CHAPTER II

AMINO ACID SEQUENCE DETERMINATION
OF
Sclerotium rolfsii LECTIN
BY
MASS SPECTROMETRY
A novel TF-antigen specific lectin from *Sclerotium rolfsii* (SRL) has been crystallized by the hanging drop vapour–phase diffusion method and its preliminary X-ray crystallographic data reported earlier, from this lab jointly with Leonidas (Leonidas *et al.*, 2003). The three-dimensional crystal structure of SRL in free form and in complex with N-acetyl-D-galactosamine (GalNAc) and N-acetyl-D-glucosamine (GlcNAc) has been determined at 1.1 Å, 2.0 Å and 1.7 Å resolutions respectively. The protein structure is composed of two β-sheets, which consist of four and six β-strands, connected by two α-helices. Sequence and structural comparisons reveal that SRL is the third member of a newly identified family of fungal lectins, which include lectins from *Agaricus bisporus* (Crenshaw *et al.*, 1995) and *Xerocomus chrysenteron* (Trigueros *et al.*, 2003) that share a high degree of structural similarity and carbohydrate specificity. The data of the free SRL are of the highest resolution obtained for any protein of this family. The crystal structures of the SRL in complex with two carbohydrates, GalNAc and GlcNAc, which differ only in the configuration of a single epimeric hydroxyl group, provide the structural basis for its carbohydrate specificity. SRL has two distinct binding sites, a primary and a secondary. GalNAc binds at the primary site, whereas GlcNAc binds only at the secondary site. Moreover, SRL has the ability to recognize and probably bind at the same time two different carbohydrate structures. Structural comparison with *Agaricus bisporus* lectin-carbohydrate
complexes by super positioning revealed that the primary site is also able to bind the Thomsen-Friedenreich antigen (Galβ1-3GalNAc-α-glycan structures) whereas secondary site cannot (Leonidas et al., 2007).

The primary structure of SRL has not been previously determined but the high resolution of the diffraction data and the parallel independent refinement of several structures allowed us to deduce the sequence directly from the electron density maps. The potential problems in X-ray crystallographic studies of proteins are the ambiguities in amino acid sequence deduced based on electron density map (EDM). Further more, amino acid side chains often cannot be reliably identified from the electron density maps. However, from those crystal structures solved and refined at very high resolution, the amino acid sequence deduced from the electron density maps could be reliable. Indeed the ambiguities such as differentiating Asn /Asp and Gln / Glu cannot be ascertained conclusively, since the difference in density between the COOH and CONH₂ groups is insufficient (Rees and Offord, 1972). This problem was encountered for interpreting the EDM data of SRL from its X-ray crystal structure and warranted for confirmation of the amino acid sequence.

Over the past two decades, a vast majority of primary structures of Proteins have been determined either by the stepwise, automated Edman degradation or indirectly, by the translation of the DNA sequence of the corresponding gene coding for the protein (Johnson and Biemann, 1987).
Traditionally, Edman sequencing was sole tool used for the de novo sequence analysis of unknown proteins, largely because the sequencing could be carried out directly on the intact protein (Walsh et al., 1981), however, it has serious limitations with regard to characterize structurally modified peptides and proteins and impure proteins (Sharon and Lis, 1982; Tsunasawa and Sakiyam, 1984; Yazdanparast et al., 1987). Many commonly occurring modifications are lost or destroyed by the Edman degradation because they do not survive the harsh cleavage and derivatization conditions employed (Allen, 1981; Glazer et al., 1982; Wold and Moldave, 1984). Since the N-terminal residue of SRL was blocked due to acetylation our attempts to sequence by Edman’s method failed.

In contrast to the conventional methodologies, alternative method based on entirely different approach are highly appreciated as they provide a way to overcome obstacles in the course of the conventional approach. Mass spectrometric identification and sequencing of protein is such an alternative, which has been particularly useful for sequencing of N-blocked and post translationally modified proteins, which are not suitable for the Edman’s method (Hunt et al., 1986; Biemann, 1990).

With the introduction of new soft ionization techniques, such as matrix-assisted laser desorption (MALDI) and electro spray ionization (ESI) mass spectrometry is rapidly developing as a powerful and convenient method for structural analysis of peptides and proteins. Because
of its unique features such as flexibility, increased speed, sensitivity, reliability, accuracy and the ability to deal with peptide mixtures make mass spectrometry an advantageous method to determine molecular weight information at picomole concentrations of peptides and proteins independent of covalent modifications that may be present (Aebersold and Mann, 2003), locate and correct errors in DNA and cDNA deduced protein sequences (Levine et al., 1987; Giffin et al., 1989; Furuya et al., 1989; Pearson et al., 1990; Stults et al., 1990), identify numerous types of posttranslational modifications as well as to verify the fidelity in translation of recombinant proteins (Carr et al., 1988; Schweppe et al., 2003).

To confirm and verify the amino acid sequence deduced from X-ray crystallography as well as to resolve ambiguities in assigning Asn/Asp and Gln/Glu, matrix-assisted laser desorption ionization (MALDI) and electrospray ionization (ESI) techniques have been employed. Keefe and co-workers have successfully demonstrated the use of matrix-assisted laser desorption/ionization mass spectrometry (MALDI-MS) for resolving the ambiguities of amino acid sequence in X-ray crystallographic structure of mutant Staphylococcal nuclease (Keefe et al., 1992).

The present chapter describes the mass spectrometric strategies using matrix-assisted laser desorption/ionization (MALDI) and electrospray ionization (ESI) employed for the determination of amino acid sequence of SRL, which includes:
• In-gel digestion of SRL with trypsin and chymotrypsin.
• Analysis of tryptic peptides using MALDI-MS.
• Determination of amino acid sequence of tryptic and chymotryptic peptides of SRL using tandem mass spectrometry (LC-ESI-MS/MS).

Based on the mass spectrometric results, the resolved and confirmed amino acid sequence of Sclerotium rolfsii lectin is presented.

2.1. Materials and methods

Sclerotium rolfsii lectin (SRL) was purified from the sclerotial bodies by the established purification protocol as described earlier (Swamy et al., 2001). TPCK-treated trypsin (bovine) and TLCK-treated α-chymotrypsin (bovine) were obtained from Sigma Chemical Co., St.Louis, MO, USA. 2, 5-dihydroxybenzoic acid (DHB, 98% pure), acetonitrile (ACN >99.93% pure, HPLC grade), formic acid (FA 96% pure) and trifluoroacetic acid (TFA, >99% pure) were purchased from Aldrich (Milwaukee, WI, USA). Ammonium bicarbonate (NH$_4$HCO$_3$) was obtained from Fischer (Fair Lawn, NJ). All aqueous solutions were prepared using deionized water from Milli-Q Ultrapure Water System (Bedford, MA) and filtered through 0.22μm nylon filters from Fischer Scientific (Pittsburgh, PA, USA).

2.2.1. Sample preparation for MALDI-TOF-MS and LC-ESI-MS/MS

Purified SRL was subjected to electrophoresis on SDS PAGE (12%) and the protein band appeared after staining the gel with Coomassie blue was excised. Lectin band in the excised gel was subjected to in-gel tryptic /
chymotryptic digestion as described by Rosenfeld, et al., (1992) with some modifications. Briefly, the excised gel was sliced to small pieces, transferred to sterile siliconized micro centrifuge tube and destained by repeated washing with 50mM NH₄HCO₃, dehydrated using acetonitrile and finally dried by speed-vac. Enzymatic digestion (enzyme: substrate; 1:10) was carried out by incubating the dried gel pieces with trypsin / chymotrypsin (100μg in 1 ml of 25 mM NH₄HCO₃ containing 0.5mM CaCl₂, pH 8.0) initially for 1 h on ice. Further incubation was carried out overnight at 37°C after adding additional 40μl of bicarbonate buffer. Digested samples were centrifuged briefly to collect the supernatants and the settled gels were extracted either with 5% formic acid in 50% acetonitrile (for ESI-MS/MS) or 5% trifluoroacetic acid in 50% acetonitrile (for MALDI-TOF-MS). Extraction step was repeated 3-4 times using fresh extraction solution and all the wash supernatants (tryptic or chymotryptic digests) were pooled and concentrated to 50μl by speed-vac and used for MS analysis.

2.2.2. Mass spectrometry

2.2.2.1. MALDI-TOF-MS: MALDI-TOF-MS analysis was carried out with Ultraflex TOF/TOF (Bruker Daltonics, Bremen, Germany) mass spectrometer, equipped with nitrogen laser (337 nm). The enzyme digest samples were mixed with equal volume of saturated matrix solution (2, 5-dihydroxybenzoic acid in 50% acetonitrile /H₂O with 0.1% trifluoroacetic acid). This mixture (1μl) was deposited on the probe plate and dried by
stream of air and the spectra were recorded in the reflectron positive ion mode using Bruker Daltonics FLEX control software and the spectral data were processed by Bruker Daltonics FLEX analysis software. A standard peptide mixture (P.N: 206195, Bruker peptide calibration standard) was used for external calibration.

2.2.2.2. HPLC-ESI-MS/MS: On-line HPLC separation of tryptic and chymotryptic digests of SRL was carried out on HP1100 (Agilent) at a flow rate of 0.150ml / min. Solvent system consisting sol. A; 0.1% formic acid in water and sol. B; acetonitrile in 0.1% formic acid was used. Tryptic peptides were separated on a C8 reverse-phase column (4.6 x 150 mm; ZORBAX RX-C8, 5μm, Agilent) using a linear gradient of sol. B 20% to 95% achieved in 45 min.

The chymotryptic peptides were fractionated on 4.6 x 250 mm, C18 reverse-phase column (Phenomenex, NAXSIL, 5μm CTB) with a linear gradient from 5% solvent B to 95% solvent B in 75 mins. ESI-MS/MS data were obtained using an Esquire 3000 plus mass spectrometer (Bruker Daltonics, Germany) consisting of two octopoles followed by an ion trap. Nitrogen and helium were used as nebulizer and collision gas for collision induced dissociation (CID) experiments respectively. Fragmentation data were acquired over a range of 50-2800 m/z in positive ion mode and analyzed using Esquire data analysis software (version 3.1). Acquired MS/MS spectra were interpreted manually.
2.3. Results and discussion

In order to verify and resolve the ambiguities in X-ray derived amino acid sequence of SRL, we sequenced the SRL by MALDI-MS and LC-ESI-MS/MS (tandem mass spectrometry), as chemical and cDNA sequence are not available. The original amino acid sequence of SRL deduced from high-resolution electron density map reported earlier (Leonidas et al., 2007) is shown in figure 1.

Ac-TYKITVRVYQTNPNAFFHPVEKTVWKYANGGTWTITDDQHVLT MGGSGTSGTLRFHADNGESFTATFGVHNYKRWCIVTNLAADET GMVINQYYQSQKNREEARERQLSNYEVKNAKGRNFEIVYTEAEG NDLHANLIIG-COOH

**Figure 1.** Amino acid sequence of SRL deduced from X-ray crystallographic studies at 1.1 Å resolution.

The expected proteolytic peptides and fragmentation ion masses of the amino acid sequence of SRL deduced from X-ray crystallography were calculated using programs http://delphi.phys.univ_tours.fr/pyrolysi/cutter.html and http://hodgkin.mbu.iisc.ernet.in/~pfia/index.htm. From the X-ray crystallographic analysis it was clear that the N-terminal of SRL is blocked which further confirmed our failure to sequence intact SRL by Edman degradation. Hence we decided to determine the sequence of SRL by mass spectrometry from the tryptic and chymotryptic in-gel digests by mass spectrometry. The purified protein was subjected to in-gel digestion with
trypsin and chymotrypsin and the digests were used for mass spectrometric analysis.

2.3.1. Peptide mapping of tryptic digest of SRL by MALDI-TOF-MS and sequencing of tryptic peptides by LC-ESI-MS/MS

Unfractionated tryptic digest after the in-gel digestion of SRL was subjected to MALDI-TOF-MS to characterize the peptide mass pattern. MALDI-TOF-MS has been used as the preferable method because of its robustness in complex samples, tolerance to many biological buffers, salts and the prevalence of singly charged peaks. The expected, observed protonated monoisotopic masses of tryptic peptides obtained by MALDI-TOF-MS and their positions in the final sequence are shown in table 1. Of the total 15 tryptic peptides expected, 12 peptides were identified in MALDI-TOF-MS. Mass peaks corresponding to designated peptides T4, T5 and T8 were not detected, probably because of suppression of ions in the mixture or they were not recovered from the gel. When assigned tryptic peptides were further subjected to fragmentation analysis to derive sequence information, due to poor fragmentation sequence information could not be obtained. Despite the fact that the spectrum showed considerable number of fully resolved peaks (Figure 2), the approach of peptide mass mapping gave no positive hit.
Figure 2. Positive ion reflectron MALDI-TOF spectra of in-gel tryptic digest of the SRL. The samples were prepared, using dried droplet method, by mixing the sample with 2,5-dihydroxybenzoic acid as matrix. Peaks arising from the tryptic digest are labeled according to the sequence position in the protein. Inset illustrates the expanded spectrum of low mass region.

Hence, the tryptic digest of SRL was subjected to on-line LC-ESI-MS/MS and the reconstructed total ion current (TIC) chromatogram is presented in figure 3. Each of the peaks is labeled according to the corresponding tryptic peptides.
Table 1. Peptide fragments obtained from a tryptic digest of the SRL by MALDI-TOF-MS

<table>
<thead>
<tr>
<th>Tryptic Peptide</th>
<th>Amino acid no. a</th>
<th>Expected (M+H)+ b (Da)</th>
<th>Observed (M+H)+ (Da)</th>
<th>Deviation (Da)</th>
</tr>
</thead>
<tbody>
<tr>
<td>T1</td>
<td>1-3</td>
<td>453.0</td>
<td>453.31</td>
<td>+ 0.31</td>
</tr>
<tr>
<td>T2</td>
<td>4-7</td>
<td>488.32</td>
<td>488.15</td>
<td>+ 0.17</td>
</tr>
<tr>
<td>T3</td>
<td>8-22</td>
<td>1790.89</td>
<td>1790.69</td>
<td>+ 0.2</td>
</tr>
<tr>
<td>T4</td>
<td>23-26</td>
<td>533.31</td>
<td></td>
<td>*</td>
</tr>
<tr>
<td>T5</td>
<td>27-54</td>
<td>2896.35</td>
<td></td>
<td>*</td>
</tr>
<tr>
<td>T6</td>
<td>55-73</td>
<td>2141.97</td>
<td>2141.77</td>
<td>- 0.20</td>
</tr>
<tr>
<td>T7</td>
<td>74-74</td>
<td>175.12</td>
<td>174.69</td>
<td>+ 0.43</td>
</tr>
<tr>
<td>T8</td>
<td>75-99</td>
<td>2890.34</td>
<td></td>
<td>*</td>
</tr>
<tr>
<td>T9</td>
<td>100-101</td>
<td>289.16</td>
<td>289.81</td>
<td>- 0.65</td>
</tr>
<tr>
<td>T10</td>
<td>102-105</td>
<td>504.24</td>
<td>504.05</td>
<td>+ 0.19</td>
</tr>
<tr>
<td>T11</td>
<td>106-107</td>
<td>304.16</td>
<td>303.86</td>
<td>+ 0.30</td>
</tr>
<tr>
<td>T12</td>
<td>108-115</td>
<td>980.51</td>
<td>980.42</td>
<td>+ 0.09</td>
</tr>
<tr>
<td>T13</td>
<td>116-118</td>
<td>332.19</td>
<td>331.89</td>
<td>+ 0.30</td>
</tr>
<tr>
<td>T14</td>
<td>119-120</td>
<td>232.14</td>
<td>231.61</td>
<td>+ 0.53</td>
</tr>
<tr>
<td>T15</td>
<td>121-141</td>
<td>2332.15</td>
<td>2331.13</td>
<td>+ 1.02</td>
</tr>
</tbody>
</table>

a Assignments were made by comparing the observed peptide masses with the expected masses of the SRL derived from X-ray crystallographic data.
b Expected mass calculated from monoisotopic weights.
* Not observed.

ESI-MS/MS spectra of protonated tryptic peptides of SRL T1, T2, T3, T4, T5, T8, T10, T11 and T13 obtained from collision induced dissociation are shown in figure 4; a, b, c, d, e, f, g, h and i respectively.
Figure 3. Total ion current (TIC) chromatogram of tryptic digests of SRL.

Figure 4. ESI-MS/MS spectra of tryptic peptides of SRL.

4 (a). ESI-MS/MS spectrum of N-terminal tryptic peptide of SRL. The spectrum revealed acetylation of the N-terminus as indicated by shift of b3 ion by 42 Da. The sequence obtained is Ac-TYK (residues 1-3).
4 (b). ESI-MS/MS spectrum of the (M+H)$^+$ ion of the peptide ITVR (residues 4-7).

4 (c). ESI-MS/MS spectrum of the doubly charged ion m/z 896.3. The spectrum revealed the substitution of D for N at the fourteenth position and the sequence is VYQTNPDAFFHPVEK (residues 8-22).
4 (d). ESI-MS/MS spectrum spectra of the singly charged tryptic peptide (T4) TVWK (residues 23-26).

4 (e). ESI-MS/MS spectrum of the doubly charged ion m/z 1071.5 obtained from ESI-ion trap mass spectrometer. The spectrum revealed the sequence FHADNGESFTATFGVHNYK (residues 55-73).
4 (f). ESI-MS/MS spectrum of tryptic peptide (T7) of SRL. The sequence obtained is EEAR (residues 102-105).

4 (g). ESI-MS/MS spectrum of the doubly charged ion m/z 490.3 obtained from ESI-ion trap mass spectrometer. The spectrum revealed the substitution of Q for E at the 113th position and the corrected sequence is QLSNYQVK (residues 108-115).
4 (h). ESI-MS/MS spectrum of tryptic peptide (T11) of SRL. The sequence obtained is NAK (residues 116-118).

4 (i). ESI-MS/MS spectrum of the doubly charged ion m/z 1165.6 obtained from ESI-ion trap mass spectrometer. The spectrum revealed the substitution of Q for E and N for D at positions 123rd and 133rd respectively and the corrected sequence is NFQIVYTEAEGNNLHANLIIG (residues 121-141).
Table 2 shows the expected and observed protonated masses of tryptic peptides of SRL with their sequences determined from CID mass spectral data and positions in the final sequence. Peptides, T1, T2, T3, T4, T5, T8, T10, T11, and T13 were successfully sequenced based on their fragmentation spectra by CID. Because of the poor quality CID spectra sequences of peptides T6, T7, T9 and T12 were confirmed solely from their molecular masses. High molecular weight peptides, having m/z 2892.31 (residues 75-99) and 2896.35 (residues 27-54) were not observed in the LC-ESI-MS indicating probably the protein part constituting these peptides was resistant to tryptic digestion or they could not be extracted from the gel. Observed m/z values of sequence specific ions from the CID mass spectral analysis for peptides T1, T2, T3, T4, T5, T8 T10, T11 and T13 are presented in table 3. Fragmentation ions of protonated peptides produced by CID are labeled according to the standard nomenclature proposed by Biemann et al., (1990). Observed monoisotopic masses of the sequence defining ion series in each spectrum are shown directly above and below the respective sequence. Fragmentation ions of low mass (< 1/3 of the peptide mass) were not observed in the CID spectra, essentially because of the inherent limitation of the ion trap mass analyzer. Observed protonated masses of tryptic peptides having m/z 1791.6, 979.5 and 2330.2 corresponding to sequence residues 8-22, 108-115 and 121-141 were different from their
calculated masses by 396 ppm, 1031 ppm and 772 ppm respectively (Table. 2).

Table 2. Tryptic peptides of SRL detected by LC-ESI-MS/MS.

<table>
<thead>
<tr>
<th>Peptide No.</th>
<th>Amino acid No.</th>
<th>Expected (M+H)+b (Da)</th>
<th>Observed (M+H)+ (Da)</th>
<th>Charge state</th>
<th>Ret. Time</th>
<th>Error (ppm)</th>
<th>Peptide sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>T1</td>
<td>1-3</td>
<td>453.30</td>
<td>453.3</td>
<td>+1</td>
<td>26.1</td>
<td>0.00</td>
<td>Ac-TYKc</td>
</tr>
<tr>
<td>T2</td>
<td>4-7</td>
<td>488.32</td>
<td>488.3</td>
<td>+1</td>
<td>11.8</td>
<td>40.95</td>
<td>ITVRC</td>
</tr>
<tr>
<td>T3</td>
<td>8-22</td>
<td>1790.89</td>
<td>1791.6</td>
<td>+2</td>
<td>23.8</td>
<td>150.70*</td>
<td>VYQTPDAFFHPVEKc</td>
</tr>
<tr>
<td>T4</td>
<td>23-26</td>
<td>533.31</td>
<td>533.3</td>
<td>+1</td>
<td>15.2</td>
<td>18.75</td>
<td>TVWKc</td>
</tr>
<tr>
<td>T5</td>
<td>55-73</td>
<td>2141.97</td>
<td>2142.0</td>
<td>+2, +3, +4</td>
<td>21.5</td>
<td>14.00</td>
<td>FHADNGESFTATFGVHNYKc</td>
</tr>
<tr>
<td>T6</td>
<td>74-74</td>
<td>175.12</td>
<td>175.10</td>
<td>+1</td>
<td>39.0</td>
<td>114.22</td>
<td>Rm</td>
</tr>
<tr>
<td>T7</td>
<td>100-101</td>
<td>289.16</td>
<td>289.1</td>
<td>+1</td>
<td>41.0</td>
<td>207.54</td>
<td>NRm</td>
</tr>
<tr>
<td>T8</td>
<td>102-105</td>
<td>504.24</td>
<td>504.2</td>
<td>+1</td>
<td>10.5</td>
<td>79.33</td>
<td>EEARd</td>
</tr>
<tr>
<td>T9</td>
<td>106-107</td>
<td>304.16</td>
<td>304.2</td>
<td>+1</td>
<td>10.6</td>
<td>131.49</td>
<td>ERm</td>
</tr>
<tr>
<td>T10</td>
<td>108-115</td>
<td>980.51</td>
<td>979.5</td>
<td>+1, +2</td>
<td>13.4</td>
<td>20.41*</td>
<td>QLSNYQVKc</td>
</tr>
<tr>
<td>T11</td>
<td>116-118</td>
<td>332.19</td>
<td>332.2</td>
<td>+1</td>
<td>10.5</td>
<td>30.10</td>
<td>NAKc</td>
</tr>
<tr>
<td>T12</td>
<td>119-120</td>
<td>232.14</td>
<td>232.1</td>
<td>+1</td>
<td>10.6</td>
<td>172.33</td>
<td>GRm</td>
</tr>
<tr>
<td>T13</td>
<td>121-141</td>
<td>2332.15</td>
<td>2330.20</td>
<td>+2, +3</td>
<td>27.5</td>
<td>8.58*</td>
<td>NFQIVYTEAEGNLHANLIIGc</td>
</tr>
</tbody>
</table>

* Assignments were made by comparing observed peptide masses with a theoretical digest of the SRL derived from the X-ray crystallographic data.

b Expected mass calculated from monoisotopic molecular weights.

c Acetylation at the N terminus.

d Sequence deduced from CID spectrum and the corrected amino acids are indicated in bold.
m Sequences confirmed solely on molecular mass.

* Obtained after modification.
In order to account for these observed differences, changes were made with the residues N (14) → D, E (113) → Q, E (123) → Q and D (133) → N. With these changes made in each of the peptide, the error reduced to 150 ppm, 20 ppm and 8.5 ppm for peptides m/z 1791.6, 979.5 and 2330.2 respectively. The replacements at these positions were further confirmed from their CID spectral data, which yielded numerous sequence-specific daughter ions (Table 3. G, F, and I). Also the observed mass accuracy for all the tryptic peptides lie within + < 0.3 Dalton threshold limit required to differentiate unambiguously the Asn / Asp or Gln / Glu, which differ by 1 Dalton (Biemann, 1992). The CID spectrum of the peptide having m/z 453.3, represent the acetylated N-terminus peptide, shows mass of b₃ ion shifted by 42 Da accounting for N terminal acetylated threonine (Table 3. A). This is in agreement with our earlier observation that SRL is not amenable to Edman degradation, attempted during the earlier phase of its characterization.
Table 3. CID mass spectral data for tryptic peptides. The m/z values (found) for particular ion series are listed by rows.

<table>
<thead>
<tr>
<th>Ion Series</th>
<th>m/z Values</th>
<th>Peptide Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>(A) (M+H)⁺</td>
<td>m/z 453.3</td>
<td>Ac-The - Tyr - Lys - Asn - Ala - Lys - Thr - Val - Trp - Lys</td>
</tr>
<tr>
<td>(B) (M+H)⁺</td>
<td>m/z 332.2</td>
<td>Glu - Glu - Ala - Arg</td>
</tr>
<tr>
<td>(C) (M+H)⁺</td>
<td>m/z 488.3</td>
<td>Val - Tyr - Glu - Thr - Asn - Pro - Asp - Ala - Phe - Phe - His - Pro - Val - Glu - Lys</td>
</tr>
<tr>
<td>(D) (M+H)⁺</td>
<td>m/z 533.3</td>
<td>Phe - HU - Ala - Asp - Asn - Phe - Gly - Glu - Asp - Ser - Thr - Ala - Thr - Phe - Thr - Val - Asn - Tyr - Lys</td>
</tr>
<tr>
<td>(E) (M+H)⁺</td>
<td>m/z 504.2</td>
<td>Glu - Glu - Ala - Arg</td>
</tr>
<tr>
<td>(F) (M+H)⁺</td>
<td>m/z 490.3</td>
<td>Glu - Leu - Ser - Asn - Tyr - Glu - Val - Lys</td>
</tr>
<tr>
<td>(G) (M+H)⁺</td>
<td>m/z 396.3</td>
<td>Asn - Phe - Glu - Ile - Val - Tyr - Thr - Glu - Leu - Asp - Asn - Leu - Val - Gly</td>
</tr>
<tr>
<td>(H) (M+H)⁺</td>
<td>m/z 1071.5</td>
<td>741.4 830.3 977.2 1144.7 1555.6</td>
</tr>
<tr>
<td>(I) (M+2H)⁺</td>
<td>m/z 1165.6</td>
<td>727.5 1038.2 1565.5 1972.0</td>
</tr>
</tbody>
</table>
The results of sequencing of tryptic peptides by MALDI-TOF-MS and LC-ESI MS/MS indicated errors in the X-ray sequence, i.e., N → D at residue 14, E → Q at residue 113, E → Q at residue 123 and D → N at residue 133. Apart from these corrections the results also confirmed the X-ray crystallographically deduced sequence for residues 1-26, 55-74, 100-141. However amino acid residues 27-54 and 75-99 could not be resolved from the tandem mass spectrometry data of tryptic peptides although these two large stretches accounting to a total of 53 residues are well defined in the X-ray crystal structure (Leonidas et al., 2007). It may be concluded that these peptides could not be extracted from the gel or retained on the reverse phase column throughout the LC-ESI-MS/MS analysis. Such difficulties have been reported previously and the sequence for the peptide was confirmed solely based on the well defined X-ray data (Lunin et al., 2004).

2.3.2. Sequencing of chymotryptic peptides of SRL by LC-ESI MS/MS

In order to obtain the sequence for these two stretches that were not resolved from the tryptic peptides as well as to provide the necessary overlap to establish the order of the sequenced tryptic peptides, the SRL was subjected to chymotryptic digestion and the resulting peptides were analyzed by LC-ESI-MS/MS. The reconstructed total ion current (TIC) chromatogram of the chymotryptic digest of SRL is shown in figure 4. Each
of the peaks is labeled according to the corresponding chymotryptic peptides.

Figure 4. Total ion current (TIC) chromatogram of Chymotryptic digest of SRL.

ESI-MS/MS spectra of protonated chymotryptic peptides of SRL C1, C2, C3, C4, C5, C6, C7, C8, C9, C10, C12, C13, C14 and C15 obtained from CID are shown in figure 5; a, b, c, d, e, f, g, h, i, j, k, l, m and n respectively. Expected, observed protonated masses, deduced sequences of chymotryptic peptides from their CID and their positions in the final sequence are presented in table. 4.
Figure 5. ESI-MS/MS spectra of chymotryptic peptides of SRL

5 (a). ESI-MS/MS spectrum of the singly charged chymotryptic peptide (C1), KITVRVY (residues 3-9).

5 (b). ESI-MS/MS spectrum of peptide ion (M+H)+ = 996.2 of SRL. The sequence deduced is HPVEKTVW (residues 18-25).
5 (c). ESI-MS/MS spectra of (M+H)^+ of peptide C3 and the sequence is KY (residues 26-27).

5 (d). ESI-MS/MS spectrum of the peptide of m/z 1041.3. The spectrum revealed the sequence TITDDQHVL (residues 34-42).
5 (e). ESI-MS/MS spectrum of chymotryptic peptide (C5) of SRL. The sequence obtained is GGSGTSGTL (residues 45-53).

5 (f). ESI-MS/MS spectrum of peptide ion (M+H)$^+$ = 322.1 of SRL. The sequence deduced is RF (residues 54-55). The immonium ion of phenylalanine is indicated by the one letter code (F).
5 (g). ESI-MS/MS spectrum of the singly charged ion m/z 439.1 obtained from ESI-ion trap mass spectrometer. The spectrum revealed the sequence TATF (residues 64-67).

5 (h). ESI-MS/MS spectrum of peptide ion (M+H)+ of the m/z 589.2 of SRL. The sequence deduced is GVHNY (residues 68-72).
5 (i). ESI-MS/MS spectrum of peptide ion (M+H)$^+$ = 489.2 of SRL. The sequence deduced is KRW (residues 73-75).

5 (j). ESI-MS/MS spectrum of the singly charged ion m/z 764.3 obtained from ESI-ion trap mass spectrometer. The spectrum revealed the sequence VINQQV (residues 90-95).
5 (k). ESI-MS/MS spectrum of peptide ion (M+H)^+ = 383.0 of SRL. The sequence deduced is SNY (residues 110-112). The immonium ion of tyrosine is indicated by the one letter code (Y).

5 (l). ESI-MS/MS spectrum of peptide ion (M+H)^+ = 522.2 of SRL. The spectrum revealed the substitution of Q for E at the 123rd position and the corrected sequence is QIVY (residues 123-126).
5 (m). ESI-MS/MS spectrum of the doubly charged ion m/z 783.6 obtained from ESI-ion trap mass spectrometer. The spectrum revealed the substitution of N for D at the 133rd position and the corrected sequence is TEAEGNLHANLIIG (residues 127-141).

5 (n). ESI-MS/MS spectrum of peptide ion (M+H)+ = 302.1 of SRL. The sequence deduced is IIG (residues 139-141). The immonium ion of isoleucine is indicated by the one letter code (I).

Expected, observed protonated masses, deduced sequences of chymotryptic peptides from their CID and their positions in the final sequence are presented in table 4.
Table 4. Peptide fragments observed from a chymotryptic digest of SRL in LC-ESI-MS/MS.

<table>
<thead>
<tr>
<th>Peptide No.</th>
<th>Amino acid No.</th>
<th>Expected mass (M+H)+b (Da)</th>
<th>Observed mass (M+H)+ (Da)</th>
<th>Charge state</th>
<th>Ret. Time (min)</th>
<th>Error (ppm)</th>
<th>Peptide sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1</td>
<td>3-9</td>
<td>878.55</td>
<td>878.5</td>
<td>+1</td>
<td>35.2</td>
<td>56.91</td>
<td>KITVRYY*</td>
</tr>
<tr>
<td>C2</td>
<td>18-25</td>
<td>995.53</td>
<td>996.2 (439.7)2+</td>
<td>+1,+2</td>
<td>35.9</td>
<td>672.55</td>
<td>HPVEKTVW*K</td>
</tr>
<tr>
<td>C3</td>
<td>26-27</td>
<td>310.18</td>
<td>310.1</td>
<td>+1</td>
<td>17.7</td>
<td>257.98</td>
<td>KY*</td>
</tr>
<tr>
<td>C4</td>
<td>34-42</td>
<td>1041.52</td>
<td>1041.3 (521.3)2+</td>
<td>+1, +2</td>
<td>37.7</td>
<td>211.27</td>
<td>TTTDDQHVL*</td>
</tr>
<tr>
<td>C5</td>
<td>45-53</td>
<td>736.35</td>
<td>736.3</td>
<td>+1</td>
<td>32.9</td>
<td>67.90</td>
<td>GGSGTSGTL*</td>
</tr>
<tr>
<td>C6</td>
<td>54-55</td>
<td>322.19</td>
<td>322.1</td>
<td>+1</td>
<td>27.4</td>
<td>279.41</td>
<td>RF*</td>
</tr>
<tr>
<td>C7</td>
<td>64-67</td>
<td>439.22</td>
<td>439.1</td>
<td>+1</td>
<td>37.0</td>
<td>273.28</td>
<td>TATF*</td>
</tr>
<tr>
<td>C8</td>
<td>68-72</td>
<td>589.27</td>
<td>589.2</td>
<td>+1</td>
<td>29.8</td>
<td>118.805</td>
<td>GVHN*</td>
</tr>
<tr>
<td>C9</td>
<td>73-75</td>
<td>489.29</td>
<td>489.2</td>
<td>+1</td>
<td>28.3</td>
<td>183.97</td>
<td>KRW*</td>
</tr>
<tr>
<td>C10</td>
<td>90-95</td>
<td>764.39</td>
<td>764.3</td>
<td>+1</td>
<td>35.6</td>
<td>117.75</td>
<td>VINQQY*</td>
</tr>
<tr>
<td>C11</td>
<td>97-104</td>
<td>961.47</td>
<td>962.3</td>
<td>+1</td>
<td>47.7</td>
<td>155.87*</td>
<td>SEKNREEA*</td>
</tr>
<tr>
<td>C12</td>
<td>110-112</td>
<td>383.16</td>
<td>383.0</td>
<td>+1</td>
<td>30.0</td>
<td>417.75</td>
<td>SNY*</td>
</tr>
<tr>
<td>C13</td>
<td>123-126</td>
<td>523.28</td>
<td>522.2</td>
<td>+1</td>
<td>39.3</td>
<td>172.3*</td>
<td>QIVY*</td>
</tr>
<tr>
<td>C14</td>
<td>127-141</td>
<td>1566.78</td>
<td>1566.2 (783.6)2+</td>
<td>+1, +2, +3</td>
<td>41.4</td>
<td>261.78*</td>
<td>TEAEGNNLHANLIIG*</td>
</tr>
<tr>
<td>C15</td>
<td>139-141</td>
<td>302.21</td>
<td>302.1</td>
<td>+1</td>
<td>32.0</td>
<td>364.11</td>
<td>IIG*</td>
</tr>
</tbody>
</table>

*a Assignments were made by comparing observed peptide masses with a theoretical digest of the SRL derived from the X-ray crystallographic data.

*b Expected mass calculated from monoisotopic molecular weights.

*n Acetylation at the N terminus.

*c Sequence deduced from CID spectrum and the corrected amino acids are indicated in bold.

*m Sequences confirmed solely on molecular mass.

* Obtained after modification.
The data presented in table 4 show the observed protonated monoisotopic masses of chymotryptic peptides of m/z 962.30 (residues 97-104), 522.20 (residues 123-126) and 1566.2 (residues 127-141) were differing from their theoretical masses by 862 ppm, 2068 ppm and 370 ppm respectively. While accounting for these deviations, corrections were made with regard to Asn/Asp and Gln/Glu. Replacement of Q→E at residue 98, E→Q at residue 123 and D→N at residue 133, gave greatly reduced error by 155.87 ppm for m/z 962.3, 172.3 ppm for m/z 522.20 and 261 ppm for 1566.2.

Replacements at these positions were justified and further confirmed from their CID spectral data, which yielded numerous sequence-specific daughter ions presented in table 5. M and N. The mass accuracy of the chymotryptic peptides are within the admissible limit of +/<0.3 Dalton (Biemann, 1992) except for peptides C2 and C14, for which the mass accuracy value was found to be 0.67 and 0.41 Daltons respectively and cannot be explained. However the sequences for these two peptides C2 (18-25) and C14 (127-141) were established from the complimentary sequence data of tryptic peptides T3 (8-22), T4 (23-26) and T13 (121-141) obtained by LC-ESI-MS/MS analysis.
Table 5. CID mass spectral data for Chymotryptic peptides. The m/z values (found) for particular ion series are listed by rows.

<table>
<thead>
<tr>
<th>(A) (M+H)^+ m/z 878.5</th>
<th>(B) (M+H)^+ m/z 996.2</th>
<th>(C) (M+H)^+ m/z 201.10</th>
</tr>
</thead>
<tbody>
<tr>
<td>a 315.3 416.3 576.5 680.6</td>
<td>s 664.2 763.8</td>
<td>a</td>
</tr>
<tr>
<td>b 129.0 342.1 412.2 588.4 697.1</td>
<td>b 214.9 646.1 591.0 456.4 791.4</td>
<td>b 129.0</td>
</tr>
<tr>
<td>c 248.3 497.4 716.4</td>
<td>e 351.0 400.1 477.0</td>
<td>e 147.0</td>
</tr>
<tr>
<td>Ly - Ly - Thr - Val - Arg - Val - Tyr</td>
<td></td>
<td></td>
</tr>
<tr>
<td>a 483.6 482.4</td>
<td>s 482.8 599.3 421.9</td>
<td>x</td>
</tr>
<tr>
<td>y 786.3 477.4 436.2 457.2 218.0</td>
<td>y 892.1 761.3 662.3 553.4</td>
<td>y 323.2 154.8</td>
</tr>
<tr>
<td>z 735.6 493.3 512.3 244.3</td>
<td>z 742.3 643.3 516.9 308.5</td>
<td>z 187.9</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>(D) (M+H)^+ m/z 1041.3</th>
<th>(E) (M+H)^+ m/z 736.3</th>
<th>(F) (M+H)^+ m/z 322.1</th>
</tr>
</thead>
<tbody>
<tr>
<td>a 156.0 782.5 881.4</td>
<td>s 478.1</td>
<td>a</td>
</tr>
<tr>
<td>b 218.0 361.0 453.2 568.1 672.0 811.3 796.4 1023.2</td>
<td>b 306.8 477.2 504.3 465.2</td>
<td>b 177</td>
</tr>
<tr>
<td>c 492.3 1522.4</td>
<td>c 373.3 464.0 521.1 472.1</td>
<td>c</td>
</tr>
<tr>
<td>Thr - Be - Thr - Asp - Glu - His - Val - Leu</td>
<td>Gly - Gip - Ser - Gip - Thr - Ser - Gip - Thr - Lys - Arg - Phe</td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>a 563.1</td>
<td></td>
</tr>
<tr>
<td>y 472.1 706.5 612.4 496.1 368.2</td>
<td>y 412.2 535.1 478.2 377.3</td>
<td></td>
</tr>
<tr>
<td>z 810.6 708.6 479.1</td>
<td>z 682.6 518.2 411.1 368.8</td>
<td>x</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>(G) (M+H)^+ m/z 459.1</th>
<th>(H) (M+H)^+ m/z 597.2</th>
<th>(I) (M+H)^+ m/z 609.2</th>
</tr>
</thead>
<tbody>
<tr>
<td>a 156.0 782.5 881.4</td>
<td>s 478.1</td>
<td>a</td>
</tr>
<tr>
<td>b 218.0 361.0 453.2 568.1 672.0 811.3 796.4 1023.2</td>
<td>b 306.8 477.2 504.3 465.2</td>
<td>b 177</td>
</tr>
<tr>
<td>c 492.3 1522.4</td>
<td>c 373.3 464.0 521.1 472.1</td>
<td>c</td>
</tr>
<tr>
<td>Thr - Ala - Thr - Phe</td>
<td>Gip - Val - His - Asp - iyr</td>
<td>Lys - Arg - Thr</td>
</tr>
<tr>
<td>a</td>
<td>a 305.1</td>
<td></td>
</tr>
<tr>
<td>y 247.0 326.9</td>
<td>y 522.1 423.1 294.0</td>
<td>y 344.1</td>
</tr>
<tr>
<td>z 549.0 480.9</td>
<td>z 515.0 414.0 376.9</td>
<td>z 344.0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>(J) (M+H)^2 m/z 746.3</th>
<th>(K) (M+H)^2 m/z 832.0</th>
<th>(L) (M+H)^2 m/z 302.10</th>
</tr>
</thead>
<tbody>
<tr>
<td>a 555.3</td>
<td>s 170.9 327.0</td>
<td>a 86.3 198.0 256.9</td>
</tr>
<tr>
<td>b 213.1 372.7 485.2 582.1</td>
<td>b 201.9 365.0</td>
<td>b 137.0</td>
</tr>
<tr>
<td>c 343.9 472.0</td>
<td>c 373.3</td>
<td>c 373.3</td>
</tr>
<tr>
<td>Val - His - Asp - Gln - Gln - Tyr</td>
<td>Ser - Asp - Tyr</td>
<td>Ser - Be - Gip</td>
</tr>
<tr>
<td>a</td>
<td>x 102.1</td>
<td>x 287.8</td>
</tr>
<tr>
<td>y 521.1 439.1 309.5</td>
<td>y 295.9</td>
<td>y 109.0</td>
</tr>
<tr>
<td>z 648.2 551.1 451.2 353.9</td>
<td>z 218.3 144.8</td>
<td>z 344.0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>(N) (M+2H)^2 m/z 783.6</th>
</tr>
</thead>
<tbody>
<tr>
<td>a 468.9 747.2 488.4</td>
</tr>
<tr>
<td>b 301.8</td>
</tr>
<tr>
<td>c 505.1</td>
</tr>
</tbody>
</table>

| Thr - Gip - Ala - Gip - Gip - Asp - Asp - Leu - His - Ala - Asp - Leu - Be - Ser - Gip | |
| a | 1104.3 876.3 434.0 |
| y 1264.3 1158.4 1078.9 964.0 853.4 717.2 436.0 319.4 415.2 202.2 188.9 | |
| z | 1247.3 1177.3 1081.3 | 710.3 562.3 | 398.2 286.9 |
2.3.3. Differentiation of lysine and glutamine

Since the isobaric amino acids lysine and glutamine which differ in mass by 0.04u, are difficult to distinguish by low energy CID analysis, trypsin cleavage specificity was considered for assigning Lys-3, Gln-10, lys-22, Lys-26, Gln-39, Lys-73, Gln-93, Gln-94, Lys-99, Gln-108, Lys-115, Lys-118 (Fig. 6). However, Gln-113 and Gln-123 were assigned based on the accuracy of monoisotopic masses of both precursor ion as well as sequence specific daughter ions observed in the CID spectra of tryptic peptide of m/z 979.30 and, 2330.20 also chymotryptic peptide having m/z 522.2.

2.3.4. Differentiation of leucine and isoleucine

Leucine and isoleucine having identical mass cannot be differentiated by low energy CID, hence the advantage of chymotrypsin cleavage specificity was taken to differentiate these two amino acids. Based on the chymotrypsin specificity and CID spectrum of chymotryptic peptides, it was possible to assign unambiguously Ile-4, Ile-35, Leu-42, Leu-53, Ile-91, Leu-109, Ile-124, Leu-138, Ile-139 and Ile-140. Whereas Ile-78, Leu-82 and Leu-134 could not be confirmed, as the chymotryptic peptides containing these residues were not observed in LC-ESI-MS/MS. However Leu-134 was assigned by comparing the homology with related fungal lectins (Leonidas et al., 2007). The CID mass spectral data of chymotryptic peptides C1 (3-9), C2 (18-25) and C3 (26-27) and the tryptic peptides T1 (1-
3), T2 (4-7) T3 (8-22) and T4 (23-26), provided overlapping amino acid sequence information and to deduce the sequence from 1-27; Ac-TYKITVRVYQTNPDAFFHPVEKTVWKY. Similarly, the collision-induced dissociation mass spectra of C6 (54-55), C7 (64-67), C8 (68-72) and C9 (73-75) gave complementary sequence information obtained from the MS/MS spectral data of T5 (55-73), which provided sequence from 54-75; RFHADNGESFTATFGVHNYKRW. The sequence for positions 97-105 (SEKNREEAR) was deduced from the complementary CID spectral data obtained from T8 (102-105) and C11 (97-104). The MS/MS spectral data of C12 (110-112) provided overlapping amino acid sequence information of T10 (108-115) where as no complementary chymotryptic peptide was found for T11 (116-118) and T12 (119-120), however based on CID spectral data of C12, T10 and T11, we could assign the amino acid sequence QLSNYQVKNK (108-118) unambiguously. The C-terminal amino acid sequence NFQIVYTEAEGNNLHANLIIG was confirmed based on the MS/MS spectrum of C13 (123-126), C14 (127-141) and C15 (139-141) as they gave overlapping amino acid sequence information of T13 (121-141).

The combined Collision induced dissociation mass spectral data of tryptic and chymotryptic peptides was key to verify and resolve the ambiguities between Asn / Asp and Gln / Glu and enabled us to assign the amino acid sequences corresponding to the positions 1-27 (N-terminus), 54-
75, 97-106, 102-105, 108-118 and 121-141 (C-terminus) of the protein. The sequence TITDDQHVL (34-42) and GGSGTSGTL (45-53) deduced solely from the MS/MS spectral data of C4 and C5 as no corresponding complementary tryptic peptides were observed from tandem mass spectrometry.

Peptides corresponding to 28-33 (ANGGTW), 43-45 (TM) and 76-89, (CDIVTNLAADETGM) (Figure 6), were not detected neither in the tryptic nor in the chymotryptic digests, by LC-ESI-MS/MS. However these sequences were confirmed from the high electron density maps obtained at 1.1 Å resolution (Leonidas et al., 2007). Residues Asn-29, Asp-77, Asn-81, Asp-85 and Glu-86 were indeed assigned based on their differences in the B-factors and potential formation of hydrogen bonds (Leonidas et al., 2007). From these results of MS/MS spectral analysis we could assign the residues D 14, E 98, Q 113, Q 123, and N 137 unambiguously. Further the analysis data involving 13 tryptic and 15 chymotryptic peptides covered 118 residues of the total 141 residues accounted for 83.68 % of the SRL sequence that was deduced by X ray crystallography at near atomic resolution (1.1 Å).

The primary structure of SRL deduced from the data discussed above is shown in figure 6. Solid lines below and arrow lines above the amino acid sequence show Tryptic and chymotryptic peptides respectively.
Figure 6. Amino acid sequence of SRL determined by tandem mass spectrometry.
Peptides are identified by letters that indicate the enzyme used for preparation: T and C indicates tryptic peptides (→) and chymotryptic peptides (↔) respectively. Asterisks indicate ambiguous amino acids resolved by X-ray crystallography. * Corrected amino acids; X: Confirmed based on fragmentation ions. Arrows (►) indicate amino acid residues confirmed by X-ray crystallography.
In order to determine the sequence identity, similarities and differences the amino acid sequence of SRL deduced by mass spectrometry was used for multiple sequence alignments with other related fungal lectins; *Agaricus bisporus* lectin (ABL) (Crenshaw et al., 1995), *Xerocomus chrysenteron* lectin (XCL) (Trigueros et al., 2003) and *Arthrobotrys oligospora* lectin (AOL) (Rosen et al., 1996, 1997) using the program ClustalW as it provides information of conserved sequences (Thompson et al., 1994). Multiple sequence alignment of SRL compared with the sequences of ABL, XCL and AOL is presented in figure 7.

Figure 7. Multiple sequence alignment of SRL, ABL, XCL and AOL as aligned by ClustalW program.

The sequences of fungal lectins from *Agaricus bisporus*, *Xerocomus chrysenteron* and *Arthrobotrys oligospora* were compared with the SRL sequence. Identical residues in all the four lectins are shown in green color indicating the sequence homologies. Conserved and semi conserved substitutions are marked by pink and blue color respectively.
Amongst the four lectins, SRL, ABL, XCL, AOL compared with, SRL is the smallest having 141 residues also it differs from the rest by having blocked N-terminus. Largest among the group is AOL with 145 residues. Multiple sequence alignments by ClustalW revealed that SRL show highest sequence identity with ABL (65%) followed by XCL (58%) and relatively lower level of identity for AOL (48%). SRL, isolated from Sclerotium rolfsii, a deuteromycete belonging to lower fungi where as the lectins AOL, ABL and XCL isolated from Arthrobotrys oligospora (a ascomycete), Agaricus bisporus (a basidiomycete) and Xerocomus chrysenteron (a basidiomycete) all of them belong to higher fungi. Legume lectins, inspite of their considerable differences in their ligand binding specificity and quaternary structures, they posses high degree homologies and hence are considered as evolutionarily conserved molecules (Sharon, 1993; Brewer, 1996). Considering the sequence homology found in so far known fungal lectins, similar evolutionary significance could be attributed for fungal lectins too.