

Complete Amino Acid Sequence of a *Lolium perenne* (Perennial Rye Grass) Pollen Allergen, *Lol p II**

(Received for publication, February 8, 1989)

Aftab A. Ansari‡§, Ponniah Shenbagamurthi¶, and David G. Marsh‡

From ‡the Clinical Immunology Division, Department of Medicine, The Johns Hopkins University School of Medicine, Baltimore, Maryland 21239 and ¶Department of Biological Chemistry, The Johns Hopkins University School of Medicine, Baltimore, Maryland 21205

The complete amino acid sequence of a *Lolium perenne* (rye grass) pollen allergen, *Lol p II* was determined by automated Edman degradation of the protein and selected fragments. Cleavage of the protein by enzymatic and chemical techniques established an unambiguous sequence for the protein. *Lol p II* contains 97 amino acid residues, with a calculated molecular weight of 10,882. The protein lacks cysteine and glutamine and shows no evidence of glycosylation. Theoretical predictions by Fraga's (Fraga, S. (1982) *Can. J. Chem.* 60, 2606-2610) and Hopp and Woods' (Hopp, T. P., and Woods, K. R. (1981) *Proc. Natl. Acad. Sci. U. S. A.* 78, 3824-3828) methods indicate the presence of four hydrophilic regions, which may contribute to sequential or parts of conformational B-cell epitopes. Analysis of amphipathic regions by Berzofsky's method indicates the presence of a highly amphipathic region, which may contain, or contribute to, an Ia/T-cell epitope. This latter segment of *Lol p II* was found to be highly homologous with an antibody-binding segment of the major rye allergen *Lol p I* and may explain why immune responsiveness to both the allergens is associated with HLA-DR3.

Grass pollen allergy is a major human health problem throughout the world. One of the more important species causing such allergy is perennial rye grass (*Lolium perenne*). Several allergens from *L. perenne* and other plant pollens have been isolated and characterized (Marsh *et al.*, 1986). The major allergen of *L. perenne* pollen is a glycoprotein ($M_r = 27,000-32,000$), known as *Lol p I* (Rye I),¹ toward which 85-90% of grass-allergic patients are sensitive (Freidhoff *et al.*, 1986). Two additional, well characterized *L. perenne* allergens, *Lol p II* and III, are proteins ($M_r = 11,000$ in both cases), toward which about 45% of grass allergic patients are sensitive in each case (Freidhoff *et al.*, 1986).² These allergens exist in multiple "isoallergenic" forms which are variants that appear

to be immunologically indistinguishable (Johnson and Marsh, 1965a, 1965b). *Lol p II* is a group of acidic proteins with isoelectric points in the range of 5.0-5.3. The isoallergenic form *Lol p IIA* can be separated by electrophoresis or anion-exchange chromatography from the form *Lol p IIB*. Further studies have suggested that *Lol p IIB* may contain two or more isoelectric forms (pI 5.1-5.2) (Cottam *et al.*, 1986; Ansari *et al.*, 1987).

Lol p I and II provide a set of proteins that show no detectable immunological crossreactivity (Johnson and Marsh, 1965a; Ansari *et al.*, 1987), and yet human immune responsiveness to both the antigens is associated with a common HLA-D³ type, namely DR3 (Friedhoff *et al.*, 1988). Therefore, the use of these proteins has proved to be important in genetic studies of human immune responsiveness (Ansari *et al.*, 1989a, 1989b; Freidhoff *et al.*, 1988). It is now known that initiation of an immune response to a soluble protein antigen requires that a fragment of the antigen be presented to the specific helper T-cell in conjunction with an appropriate class II major histocompatibility complex (Ia) molecule, which in the case of humans is an HLA-DR or -DQ or -DP molecule. Since *Lol p I* and II are noncross-reactive at the B-cell level, yet responsiveness is associated with DR3 in both cases, the two allergens may contain a similar structure that is recognized by the same Ia molecule. In order to investigate this possibility, we have determined the complete primary structure of *Lol p II* and compared it with known partial sequences of *Lol p I*. We found a sequence in *Lol p II* that is homologous with a *Lol p I* sequence and appears to contain an important Ia/T-cell site. Furthermore, this sequence may contain, or contribute to, a B-cell site.

MATERIALS AND METHODS AND RESULTS⁴

DISCUSSION

The complete amino acid sequence of *Lol p II* (variant A) was derived from tryptic, chymotryptic, V8 protease, endoproteinase Asp-N, and CNBr peptides of the protein (Fig. 1).

* This work was supported by grants AI-19727 and AI-25372 from the National Institutes of Health and by the Johns Hopkins Institutional Grant M-53-1007. This is publication number 756 from the O'Neill Laboratories, the Good Samaritan Hospital, Baltimore, MD. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

§ To whom correspondence should be addressed: the Clinical Immunology Division, Dept. of Medicine, The Johns Hopkins University School of Medicine, 5601 Loch Raven Blvd., Baltimore, MD 21239. Tel.: 301-532-4222.

¹ *Lol p II* is the designation for the Rye II allergen given under the IUIS Allergen Nomenclature System (Marsh *et al.*, 1986).

² A. A. Ansari, L. R. Freidhoff, and D. G. Marsh, unpublished data.

³ The abbreviations used are: HLA, human leucocyte antigen; Ab, antibody; HPLC, high performance liquid chromatography; rHPLC, reverse-phase HPLC; *Lol p I*, II, and III, *L. perenne* pollen allergens previously known as Rye I, II, and III; TFA, trifluoroacetic acid; TPCK, tosylphenylalanine chloromethyl ketone; TLCK, tosyllysyl chloromethyl ketone; DTT, dithiothreitol; SDS-PAGE, sodium dodecyl sulfate-polyacrylamide gel electrophoresis; PTH, phenylthiohydantoin.

⁴ Portions of this paper (including "Materials and Methods," "Results," Tables S-I and S-II, and Figs. S-1-S-3) are presented in miniprint at the end of this paper. Miniprint is easily read with the aid of a standard magnifying glass. Full size photocopies are included in the microfilm edition of the Journal that is available from Waverly Press.

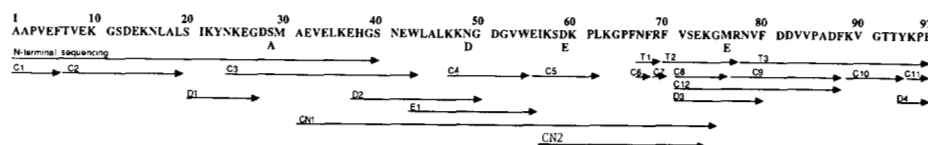


FIG. 1. Amino acid sequence of *Lol p IIA* and schematic outline of the data supporting the sequence. Amino acid residues are given in the single letter code. The preponderant sequence is shown in a contiguous horizontal row; alternate residues at four positions (29, 49, 59, and 76) are shown below this sequence. Peptides obtained by cleavage with trypsin (T), chymotrypsin (C), endoproteinase (Asp-N (D)), *Staphylococcus aureus* V8 protease (E), and CNBr (CN) are indicated, with numbering from the NH₂ terminus (T1, T2, etc.). Residues identified by Edman degradation and the direction of sequencing are indicated by arrows. Data supporting the sequences are given in Miniprint at the end of this paper. The peptides CN1 and CN2 were obtained from two different CNBr digests of the protein; the high performance liquid chromatography profile of only one of the digests is shown in Fig. S-1.

The sequence presented in this paper is based on multiple sequencing of each region. Although data on 24 peptides are presented, a total of 52 peptides were sequenced, with the result that the sequence of each region was confirmed in multiple experiments, sometimes in quadruplicate. *Lol p II* contains 97 amino acid residues, lacks cysteine and glutamine, and shows no evidence of glycosylation. The amino acid

TABLE I
Amino acid composition of *Lol p II* calculated from
amino acid analysis and sequencing

Amino acid	Amino acid analysis	Sequencing ^a
Asp		7
Asn	13.12 (Asp + Asn)	6
Glu		11
Gln	13.18 (Glu + Gln)	0
Ser	5.80	6
Gly	8.26	8
His	0.91	1
Arg	2.43	2
Thr	2.81	3
Ala	7.60	6
Pro	5.14	5
Tyr	1.79	2
Val	7.77	9
Met	2.10	2
Cys	0.08	0
Ile	1.89	2
Leu	6.42	6
Phe	4.70	6
Lys	11.01	13
Trp	ND ^b	2
Total		97
<i>M_r</i>		10,882 ^a
Calculated pI		4.97 ^a

^a These values are for the predominant sequence of the *Lol p IIA* preparation.

^b ND, not determined.

composition derived from the preponderant sequence of *Lol p IIA* is in accord with that determined from the acid hydrolysate of the protein (Table I). The experimental values for some of the amino acid residues do not agree precisely with those predicted from the sequence, because of minor contaminants and lack of accuracy in amino acid analysis. The relative molecular mass of *Lol p II*, calculated from its sequence, is 10,882, which is in agreement with the estimate of 11,000 by physical methods (Johnson and Marsh, 1965a, 1965b; Marsh, 1974, 1975; Cottam *et al.*, 1986). The charged amino acids in *Lol p II* are not clustered in any orderly way in the molecule. Based on the number of acidic and basic residues in *Lol p II*, a net charge of -2.5 at pH 6.5 is calculated, assuming that histidine carries a charge of +0.5 (Offord, 1966). The isoelectric point of *Lol p IIA*, calculated from its amino acid composition using the program in the Bionet, is 4.97, which is in agreement with the value of 5.1-5.2 determined by isoelectric focusing.

A search of the protein sequence database (NBRF, release No. 17.0 and Swiss Protein Database, release No. 7.0) showed that no sequence similar to that reported here for *Lol p II* exists in the database. Only sporadic, short-segment similarities were found.

The *Lol p II* sequence was compared with that of *Lol p I*, for which the following partial sequences are known: NH₂-terminal 30-residue sequence of *Lol p I* (Cottam *et al.*, 1986) and a 28-residue internal sequence (Esch and Klapper, 1989). From studies using monoclonal antibodies, the latter fragment clearly contains a major sequential determinant which is also present in the Group I (*Lol p I*-related) homologues in the pollens of several other grass species (Esch and Klapper, 1989). The NH₂-terminal sequence of *Lol p I* showed no similarity with the *Lol p II* sequence, but the antibody-combining fragment of *Lol p I* showed a high degree of homology with a portion of the *Lol p II* sequence (Fig. 2). A part (residues 75-

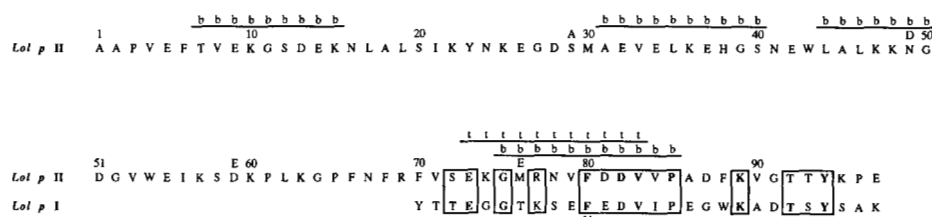


FIG. 2. Alignment of amino acid sequence of *Lol p IIA* with an antibody combining fragment of *Lol p I*. The partial sequence of *Lol p I* is from Esch and Klapper (1989). Alternate residues are shown above the *Lol p II* sequence and below the *Lol p I* sequence. Identical residues are shown in bold letters. These residues and other similar residues in the two sequences are indicated by boxes (the following residues were considered similar: E = D, K = R, S = T, and V = I). The figure also shows postulated B-cell epitopes (indicated by letters b), predicted by the methods of Fraga (1982) and Hopp and Woods (1981), and a postulated Ia/T-cell epitope (indicated by letters t), predicted by the amphipathicity analysis of Margalit *et al.* (1987).

85) of this homologous region in *Lol p II* is a predicted B-cell epitope (see "Materials and Methods" and "Results" for details). Although this segment of *Lol p II* bears sufficient similarity to the *Lol p I* sequence to expect cross-reactivity between the two molecules, we have never detected such a cross-reactivity. This is not surprising considering that single amino acid differences between homologous protein antigens can result in the production of monospecific antibodies.

In addition to containing, or contributing to, a possible B-cell site, the homologous region of *Lol p I* and *II* may also contain an important Ia/T-cell site. Thus, a part (residues 73–83) of the homologous region in *Lol p II* was also found to be highly amphipathic when the *Lol p II* sequence was analyzed by the method of Margalit *et al.* (1987). According to these authors, amphipathic regions are considered to contain Ia/T-cell binding sites; by applying their algorithm, 18 of 23 known Ia/T-cell sites recognized by mice were predicted correctly. Thus, it is likely that the segment 73–83 in *Lol p II* and the homologous segment in *Lol p I* may represent an important Ia/T-cell epitope. Analysis of the 28-residue *Lol p I* peptide sequence did not show any amphipathic region, probably because of the short size of the peptide. Since human immune responsiveness to both *Lol p I* and *II* is associated with HLA-DR3 (Freidhoff *et al.*, 1988), it is possible that this region of the *Lol p I* molecule may also contain a binding site either for the DR3 molecule or for the DQw2.3 molecule that is encoded by DQ genes which are in linkage disequilibrium with genes encoding DR3 in Caucasoid populations.

In conclusion, our data suggest the presence of a closely related Ia/T-cell site on *Lol p I* and *II*. In the case of *Lol p I*, this region has already been proved to contain an important B-cell site, and our analysis suggests that the same may also be true for *Lol p II*. We hypothesize that these homologous sequences contain Ia/T-cell epitopes, which may bind to HLA-DR3 or with DQw2.3 in Caucasoids.

Acknowledgment—We are grateful to Dr. E. Appella for the enzyme endoproteinase Asp-N.

REFERENCES

- Ansari, A. A. (1985) *J. Biosci.* **7**, 315–322
- Ansari, A. A., Carta-Sorcinì, M., Mage, R. G., and Appella, E. (1976) *J. Biol. Chem.* **251**, 6798–6806
- Ansari, A. A., Kihara, T. K., and Marsh, D. G. (1987) *J. Immunol.* **139**, 4034–4041
- Ansari, A. A., Marsh, D. G., and Freidhoff, L. R. (1989a) *Int. Arch. Allergy Appl. Immunol.* **88**, 164–169
- Ansari, A. A., Freidhoff, L. R., Meyers, D. A., Bias, W. B., and Marsh, D. B. (1989b) *Human Immunol.* **25**, 59–71
- Bidlingmeyer, B. A., Cohen, S. A., and Tarvin, T. L. (1984) *J. Chromatogr.* **336**, 93–104
- Blumenthal, K. M., Moon, K., and Smith, E. L. (1975) *J. Biol. Chem.* **250**, 3644–3654
- Cottam, G. P., Moran, D. M., and Standing, R. (1986) *Biochem. J.* **234**, 305–310
- Croft, L. R. (1980) in *Handbook of Protein Sequence Analysis*, 2nd Ed., Vol. 2, pp. 84–85, John Wiley & Sons Ltd., New York
- Drapeau, G. R. (1977) *Methods Enzymol.* **47**, 189–191
- Drapeau, G. R. (1980) *J. Biol. Chem.* **255**, 839–840
- Esch, R. E., and Klapper, D. G. (1989) *Mol. Immunol.*, in press
- Fraga, S. (1982) *Can. J. Chem.* **60**, 2606–2610
- Freidhoff, L. R., Kautzky, E. E., Grant, J. H., Meyers, D. A., and Marsh, D. G. (1986) *J. Allergy Clin. Immunol.* **78**, 1190–1201
- Freidhoff, L. R., Kautzky, E. E., Meyers, D. A., Ansari, A. A., Bias, W. B., and Marsh, D. G. (1988) *Tissue Antigens* **31**, 211–219
- Hopp, T. P., and Woods, K. R. (1981) *Proc. Natl. Acad. Sci. U. S. A.* **78**, 3824–3828
- Hunkapiller, M. W. (1985) *PTH Amino Acid Analysis, User Bulletin*, Applied Biosystems Protein Sequencer, Issue 14, pp. 23–25, Applied Biosystems, Foster City, CA
- Hunkapiller, M. W. (1987) in *Protein/Peptide Sequence Analysis: Current Methodologies* (Bhown, A. S., ed) pp. 87–117, CRC Press, Boca Raton, FL
- Johnson, P., and Marsh, D. G. (1965a) *Eur. Polym. J.* **1**, 63–77
- Johnson, P., and Marsh, D. G. (1965b) *Nature* **206**, 935–937
- Margalit, H., Spouge, J. L., Cornette, J. L., Cease, K. B., DeLisi, C., and Berzofsky, J. A. (1987) *J. Immunol.* **138**, 2213–2229
- Marsh, D. G. (1974) in *Allergology: Proceedings of the 8th Congress of the International Association in Allergology* (Yamamura, Y., Frick, O. L., Horuchi, Y., Miyamoto, S., Naranjo, P., and de Weck, A. L., eds) pp. 381–393, Excerpta Medica, Amsterdam
- Marsh, D. G. (1975) in *The Antigens* (Sela, M., ed) Vol. 3, pp. 271–359, Academic Press, FL
- Marsh, D. G., Goodfriend, L., King, T. P., Löwenstein, H., and Platts-Mills, T. A. E. (1986) *Bull. WHO* **64**, 767–770, reprinted in (1987) *Mol. Immunol.* **24**, 1383–1388
- Offord, R. E. (1966) *Nature* **211**, 591–593

SUPPLEMENTARY MATERIAL TO

Complete Amino Acid Sequence of a *Lolium perenne* (Perennial Rye Grass) Pollen Allergen, *Lol p II*

by

Aftab A. Ansari, P. Shenbagamurthi and David G. Marsh

MATERIALS AND METHODS

***Lol p II*.** The less acidic form of *Lol p II*, *Lol p IIA* (pI 5.1), was purified as described earlier (Ansari *et al.*, 1987). On SDS-PAGE, the *Lol p IIA* preparation gave one band corresponding to its reported molecular weight of 11,000. On agarose electrophoresis and isoelectric focusing, the protein gave a major band with evidence of contamination from its isoallergenic form (Johnson and Marsh, 1965a, 1965b). No contamination with *Lol p I* or *II* was detectable by Coomassie Blue staining following agarose electrophoresis, SDS-PAGE, or isoelectric focusing.

Citraconylation and decitraconylation of the proteins. These procedures were performed essentially as described elsewhere (Ansari *et al.*, 1976).

Trypsin digestion. *Lol p II* (20 nmol) was citraconylated in order to limit the digestion to arginine peptide bonds. The citraconylated protein was digested in 10 mM ammonium bicarbonate, pH 7.8, with TPCK-treated trypsin from bovine pancreas (Sigma T8642, batch # 55F-8065). The enzyme-to-substrate weight ratio was 1:100, and the digestion was carried out at 37°C for 90 min. After the digestion, the citraconyl groups were removed by suspending the digest in 0.1 M pyridine-acetate buffer, pH 3.5, for 24 hr at 25°C. The mixture was lyophilized and then reconstituted in water containing 0.05% TFA. Some peptides were lost as precipitates.

Chymotryptic digestion. *Lol p II* (10 nmol) was digested in 10 mM ammonium bicarbonate, pH 7.8, with TLCK-treated alpha-chymotrypsin from bovine pancreas (Sigma C3142, batch # 75F-8025). The enzyme-to-substrate weight ratio was 1:50, and the digestion was carried out at 37°C for 18 hr. The mixture was lyophilized and then reconstituted in water containing 0.05% TFA. No insoluble material was noticed.

Digestion with Glu-specific Staphylococcal V8 protease. *Lol p II* (12 nmol) was digested in 0.1 M sodium phosphate/2 mM EDTA, pH 7.8, with the V8 protease (Sigma) at 37°C for 18 hr using an enzyme-to-substrate ratio of 1:30. The digest was clear, with no visible insoluble material. No lyophilization was performed before HPLC. EDTA was included in the digestion buffer to inactivate any possible contaminating "neutral protease" activity from *Staphylococcus aureus* (Drapeau, 1977). In the phosphate buffer used, the V8 protease is known to cleave both Asp and Glu bonds; in other buffers (ammonium bicarbonate or acetate), the enzyme more selectively attacks Glu bonds (Drapeau, 1977).

Digestion with endoproteinase ASP-N. *Lol p II* (18 nmol) was digested in 0.1 M ammonium bicarbonate, pH 7.8, with the endoproteinase ASP-N (Drapeau, 1980), which was a kind gift from Dr. E. Appella of the NIH, Bethesda, Maryland. The enzyme-to-substrate ratio was 1:50 (w/w), and the reaction was carried out at 37°C for 24 hr.

CNBr digestion. To a solution of *Lol p II* (5.7 nmol) in 100 µl of 50 mM ammonium bicarbonate, pH 7.8, was added 5 µl of 100 mM DTT. The tube was flushed with nitrogen, and incubated in the dark for 30 min. CNBr (1mg) solution in 105 µl of 0.2 M HCl was added, the tube was flushed with nitrogen again, and the reaction allowed to proceed at 25°C for 14 hr. The solution was diluted with 200 µl of 0.1% TFA in water and applied directly on the HPLC column. Four major peaks were obtained (Fig. S-1); sequencing data on one of them, CN1, is presented in this paper. Another CNBr digest of the protein, where no nitrogen flush was used, yielded six peaks (profile not shown); sequencing data on one of them, CN2, is also presented.

Peptide purification. Peptides were purified by reverse-phase HPLC using Waters C18 columns. Experimental details are presented with individual chromatograms. HPLC effluents were monitored at 214 nm; in later part of the study, the effluents were also monitored at 280 nm in order to discriminate between peptides containing different proportions of aromatic amino acids.

Nomenclature of the peptides. Peptides are designated by a serial number prefixed by a letter. The letters indicate the type of digestion: T, trypsin; C, chymotrypsin; E, Glu-specific protease; D, endoprotease ASP-N; CN, CNBr. The numbers in the peptide designations do not correspond to the order of their elution in HPLC, but to their position in the protein sequence, starting from the N-terminus. This system of nomenclature is followed for all those peptides for which data are presented. There are some peptides which were sequenced, but data on them are not presented in detail, as these peptides cover regions of the proteins that were already sequenced using the numbered peptides. The peptides for which sequencing data are not presented, are indicated in the HPLC profiles (Fig. S-1), in which the corresponding peaks carry a designation showing the amino acid residue positions spanned by the peptides.

Amino acid analysis. This was performed using the PICO-TAG amino acid analysis system of Waters-Millipore. The protein was hydrolyzed with 6 N HCl containing 1% phenol by volume at 110°C for 24 h. The hydrolysate was dried and the amino acids derivatized with phenylisothiocyanate to yield the corresponding phenylthiocarbonyl derivatives following standard procedures (Bidlemeier *et al.*, 1984). These derivatives were analyzed on the PICO-TAG amino acid analysis system which had been previously calibrated with a standard mixture of amino acids.

Amino acid sequence analysis. Automated N-terminal sequence analysis of peptides was performed on a Model 470A protein sequencer (Applied Biosystems) using the O3RPTH program. BioBren Plus (3 mg) was applied onto the trifluoroacetic acid-etched glass fiber filter and subjected to three precycles of Edman degradation prior to sample application. Conversion of the thiazolinone derivatives to phenylthiohydantoin amino acids was carried out with 25% TFA. PTH-amino acid derivatives were separated, on-line, by reversed-phase HPLC on a PTH C-18 column (2.1 × 220 mm; 5 µm; Applied Biosystems) with a gradient of acetonitrile and sodium acetate-trimethylamine buffer, containing 5% tetrahydrofuran as outlined by Applied Biosystems on a Model 120A PTH analyzer.

Prediction of B-cell determinants. Analyses of hydrophilicity and recognition factors (based on amino acid-to-amino acid interaction energies) were carried out by the methods of Fraga (1982) and Hopp and Woods (1981) as described earlier (Ansari, 1985). The analysis was then used to predict the regions that may include, or form part of, B-cell determinants according to the following criteria: The region of highest hydrophilicity in a protein molecule is considered antigenic; regions of high hydrophilicity and low recognition are also considered antigenic; regions of negative hydrophilicity or high recognition or regions less than six residues long are not considered antigenic. This approach is only a predictive method, the results of which must be subsequently tested by experiment.

Prediction of Ia/T-cell determinants. Amphipathicity analysis was performed by the Berzofsky's method of Margalit *et al.* (1987). According to this analysis, amphipathic regions are considered to contain Ia/T-cell binding sites.

RESULTS

Reconstruction of the amino acid sequence. Fig. 1 (main section) depicts the complete amino acid sequence of Lol p II, indicating the alignment of the peptides. These peptides were obtained by HPLC of various digests of Lol p II (Fig. S-1). Most of the major peaks, and some minor peaks were subjected to the automated gas-phase sequencing procedure (see Fig. S-2 for yield data on selected peptides). Additional peptides from each digestion were sequenced but data on them are not shown (except in the chromatographic profiles, Fig. S-1), as these peptides turned out to cover regions that were already sequenced using one or more peptides. These additional peptides gave sequences identical to the respective sequences presented. Because of this replication in sequencing, each region of the proteins was ultimately confirmed by more than one, some times up to four, independent sequencings.

Lol p II-derived chymotryptic peptides C1–C11, including the two dipeptides C6 and C7, yielded sequence information on 91% of the protein. Overlaps between C1, C2 and C3 were obtained from the N-terminal sequencing of the intact protein. A large CNBr fragment CN1 provided overlaps spanning C3 through C8. Overlaps between C8 and C9 were provided by T2 and C12. The tryptic peptide T3 provided the final overlaps between C9, C10, and C11. The carboxy-terminal sequence of Lol p II was obtained from the peptide D3 derived by endoprotease ASP-N digestion of the protein. This peptide yielded the following sequence: Y K P E. Upon acid hydrolysis and amino acid analysis (Table S-1), only these four amino acids were obtained, and the proportion of these amino acids with respect to each other was close to unity. These sequencing and amino acid analysis data on peptide D3 confirmed the carboxy-terminal sequence of Lol p II.

Difficult regions. Lol p II presented one region that was relatively difficult to solve, namely alignment of the peptides C4 and C5 (Fig. 1). Peptide E1 had given one amino acid overlap between C4 and C5. We needed a peptide that covered the region between residues 51 and 59. Endoprotease ASP-N was tried because of the strategic location of two Asp residues at positions 51 and 59. This digestion resulted in a peptide that started from Asp51, but automated sequencing failed to show any residues beyond Trp54. Tryptic digestion did not give a peptide appropriate for the overlap. Cleavage at Met53 and Met56 was tried using CNBr; in one attempt (with no DTT and no nitrogen flush), a peptide was obtained that started with Glu55, obviously as a result of cleavage at Trp54–Glu55. In another attempt, the CNBr digestion was carried out in the presence of DTT in the dark, with a nitrogen flush. The HPLC effluent was monitored at 280 nm in addition to the usual 214 nm, and the peak showing the highest ratio of 280/214 readings was selected for sequencing. This turned out to be the desired peptide CN1, which did not have breaks at the two Trp residues at 43 and 54.

Variants of Lol p II. Lol p II was found to show alternate amino acids at four positions: 29, 49, 59, and 76. In Fig. 2, the major sequence is shown in one contiguous line, with the alternate amino acids shown above it. N-terminal sequencing of the undigested protein showed two amino acids at cycle 29, Ser and Ala in the ratio 532/376 (Fig. S-2), however this may not be a true reflection on the relative proportion of the two forms of the protein, as the sequencing yield of Ser is usually low (Croft, 1980; Hunkapiller, 1985, 1987). The presence of two alternate amino acids at position 29 was confirmed by two chymotryptic peptides C3 and C3a; the former yielded only Ser, while the other peptide showed only Ala at this position. At position 49, most peptides including C4, CN1, D2, and E1 showed only Asn; but one peptide, E1a, showed Asp in addition to Asn in the cycle corresponding to position 49 (Fig. S-2). Position 59 showed Asp in the major peptide CN1 and in D3; but peptide C4 yielded Glu at this position. Position 76 was found to be Met in the following peptides: T2, C8, C9, C12 and four additional peptides for which data are not shown; however, peptide D3 showed Glu in addition to Met at this position. Thus, a total of 8 peptides showed only Met at position 76, while only one peptide, D3, showed Glu (in addition to Met) at this position. Obviously, in our protein preparation, the Met57 containing protein is present in predominance compared with Glu57 containing protein variant. It is not clear whether one or more alternative forms of Lol p II contain all these four substitutions. In addition to a major band, the Lol p II preparation contained two minor protein bands, which may explain the substitutions.

Specificity of enzymic and chemical cleavages. We consider it important to make a note of the unexpected peptide-bond cleavages observed during the course of this investigation. Table S-II lists those cleavages that are usually not expected from the specificities of the respective reagents. This table shows the bond cleaved, position in the protein, and yield of the peptide relative to the yield (based on sequencing data) of the most abundant peptide in the group; this latter information gives an idea of whether the "unexpected" cleavage occurred at a low or high yield relative to an "expected" cleavage.

Prediction of B-cell sites and Ia/T-cell sites. Fig. S-3 shows the hydrophilicity and recognition profiles for Lol p II. Based on these profiles, four segments that may contain, or contribute to, antibody combining sites or B-cell determinants were predicted by Fraga's method (Fraga, 1982), as shown in Fig. 2 (main section of the paper).

Amphipathicity analysis was performed by Berzofsky's method using blocks of both 7 residues and 11 residues. Only those segments that showed amphipathicity in both the analyses and were at least 5 residues long were considered to contain potential Ia/T-cell sites. For Lol p II, the analysis using 7-residue blocks showed two amphipathic regions; 60–64 (amphipathicity score, AS = 11.6) and 74–83 (AS = 25.4); however, using 11-residue blocks, the first of these two regions could not be confirmed as amphipathic (a low AS of 4.3 was obtained for residues 62–63). The second region was confirmed to be amphipathic in the analysis using 11-residue blocks (residues 73–81, AS = 22.5), and is shown in Fig. 2.

Table S-I. Confirmation of the carboxy-terminal sequence of Lol p II using the peptide D3

Amino acid	AA analysis pMole	Number	Number obsd. by sequencing
Tyr	272	0.95	1
Lys	258	0.90	1
Pro	288	1.01	1
Glu	326	1.14	1 (Glu)

Table S-II. Unexpected peptide-bond cleavages observed

Enzyme/ Chemical	Unexpected bond cleaved	Position in the protein	Yield (%)
Chymotrypsin	K – G ¹	74 – 75	1
Trypsin	F – N ²	66 – 67	>90
V8 protease	T – Y	93 – 94	31
Endoprotease ASP-N (from <i>Pseudomonas fragi</i>)	L – S	19 – 20	32
	A – E	31 – 32	3
	K – E	36 – 37	5
	F – N	66 – 67	27
	F – V	70 – 71	21
	S – E	72 – 73	1
	E – K	73 – 74	2
	T – T	92 – 93	9
	T – Y	93 – 94	>90
CNBr	W – E ³	54 – 55	52
	G – T	91 – 92	34

¹ These cleavages could be the result of trypsin contamination in the "TLCK treated" chymotrypsin preparation used (Sigma C3142, batch # 75F-8025).

² This cleavage could be caused by contaminating chymotrypsin activity in the "TPCK treated" trypsin preparation used (Sigma T8642, batch # 55F-8065).

³ Cleavage of tryptophanyl bonds has been reported earlier (Blumenthal *et al.*, 1975).

Details of digestion conditions are described in the Materials and Methods section.

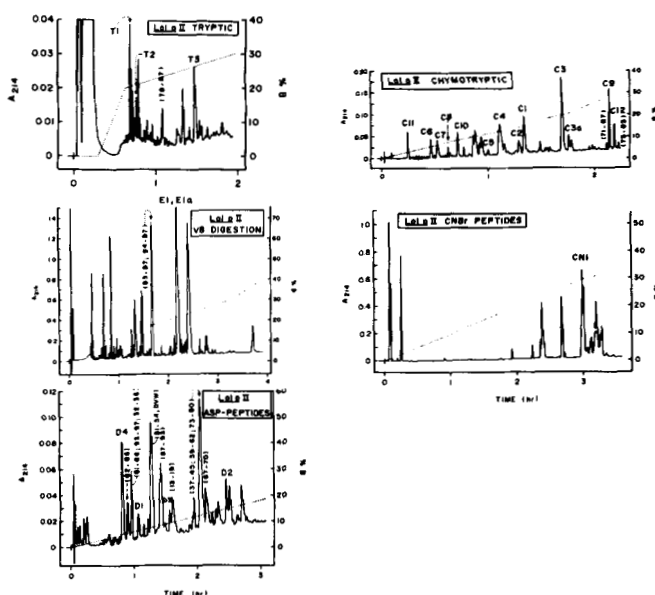


Fig. S-1. HPLC profiles of various digests of Lol p II. Digestion conditions are described in Materials and Methods. Conditions for HPLC were as follows. Tryptic digest: column, Waters C18 Novapak radially compressed/5 mm × 100 mm/5 µm; solvent A, 0.05% TFA in water; solvent B, 0.05% TFA in acetonitrile; flow rate, 1 ml per min. Chymotryptic digest: same as tryptic. Asp digest: column, Waters C18 Novapak radially compressed/5 mm × 100 mm/5 µm; solvent A, 0.109% TFA in water; solvent B, 0.100% TFA in acetonitrile; flow rate, 1 ml per min. V8 digest: column, Waters C18 Deltapak/3.9 mm × 150 mm/5 µm; solvent A, 0.1% TFA in water; solvent B, 0.09% TFA in acetonitrile; flow rate, 1 ml per min. CNBr digest: column, Waters C18 Deltapak/3.9 mm × 150 mm/5 µm; solvent A, 0.1% TFA in water; solvent B, 0.09% TFA in acetonitrile; flow rate, 0.8 ml per min.

Fig. S-2. Amino acid sequencing yield data on *Lol p II*. Residue positions are shown in the left column, followed by the amino acids in *one-letter code* in the second column. In all the subsequent columns, amino acid yields are shown in pmol. The top row for these columns shows the peptide designations, and the second row shows amount, in nmol, of the starting material for each peptide, without making any corrections for losses in digestion, insolubility, and HPLC. For those residue positions (29, 49, 59, and 76), which are occupied by alternative amino acids, the yield data are shown along with the amino acid found. The peptide CN2 was obtained from a second CNBr digest of the protein (HPLC profile not shown in this paper).

Res	A	N-term/18	Y1	Y2	Y3	C1	C2	C3	C3a	C4	C5	C6	C7	C8	C9	C10	C11	C12	CN1	CN2	D1	D2	D3	D4	E1	E2
1	A	11855																								
2	A	11855																								
3	P	7956																								
4	V	4862																								
5	E	3331																								
6	F	7090																								
7	T	7531																								
8	V	6784																								
9	Z	4884																								
10	K	7597																								
11	G	4358																								
12	S	2899																								
13	D	2551																								
14	E	2823																								
15	K	4957																								
16	N	2814																								
17	L	2928																								
18	A	3239																								
19	L	3175																								
20	S	1114																								
21	I	1231																								
22	K	2831																								
23	Y	1538																								
24	N	1417																								
25	K	1682																								
26	E	185																								
27	G	1125																								
28	D	718																								
29	S/A	5532/A374																								
30	M	617																								
31	A	1838																								
32	E	944																								
33	V	715																								
34	E	557																								
35	L	821																								
36	K	1857																								
37	E	357																								
38	I	112																								
39	G	452																								
40	S	197																								
41	N																									
42	E																									
43	W																									
44	L																									
45	A																									
46	L																									
47	K																									
48	K																									
49	N/D																									
50	G																									
51	D																									
52	G																									
53	V																									
54	W																									
55	E																									
56	K																									
57	K																									
58	S																									
59	D/E																									
60	K																									
61	P																									
62	L																									
63	K																									
64	G																									
65	P																									
66	P																									
67	N																									
68	P																									
69	N																									
70	P																									
71	V																									
72	S																									
73	E																									
74	K																									
75	G																									
76	ME																									
77	K																									
78	N																									
79	V																									
80	P																									
81	D																									
82	D																									
83	V																									
84	V																									
85	P																									
86	A																									
87	D																									
88	P																									
89	K																									
90	V																									
91	G																									
92	T																									
93	T																									
94	Y																									
95	K																									
96	P																									
97	E																									

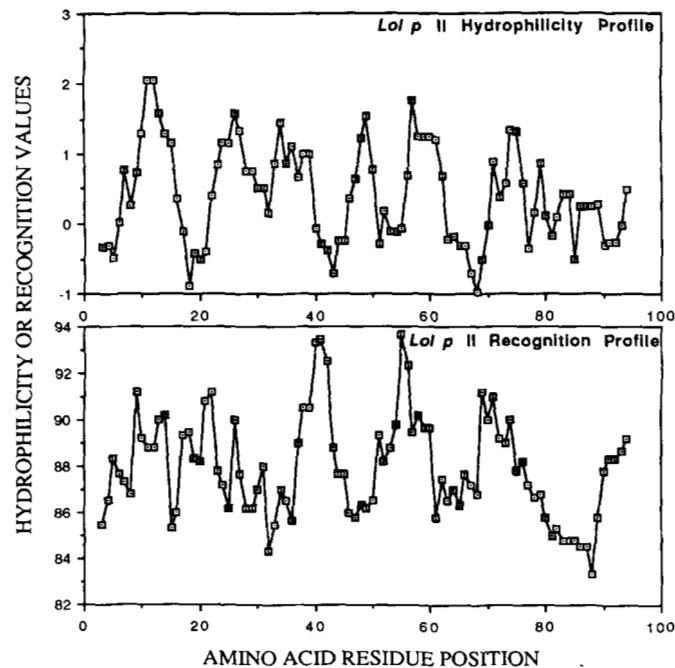


Fig. S-3. Hydrophilicity and recognition profiles for *Lol p II*. These profiles were obtained by the method of Fraga using blocks of 6 residues (Fraga, 1982).