
LEC_BOWMI       RLSVVSSYPN------SSPVVVSFDVELNNVGPPDVRVGFSATTG---QYTQTNNILAWSFRSSLMGYQAN---------------
LEC_ERYCO       ILHAVLVYPS-----SGAIYTIAEIVDVKQVLPEWVDVGLSGATGAQRDAAETHDVYSWSFQASLPETNDAVIPTSNHNTFAI---
LEC_SOYBN       LLVASLVYPS-----QRTSNILSDVVDLKTSLPEWVRIGFSAATG-LDIPGESHDVLSWSFASNLPHASSNIDPLDLTSFVLHEAI
ARC2_PHAVU      DLRVSLLYPS-----SEEKCHVSATVPLEKEVEDWVSVGFSATSGSKKETTETHNVLSWSFSSNFIN-FEGKKSERSNILLNKIL-
```

Fig. 3 Comparison of the amino acid sequence of DLL- II with other legume lectins. Comparison of the complete amino acid sequence of DLL- II β subunit with those of other leguminous lectins from (LEC5_DOLBI, LEC1_DOLBI) *Dolichos biflorus*, (PHAE_PHAVU, ARC2_PHAVU) *Phaseolus vulgaris*, (*LEC*_PEA, LEC_VICFA) *Vicia faba*, (LECS_SOPJA,) *Sophora japonica*, (LEC_LENCU) *Lens culinaris*, (LEC1_LABAL) *Laburnum alpinum*, (LECA_DOLLA) *Dolichos lablab*, (LEC1_LABAL) *Laburnum alpinum*, (LEC_ERYCO) *Erythrina corallodendron*, (LEC_BOWMI) *Bowringia mildbraedii* (LEC_SOYBN) *Glycine max*

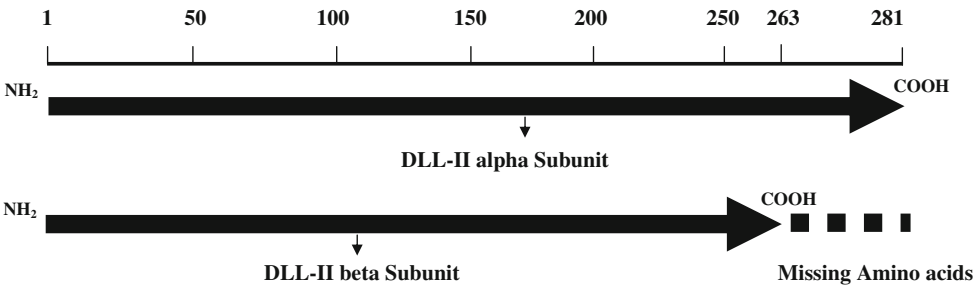

Fig. 4 Diagrammatic representation of DLL-II lectin structure. The α and β subunits are shown separately

Having the primary sequence at hand allowed us to make a comparison of the DLL-II sequences with the glucose/mannose specific lectin (DLL-I) that has been sequenced from field beans and also with the other legume lectins whose sequences are known [10]. Most lectins contain subunits that are identical or nearly identical in molecular weight; these are called one-chain lectins, which are mostly tetramers. Some lectins are composed of two subunits, one heavy and light chain, forming tetramers of the type $\alpha_2\beta_2$ and are called two chain lectins. Subunit architecture of *Dolichos lablab* lectins revealed that they are two chain lectins, DLL- I with non-identical subunit pattern of molecular mass (α, 15,000; β, 12,000) [10] and DLL- II with a pattern (α, 31,000; β, 29,000) differing at the carboxy terminus and being probably present as tetramers of the type $\alpha_2\beta_2$. Subunit heterogeneity is a common feature to a number of legume lectins [3, 21] and must be considered in studies concerning the analysis of their structure and function. The isolation of the subunits of DLL-II in the present study has enabled the direct characterization of these two very similar polypeptides. The strategy adopted in this study clearly demonstrates the close identity of both subunits of DLL-II. This finding is supported by the almost identical amino acid composition of the two subunits and our previous report wherein we have shown that an antibody raised against the α-subunit also cross-reacts with the β-subunit suggesting possible antigenic/sequence similarities between the two subunits [11].

De novo and manual sequencing of the spectra's of the proteolytic and chemical digests of the α and β subunits and subsequent database searches allowed us to look for the conserved regions of the lectin. From Fig. 3, it is evident that the sequence of DLL-II is highly homologous to a number of legume lectins (the homologous regions are marked in green). Of specific interest are the long stretches

Table 1 Sequence identity of DLL-II with other legume lectins

S. No	Species	Abbreviation	Sugar Group Specificity	Tribe	Identity (%)
1	*Dolichos biflorus* (Horse gram)	–	I and VII	*Phaseoleae*	62.6
2	*Dolichos biflorus* (Horse gram)	SL	I and VII	*Phaseoleae*	59.3
3	*Leucoagglutinating phytohemagglutinin*	PHA-L	II	*Phaseoleae*	58.1
4	*Erythroagglutinating phytohemagglutinin*	PHA-E	II	*Phaseoleae*	55.3
5	*Leucoagglutinating phytohemagglutinin*	PHA-L	II	*Phaseoleae*	56.8
6	*Glycine max* (Soybean)	SBA	I & VII	*Phaseoleae*	50.8
7	*Cytisus scoparius* Scotch broom)	CSII	IV	*Genisteae*	50.6
8	*Lathyrus ochrus* (Yellow-flowered pea)	–	II	*Vicieae*	49.5
9	*Pisum sativum* (Garden pea)	PSL1	II	*Vicieae*	46.3
10	*Lens culinaris* (Lentil)	–	II	*Vicieae*	45.9
11	*Vicia faba* (Broad bean)	Favin Lectin	V	*Vicieae*	45.7
12	*Griffonia simplicifolia*	GS4	V	*Cercideae*	44.0
13	*Erythrina corallodendron* (Coral tree)	ECorL	I & VII	*Phaseoleae*	41.9
14	*Laburnum alpinum* (Scotch laburnum)	LAA-I	III	*Genisteae*	41.9
15	*Cladrastis lutea* (Yellow wood)	LecClAII	II	*Sophoreae*	40.8
16	*Bowringia mildbraedii*	BMA	X	*Sophoreae*	39.2
17	*Dolichos lab lab* (Field bean)	–	II	*Phaseoleae*	36.4
18	*Phaseolus vulgaris* (Kidney bean) (French bean)	–	V	*Phaseoleae*	46.5
19	*Sophora japonica* (Japanese pagoda tree)	LECSJABG	VII	*Sophoreae*	49.2
20	*Canavalia ensiformis* (Jack bean)	Con A	II	*Diocleae*	35.7

I Galactose, *II* Glucose/Mannose, *III* Di-*N*-acetylchitobiose, *IV* 2-deoxygalactose, *V* *N*-acetylglucosamine, *VI* *N*-acetyllactosamin, *VII* *N*-acetylgalactosamine, *X* oligosaccharides

of sequences, such as the AFYTTPI, ADGLAFVPVG, and HIGIDVNSIK, shown in the Fig. 3. These results and those presented in Table 1, that describes the percentage of sequence identity of the DLL-II to other legume lectins, suggests strongly that the DLL-II belongs to the class of legume lectins that are known to be conserved. Furthermore, in our earlier studies by Edman degradation of the isolated subunits of DLL-II, we reported that the first ten amino acids of both subunits are identical [11].

From literature it is known that several other legume lectins, including concanavalin A (Con A) and the pea lectin have been found to undergo posttranslational processing of amino acids at their COOH termini during biosynthesis [21]. In ConA, this processing accompanies the posttranslational ligation that results in the formation of the circularly permuted mature form of this lectin [22]. The seed lectin from *Dolichos biflorus* is a heterotetramer and these heteromeric subunits 1 and 2 appear to differ only at the carboxy termini, and subunit 2 arises by proteolytic cleavage of the subunit 1. It has been suggested that differential cleavage of the COOH terminus of the α chain of the pea lectin may give rise to the molecular forms of this lectin that are commonly found in peas [23]. DLL-II might also undergo posttranslational processing as, addition of protease inhibitors during isolations and variations in time and type of extraction procedures have failed to alter the subunit stoichiometry, leading to the conclusion that the conversion of these subunits possibly occurs *in vivo*. The derivation of the β subunit of DLL-II from its α subunit resembles the origin of subunit II of the *Dolichos biflorus* seed lectin and stem/leaf lectins from subunit I [24–29]. In the case of these both lectins, only about half of the subunits are converted resulting in the production of the heterooligomeric proteins [24, 30]. These results suggested that the β subunit was derived from the α subunit by posttranslational modification/processing. Truncation at the carboxyl end is well known in certain leguminous lectins, resulting in a mixture of isoforms with different chain lengths. In these cases, cleavage performed by carboxypeptidase(s) and the digested products are sometimes heterogeneous. The present data support this interpretation, and heterogeneity of the COOH terminus of the β subunit suggesting that proteolytic conversion may occur by sequential removal of individual amino acids from its COOH terminus rather than a single endoproteolytic cleavage.

Previous studies from our laboratory have shown that both subunits of the purified galactose-specific lectin showed positive staining for carbohydrate when analyzed by PAS reagent, suggesting that both subunits contain carbohydrate. Supporting this notion we noticed two motifs for *N*-glycosylation consensus sites in each subunit, Asn^{25}–Tyr^{28} and Asn^{139}–Asn^{142}, when analyzing the primary sequences of both subunits for biologically significant sites and patterns with the ScanProsite tool available at http://www.expasy.org.

In addition, the partial amino acids sequences of both α and β subunits indicate that they are identical at the amino terminal region with a sequence NNLISFTMKR [11]. Trypsin treatment of the α and β subunits also gave the same peptide NNLISFTMK that was overlapped with the C^1 peptide which was similar to the NH_2 terminal sequence reported earlier. Chemical modification studies with native lectin carried out earlier revealed that histidine is important for agglutinating activity [11]. On the other hand the DLL-I that was extensively characterized shows the importance of aromatic amino acids such as the tyrosine and tryptophan for agglutinating activity [10 and references therein]. It is interesting to note that the DLL-I and DLL-II differ in their sugar specificity as well as in the role of specific amino acids in defining the agglutinating activity although both were isolated from the same seeds. A comparison of the amino acid sequences of a variety of legume lectins belonging to Gal/GalNAc specific category have shown extended carboxy-terminal sequences [31]. The isolated subunits showed much weaker agglutinating activity with untreated and trypsin treated rabbit erythrocytes (data not shown) than the intact protein, thus confirming the importance of interaction of both subunits for this activity and other functions of the lectin.

Crystal structures of many legume lectins with their ligand specificity (ConA, LCA, PSA for Man/Glc type; DBA, EcorL, for Gal/GalNAc type etc) have been solved and their structure-function relationships; *i.e.*, assignment of key amino acid residues for binding to a specific sugar, have been discussed well in literature [32].

The protein sequence obtained for DLL-II reveals, maximum homology with the leguminous family member *Dolichos biflorus* and *phaseolus vulgaris* which have been crystallized and well characterized. The near sequence identity of both the subunits except at the carboxy-terminus lead to an understanding of the conservation of the amino acids among the legumes which is quite a common phenomenon in lectins and other vacuolar proteins [33].

A recent study reported that the *Dolichos biflorus* seed lectin also exhibited lipoxygenase activity [34]. This prompted us to analyze the purified DLL-II for lipoxygenase activity and it is interesting to note from the preliminary studies that the DLL-II also exhibited this enzyme activity (data not shown). In view of the structural relatedness of the DLL-II to the *Dolichos biflorus* lectin, it would be interesting to validate these observations further.

In summary, the present study is the first report on the primary sequence determination of the newly characterized DLL-II from the seeds of *Dolichos lablab*. The study presents the potential application of utilizing the advances in mass spectrometry, proteolytic digestions, MS/MS

analysis for *de novo* sequencing and database search with different programs as powerful tools for the identification, characterization and determination of the primary sequences of proteins. Furthermore the study also gave new insights towards understanding the structural and functional differences between the DLL-I and II lectins, and suggests that DLL-II exhibits unique properties in its structure and might exhibit distinct functional properties as well which need to be evaluated. In a preliminary study carried out in our laboratory we found that both lectins together with specific glycosidases such as the α-mannosidase are localized in the protein bodies of the seeds. This common localization of these proteins suggests possible specific *in vivo* interactions of these proteins which may contribute towards the physiological significance of the lectins during plant growth and development. With the available information on the DLL lectins, and to precisely understand their physiological significance, a detailed study on the protein–protein interactions of the protein body components from the seeds need to be carried out which is the future direction of work in our laboratory, that will eventually form the basis to analyze the potential applications of the DLL lectins.

Acknowledgements The financial support by the Deutscher Akademischer Austausch Dienst (DAAD) to R. Nagender Rao (April 2006 to June 2006) at the University of Greifswald, Germany, is gratefully acknowledged. We are indebted to Dr. Elke Hammer for support in some of the MS experiments, Dr. Holger Hippe (Chromatec, Greifswald) for some of the AA-composition-analyses and also thank the proteomics staff of CCMB for the support in some MS experiments. RNR thanks CSIR (New Delhi), University of Hyderabad.

References

1. Boyd, W.C., Shapleigh, E.: Specific precipitating activity of plant agglutinins (Lectins). Science **119**, 419 (1954). doi:10.1126/science.119.3091.419
2. Goldstein, I.J., Hughes, R.C., Monsigny, M., Osawa, T., Sharon, N.: What should be called a lectin? Nature **285**, 66 (1980). doi:10.1038/285066b0
3. Goldstein, I.J., Poretz, R.D.: The Lectins. In: Liener, I.E., Sharon, N., Goldstein, I.J. (eds.) Properties, functions and applications in biology and medicine. Acadmeic, New York (1986)
4. Etzler, M.E.: Plant lectins: molecular and biological aspects. Annu. Rev. Plant Physiol. **36**, 209–234 (1985). doi:10.1146/annurev.pp.36.060185.001233
5. Hoffman, L.M., Donaldson, D.D.: Characterization of two Phaseolus vulgaris phytohemagglutinin genes closely linked on the chromosome. EMBO J. **4**, 883–889 (1985)
6. Etzler, M.E.: Isolation and characterization of subunits of DB58, a lectin from the stems and leaves of *Dolichos biflorus*. Biochemistry **33**, 9778–9783 (1994). doi:10.1021/bi00198a049
7. Hemperley, J.J., Mostov, K.E., Cunningham, B.A.: *In vitro* translation and processing of a precursor form of Favin, a lectin from Vicia faba. J. Biol. Chem. **257**, 7903–7909 (1982)
8. Foriers, A., Lebrum, E., van Rapenbusch, R., de Neve, R., Strosberg, A.D.: The structure of the lentil (*Lens culinaris*) lectin. Amino acid sequence determination and prediction of the secondary structure. J. Biol. Chem. **256**, 5550–5560 (1981)
9. Higgins, T.J.V., Chandler, P.M., Zurawski, G., Button, S.C., Spencer, D.: The biosynthesis and primary structure of pea seed lectin. J. Biol. Chem. **258**, 9544–9549 (1983)
10. Gowda, L.R., Savithri, H.S., Rajagopal Rao, D.R.: The complete primary structure of a unique mannose/glucose specific lectin from the field bean (*Dolichos lablab*). J. Biol. Chem. **269**, 18789–18793 (1994)
11. Latha, V.L., Rao, R.N., Nadimpalli, S.K.: Affinity purification, physicochemical and immunological characterization of a galactose-specific lectin from the seeds of *Dolichos lablab* (Indian lablab beans). Protein Expr. Purif. **45**, 296–306 (2006). doi:10.1016/j.pep.2005.06.010
12. Sultan, N.A.M., Rao, R.N., Nadimpalli, S.K., Swamy, M.J.: Tryptophan environment, secondary structure and thermal unfolding of the galactose-specific seed lectin from *Dolichos lablab:* Fluorescence and circular dichroism spectroscopic studies. Biochim. Biophys. Acta **1760**, 1001–1008 (2006)
13. Rao, R.N., Nadimpalli, S.K.: An efficient method for the purification and quantification of a galactose-specific lectin from vegetative tissues of *Dolichos lablab*. J. Chromatogr. B. **861**, 209–211 (2007)
14. Shadforth, I., Todd, K., Crowther, D., Bessant, C.: Protein and peptide identification algorithms using MS for use in high-throughput, automated pipelines. Proteomics **5**, 1787–1796 (2005). doi:10.1002/pmic.200401058
15. Perkins, D.N., Pappin, D.J.C., Creasy, D.M., Cottrell, J.S.: Probability-based protein identification by searching sequence databases using mass spectrometry. Electrophoresis **20**, 3551–3567 (1999). doi:10.1002/(SICI)1522-2683(19991201)20:18<3551::AID-ELPS3551>3.0.CO;2-2
16. Saris, C.J.M., Eenbergen, J.V., Jenks, B.G., Bloemers, H.P.J.: Hydroxylamine cleavage of proteins in polyacrylamide gels. Anal. Biochem. **132**, 54–67 (1983). doi:10.1016/0003-2697(83)90425-6
17. Thiede, B., Wittmann-Liebold, B., Michael, B., Krause, E.: A method to define the carboxyl terminal of proteins. FEBS Lett. **357**, 65–69 (1995). doi:10.1016/0014-5793(94)01323-S
18. Shevchenko, A., Sunyaev, S., Loboda, A., Shevchenko, A., Bork, P., Ens, W., *et al.*: Charting the proteomes of organisms with unsequenced genomes by MALDI-Quadrupole time-of-flight mass spectrometry and BLAST homology searching. Anal. Chem. **73** (9):1917–1926 (2001). doi:10.1021/ac0013709
19. Kaku, H., Tanaka, Y., Kiyoshi, T., Minami, E., Mizuno, H., Naota, S.: Sialylated oligosaccharide-specific plant lectin from Japanese elderberry (*Sambucus sieboldiana*) bark tissue has a homologous structure to type II ribosome-inactivating proteins, ricin and abrin. J. Biol. Chem. **271**, 1480–1485 (1996). doi:10.1074/jbc.271.3.1480
20. Matis, M., Mavric, M.Z., Katalinic, J.P.: Mass spectrometry and database search in the analysis of proteins from the fungus *Pleurotus ostreatus*. Proteomics **5**, 67–75 (2005). doi:10.1002/pmic.200400904
21. Etzler, M.E.: In: Allen, H.J., Kisailus, E.C. (eds.) Glycoconjugates: Composition, structure, and function, pp. 521–539. Marcel-Dekker, New York (1992)
22. Bowles, D.J., Marcus, S.E., Pappin, D.J.C., Findlay, J.B.C., Eliopoulos, E., Maycox, P.R., *et al.*: Posttranslational processing of concanavalin A precursors in jackbean cotyledons. J. Cell Biol. **102**, 1284–1297 (1986). doi:10.1083/jcb.102.4.1284
23. Stubbs, M.E., Carver, J.P., Dunn, R.J.: Production of pea lectin in Escherichia coli. J. Biol. Chem. **261**, 6141–6144 (1986)
24. Carter, W.G., Etzler, M.E.: Isolation and characterization of subunits from the predominant form of *Dolichos biflorus* lectin. Biochemistry **14**, 2685–2689 (1975a). doi:10.1021/bi00683a019

25. Carter, W.G., Etzler, M.E.: Isolation, characterization, and subunit structures of multiple forms of *Dolichos biflorus* lectin. J. Biol. Chem. **250**, 2756–2762 (1975b)
26. Etzler, M.E., Gupta, S., Borrebaeck, C.: Carbohydrate binding properties of the *Dolichos biflorus* lectin and its subunits. J. Biol. Chem. **256**, 2367–2370 (1981)
27. Roberts, D.M., Walker, J., Etzler, M.E.: A structural comparison of the subunits of the *Dolichos biflorus* seed lectin by peptide mapping and carboxyl terminal amino acid analysis. Arch. Biochem. Biophys. **218**, 213–219 (1982). doi:10.1016/0003-9861(82)90338-1
28. Schnell, D.J., Etzler, M.E.: Primary structure of the *Dolichos biflorus* seed lectin. J. Biol. Chem. **262**, 7220–7225 (1987)
29. Quinn, J.M., Etzler, M.E.: *In vivo* biosynthetic studies of the *Dolichos biflorus* seed lectin. Plant Physiol. **91**, 1382–1386 (1989)
30. Talbot, C.F., Etzler, M.E.: Isolation and characterization of a protein from leaves and stems of *Dolichos biflorus* that cross reacts with antibodies to the seed lectin. Biochemistry **17**, 1474–1479 (1978). doi:10.1021/bi00601a018
31. Sharon, N., Lis, H.: Legume lectins—a large family of homologous proteins. FASEB J. **4**, 3198–3208 (1990)
32. Manoj, K., Suguna, K.: Signature of quaternary structure in the sequences of legume lectins. Protein Eng. **14**, 735–745 (2001). doi:10.1093/protein/14.10.735
33. Parkin, S., Rupp, B., Hope, H.: Atomic resolution structure of A at 120 K Concanavalin. Acta Crystallogr D. **52**, 1161–1168 (1996). doi:10.1107/S0907444996009237
34. Roopashree, S., Singh, S.A., Gowda, L.R., Rao, A.G.: Dual-function protein in plant defence: seed lectin from *Dolichos biflorus* (horse gram) exhibits lipoxygenase activity. Biochem. J. **395**(3):629–639 (2006). doi:10.1042/BJ20051889