CHAPTER 4

PRELIMINARY CHARACTERIZATION
4.1 Introduction

In addition to the availability of a homogenous preparation of the protein to be used for crystallization attempts, biochemical data on the protein can aid in the subsequent structural studies. For example, the complete amino acid sequence of the protein is not only useful in the interpretation of the electron density map, but also in the choice of phase determination methods. Even if the partial sequence is known, it is used for the identification and classification of the protein based on its homology with a database of protein sequences. This chapter describes the preliminary characterization and sequence analysis of the immune protein from the hemolymph of *A. mylitta* larvae.

4.2 Purification of Induced Antibacterial Activity

The purified protein used for the sequencing and X-ray crystallographic analyses was kindly provided by Dr. J. Nagaraju of the Seribiotech Research Laboratories, Bangalore, India. The procedures involved in the purification of the protein are briefly described below.

The protein was purified from the hemolymph of *A. mylitta* larvae. *Antheraea mylitta* larvae were reared outdoor on *Terminalia arjuna* plants as per the standard rearing protocols. Fifth day larvae were immunized with a 50 μl suspension of log phase *Escherichia coli* strain JM103 washed in saline solution at a concentration of 1 x 10⁷ cells/ml. Hemolymph was collected from the immunized larvae after 48 hours of immunization. An antibacterial assay was carried out by measuring the zone of bacterial growth inhibition on thin agar plates seeded with log phase *E. coli* and
Micrococcus luteus. Antibacterial activity was also measured by a spectrophotometric assay in which the hemolymph was incubated with log phase *E. coli* and the absorbance at 580 nm was compared with that of the control hemolymph from unimmunized larvae (Hultmark *et al*, 1980).

The hemolymph was diluted in 0.1 M ammonium acetate buffer at pH 7.0 and loaded on a CM-Sephadex C-50 column equilibrated with the same buffer. After extensive washing, the bound proteins were eluted from the column using a step gradient of 0.3M and 0.8M ammonium acetate, pH 7.0. Each fraction was monitored by its absorbance at 280 nm. Each fraction was also assayed for antibacterial activity by the spectrophotometric assay previously mentioned. Fractions having antibacterial activity were pooled, concentrated and loaded on a Sephadex G-100 column equilibrated in 0.1 M phosphate buffer, pH 7.0. The fractions were again monitored both by absorbance at 280 nm and the antibacterial assay. Antibacterial activity in the eluate was restricted to a single large peak. This peak appeared as a single band corresponding to a molecular weight of 14 kDa in SDS-PAGE as seen in Fig. 4.1a. Protein profiles of the induced hemolymph and the molecular weight markers are also shown for comparison.

The activity of the protein fraction purified in the previous step was monitored by a gel overlay assay (Gabriel, 1971). Electrophoresis of the immunized hemolymph, purified protein fraction and the control hemolymph was carried out on 15% polyacrylamide gels. Bands with antibacterial activity were localized by overlaying the gel with log phase *E. coli* in rich bacterial medium containing 0.6% agarose. Figure 4.1b shows the results of the gel overlay assay. Two antibacterial
**Figure 4.1:** (a) SDS-PAGE of the *E. coli*-induced antibacterial protein. Lanes 1, 2 and 3 show standard molecular weight markers, the purified protein and the immunized hemolymph respectively. The 14 kDa band corresponding to the protein is indicated by an arrow. (b) 15% PAGE of immunized hemolymph (lane 1), purified protein (lane 2) and unimmunized hemolymph (lane 3) overlaid with agar gel containing log phase *E. coli*. The lower band corresponds to the major induced antibacterial protein.
bands could be found in the immunized hemolymph. The fast moving band had higher activity as compared to the less mobile band, probably corresponding to cecropin as identified in the case of *A. pernyi* (Qu et al, 1982). The purified hemolymph only contained the fast moving, higher activity band seen in the immunized hemolymph. On the contrary, no antibacterial activity could be localized in the unimmunized hemolymph.

The purity of the protein was verified by 12% SDS-PAGE followed by silver staining before N-terminal sequencing or crystallization experiments were started. The protein was seen as a single band after staining, thus confirming its homogeneity for use in further experiments.

### 4.3 N-terminal Sequencing

The sequencing data is used for the classification of the protein on the basis of its homology with other proteins. It also provides clues about the nature of activity or function of the unknown protein and helps to establish its evolutionary relationship with other proteins in the same superfamily. In case the total amino acid sequence is known, this information is used for the prediction of secondary structures, or in the homology modeling of the protein as an alternative to experimental structural studies. In crystallography, sequence determination helps in the identification of appropriate search models for structure solution using the molecular replacement method.

Protein sequences can either be determined by direct experimental methods or indirect determination from the experimentally derived DNA sequence of the gene. For proteins which are difficult to clone or are produced in minute quantities,
chemical methods of protein sequence determination are the only viable techniques available. The methods which are available for chemical sequencing of proteins include the Dansyl Chloride method, the Dabtc method and the more popular Edman’s degradation method.

4.3.1 Methodology of the Edman’s Degradation Method

The Edman’s degradation method is probably the most important technique for the determination of amino acid sequences in proteins and peptides. This revolutionary method of protein sequence was introduced and developed by Edman (Edman and Begg, 1967). The chemistry this sequencing method involves a cyclic procedure, by which amino acid residues are cleaved one at a time from the N-terminus of the peptide, and are then identified as the phenylthiohydantoin (Pth) derivatives by chromatography.

The major steps involved in the Edman’s degradation method are 1) coupling 2) cleavage and 3) conversion. Figure 4.2 depicts a schematic representation of the chemistry of this method.

The coupling reaction is the first step in the Edman’s degradation procedure. Phenylisothiocyanate (Pitc) reacts with the protein to be sequenced, coupling with the free amino group of the first residue on the N-terminal of the protein. The coupling of Pitc with the terminal amino acid residue leads to the formation of a phenylthiocarbamyl-peptide (Ptct). The cleavage of the N-terminal amino acid residue from the Ptct-peptide by cyclization involves the reduction of the thio-group in Ptct, and the cleavage of the peptide bond between the first and second amino acid. The
Figure 4.2: Schematic representation of the chemistry of the Edman’s degradation method for protein sequencing.
products in this reaction are the cleaved anilinothiazolinone derivative of the first amino acid and the peptide with a new amino terminal on the second amino acid. The conversion of the anilinothiazolinone derivative of the cleaved amino acid to its stable phenylthiohydantoin (Pth) derivative constitutes the last step in one cycle of this method. The Pth-amino acid derivative can be identified by chromatography.

These three steps are repeated to identify amino acids from the N-terminus of the protein. In practice, due to the increase in noise from the residuals of the previous steps, determination of long stretches of amino acids is not feasible. Large proteins are sequenced by proteolytic fragmentation followed by sequencing for each fragment.

4.3.2 Partial Sequence Determination of AmP

The partial sequencing of the purified protein was carried out using the general procedure described in the previous section on an Applied Biosystems 470A automated protein sequencer attached to an on-line phenylthiohydantoin (Pth) ABI 120A analyzer. The free α-amino group of the protein reacts with the phenyl isothiocyanate in alkaline medium and the subsequent acid hydrolysis removes the N-terminal residue as the phenylthiohydantoin derivative which is identified on a 120A analyzer by comparison with a profile of standard mixture of Pth amino acids.

The pure protein was first run on a 12% SDS-Polyacrylamide gel and then immobilized by transfer onto a PVDF ProBlot® membrane. The membrane was stained by Coomassie Brilliant Blue R250, and the bands of the immobilized protein were cut, air dried and loaded onto the automated sequencer. Forty cycles of
sequencing were carried out and the results were obtained and analyzed by comparison against standard profiles of *Pth* amino acid derivatives.

The N-terminal 32 amino acids of the purified protein were identified after the sequencing output was analyzed. The percentage recovery of each amino acid in the sequencing run is shown in Table 4.1. Residues corresponding to the cycles 6, 23, 25 and 27 could not be identified, as also the residues after the 32nd cycle due to increased noise levels from the residual background. The yields corresponding to cysteine and tryptophan are known to be very low in such sequencing experiments. It could thus, be expected that the residues corresponding to these cycles were either cysteine or tryptophan.

In order to determine the entire amino acid sequence of *AmP*, controlled proteolytic digestion protocols were designed. Digestion experiments were setup with Trypsin and Chymotrypsin proteases in the enzyme:substrate ratio of 1:100 and 1:50. While Trypsin cleaves the peptide bond at the carboxyl group of lysine and arginine, Chymotrypsin cleaves the peptide bond following an aromatic amino acid. Aliquots were collected after every hour and run on a 20% SDS-PAGE to monitor and optimize the best time for large scale digestion of the protein. Although low molecular weight fragments could be seen as faint bands during digestion with Chymotrypsin, purification of these fragments on C₄ and C₈ HPLC columns using the reverse phase strategy were not successful. The digestion of *AmP* with Trypsin resulted in the fragmentation of the protein into very small peptides which could not be separated by reverse HPLC.
Table 4.1: Amino acid yield for each cycle of the Edman’s degradation procedure during the N-terminal sequencing of AmP.

<table>
<thead>
<tr>
<th>Cycle No.</th>
<th>Amino Acid</th>
<th>Yield (pmol)</th>
<th>Cycle No.</th>
<th>Amino Acid</th>
<th>Yield (pmol)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Lysine</td>
<td>866</td>
<td>17</td>
<td>Phenylalanine</td>
<td>146</td>
</tr>
<tr>
<td>2</td>
<td>Arginine</td>
<td>34</td>
<td>18</td>
<td>Aspartic Acid</td>
<td>114</td>
</tr>
<tr>
<td>3</td>
<td>Phenylalanine</td>
<td>375</td>
<td>19</td>
<td>Glutamic Acid</td>
<td>100</td>
</tr>
<tr>
<td>4</td>
<td>Threonine</td>
<td>216</td>
<td>20</td>
<td>Tyrosine</td>
<td>59</td>
</tr>
<tr>
<td>5</td>
<td>Arginine</td>
<td>44</td>
<td>21</td>
<td>Leucine</td>
<td>118</td>
</tr>
<tr>
<td>6</td>
<td>*</td>
<td>-</td>
<td>22</td>
<td>Methionine</td>
<td>103</td>
</tr>
<tr>
<td>7</td>
<td>Glycine</td>
<td>233</td>
<td>23</td>
<td>*</td>
<td>-</td>
</tr>
<tr>
<td>8</td>
<td>Leucine</td>
<td>276</td>
<td>24</td>
<td>Aspartic Acid</td>
<td>64</td>
</tr>
<tr>
<td>9</td>
<td>Valine</td>
<td>240</td>
<td>25</td>
<td>*</td>
<td>-</td>
</tr>
<tr>
<td>10</td>
<td>Glutamine</td>
<td>188</td>
<td>26</td>
<td>Valine</td>
<td>45</td>
</tr>
<tr>
<td>11</td>
<td>Glutamic Acid</td>
<td>156</td>
<td>27</td>
<td>*</td>
<td>-</td>
</tr>
<tr>
<td>12</td>
<td>Leucine</td>
<td>216</td>
<td>28</td>
<td>Leucine</td>
<td>58</td>
</tr>
<tr>
<td>13</td>
<td>Arginine</td>
<td>36</td>
<td>29</td>
<td>Valine</td>
<td>80</td>
</tr>
<tr>
<td>14</td>
<td>Arginine</td>
<td>48</td>
<td>30</td>
<td>Glutamic Acid</td>
<td>24</td>
</tr>
<tr>
<td>15</td>
<td>Glutamine</td>
<td>141</td>
<td>31</td>
<td>Asparagine</td>
<td>38</td>
</tr>
<tr>
<td>16</td>
<td>Glycine</td>
<td>140</td>
<td>32</td>
<td>Glutamic Acid</td>
<td>33</td>
</tr>
</tbody>
</table>
While the determination of the amino acid sequence of the *Antheraea mylitta* immune protein was in progress, the complete cDNA derived amino acid sequence of the major protein induced by *Enterobacter cloacae* infection in *Bombyx mori* was reported (Lee and Brey, 1995). Interestingly, an immune protein, induced by *E.coli* infection in *Bombyx mori* had an identical N-terminal amino acid sequence (Abraham et al., 1995). Since both *Antheraea mylitta* and *Bombyx mori* are closely related, availability of the complete sequence of the corresponding protein from *B. mori* was of great utility in subsequent studies.

### 4.3.3 Sequence Analysis of AmP

The 32 amino acid N-terminal sequence of the *Antheraea mylitta* protein was compared against the Protein International Resource (PIR) protein sequence database using the program FASTA. The best homology scores for the AmP were obtained with the N-terminal regions of other insect lysozymes. It was observed that the amino acids at position 6, 27 and 25, which could not be resolved during the sequencing run, were conserved in all the other insect lysozymes as Cysteine at 6 and 27, and Tryptophan at position 25 respectively. However, the amino acid at position 23 was variable in all the other insect lysozymes and could not be assigned for AmP.

*AmP* has high sequence homology with the lysozymes of the Lepidopteran insects *Hyalophora cecropia*, *Bombyx mori*, *Manduca sexta* and *Spodoptera littoralis* in the N-terminal region. This protein has 87%, 78% and 72% sequence identity with the lysozymes from *Hyalophora cecropia*, *Manduca sexta* and *Bombyx mori* respectively in this region. The lysozymes from these insects also show about
37-40 % homology with chicken lysozyme in the first 32 residues. On the other hand, the sequence homology of AmP with the basic lysozyme of the dipteran, *Drosophila melanogaster* is only 21% in the same region. Multiple alignment of the above sequences was achieved using the program CLUSTAL V (Higgins et al, 1992), and is shown in Fig. 4.3.

The complete sequences of *Hyalophora*, *Manduca* and *Bombyx* lysozymes are available in literature (Engstrom et al., 1985; Mulnix and Dunn, 1994; Lee and Brey, 1995). These Lepidopteran proteins have high sequence homology (73-76%) between themselves. The sequence homology between these insect lysozymes in the first 32 residues is comparable with their homology with AmP in the same region. It can therefore, be expected that the sequence homology between these lysozymes and AmP would extend over the entire sequence. The lysozymes from Lepidoptera have relatively low homology with the lysozyme from *D. melanogaster* (25-28%). All the lysozymes characterized in insects have been shown to belong to the chicken-type (c-type) lysozyme family and have about 37-40% sequence identity with chicken lysozyme (Faye and Hultmark, 1995). Comparison of the Lepidopteran lysozyme sequences with chicken lysozyme shows that the number and positions of all the cysteines are conserved between the sequences.

Phylogenetic analysis of sequences can be used to determine the distances between homologous members of a family of proteins. The program TREEALIGN (Hein, 1990) was used to calculate the phylogenetic distance between AmP, Lepidopteran lysozymes, and other lysozymes from higher animals. The algorithm calculates evolutionary distances based on sequence homology between the different
**Figure 4.3**: Multiple sequence alignment of the determined amino acid sequence of AmP with chicken lysozyme and the lysozymes from *B. mori* (BMor), *H. cecropia* (Hcec), *M. sexta* (Msex) and *Drosophila* (Dmel).
Preliminary Characterization of AmP

proteins used for analysis. It provides a visual interpretation of sequence homology information. The first 32 amino acids of each protein were used for the analysis. The results of the phylogenetic analysis of the different lysozymes and AmP is shown in Figure 4.4. The classification of animal lysozymes into c-type and g-type is evident from this analysis. It has been established that animal lysozymes are of the c-type or goose-type (g-type) with little sequence homology between them. AmP could be grouped with the c-type lysozymes along with other insect lysozymes. Among the insect lysozymes, the Lepidopteran lysozymes from Bombyx mori, Hyalophora cecropia and Manduca sexta cluster together, whereas the Drosophila lysozyme is distant from this group.

It is clear from the above sequence analysis that AmP belongs to the c-type lysozyme family and has high sequence homology with the insect lysozymes from the Lepidopteran order. Whereas the Lepidopteran lysozymes are all of immune origin, constitutively expressed lysozymes from Drosophila and higher animals are of non-immune origin. The low sequence homology of AmP and the immune lysozymes with the non-immune lysozymes suggests a high degree of difference between two groups. Phylogenetic analysis of the lysozymes and AmP also highlights the distinction between these proteins of immune and non-immune origin. Given the differences in sequence composition and the immune origin of the Lepidopteran lysozymes and AmP, it is logical to expect these proteins to have characteristics distinct from the other lysozymes. Therefore, instead of referring to them as insect lysozymes, we prefer to call them lysozyme-like proteins to highlight this distinction. On the other hand, the presence of key conserved residues like cysteines and the weak homologies
Figure 4.4: Phylogenetic relationships between AmP, insect lysozymes and lysozymes of other origin.
suggest that these proteins may have a core structure and fold similar to that of c-type lysozymes.