389.
Hirs, C. H. W. (1967), *Methods Enzymol. 11,* 59.
Lemasson, C., Tandeau deMarsac, N., and Cohen-Bazire, G. (1973), *Proc. Natl. Acad. Sci. U.S.A. 70,* 3130.
Lowry, D. H., Rosebrough, N. J., Farr, A. L., and Randall, R. J. (1951), *J. Biol. Chem. 193,* 265.
MacColl, R., Habig, W., and Berns, D. S. (1973), *J. Biol. Chem. 248,* 7080.
OhEocha, C. (1965), *Annu. Rev. Plant Physiol. 16,* 415.
Schram, B. L., and Kroes, H. H. (1971), *Biochemistry 10,* 581.
Spackman, D. H. (1967), *Methods Enzymol. 11,* 3.
Troxler, R. F., Brown, A., Foster, J. A., and Franzblau, C. (1974), *Fed. Proc., Fed. Am. Soc. Exp. Biol. 33,* 1446.
VanderVelde, H. H. (1973), *Biochim. Biophys. Acta 303,* 246.
Vaughn, M. H. (1964), Ph.D. Thesis, Massachusetts Institute of Technology, Cambridge, Mass.
Weber, K., and Osborn, M. (1969), *J. Biol. Chem. 244,* 4406.
Williams, V. P., Friedenreich, P., and Glazer, A. N. (1974), *Biochem. Biophys. Res. Commun. 59,* 462.

The Amino Acid Sequence of Ragweed Pollen Allergen Ra5†

Lawrence E. Mole,‡ Lawrence Goodfriend,# Charles B. Lapkoff, J. Michael Kehoe,§ and J. Donald Capra*

ABSTRACT: The complete amino acid sequence of Ra5, a ragweed pollen allergen, has been determined. Allergen Ra5 is a low molecular weight protein of 45 residues derived from *Ambrosia elatior,* the short ragweed. It contains no detectable carbohydrate or lipid and has four disulfide bridges. The total structure was determined on 1.4 μmol of material and indicates that structural analysis is increasingly possible on relatively small amounts of highly purified material when a combination of automated and manual sequencing techniques and highly sensitive detection systems is employed. This represents the first complete amino acid sequence of a ragweed allergen and it should provide a basis for many structure–function correlative experiments in the field of immediate hypersensitivity.

Almost all wind-pollinated weeds have been implicated in human pollen allergy (Sherman, 1968; Wodehouse, 1971). The most extensively studied of these is the pollen of *Ambrosia elatior,* commonly known as short ragweed. Early work by King *et al.* (1964) suggested that antigen E was the major allergen of ragweed but, recently, several other important allergenically active ragweed pollen allergens have been isolated and characterized (Underdown and Goodfriend, 1969; Griffiths and Brunet, 1971; Lichtenstein *et al.,* 1973; Lapkoff and Goodfriend, 1971, 1974). Allergen Ra5 was of particular interest because of its low molecular weight (5000) and the absence of detectable carbohydrate and lipid. Furthermore Marsh *et al.* (1973a) have shown that IgE-mediated sensitivity to Ra5 was significantly associated with possession of the histocompatibility antigens of the HL-A7 cross-reacting group (Creg). This same association with HL-A7 Creg was reported for formation of human IgG antibodies to Ra5 by Marsh *et al.* (1973b). This represented the first evidence of a strong association between a specific immune response and possession of a specific group of closely related HL-A antigens.

All allergens studied to date have relatively low molecular weights but have otherwise shown no distinguishing chemical features which might correlate with their sensitizing property (Marsh, 1974). However, since relatively little detailed structural information is presently available for these molecules, especially with regard to amino acid sequence, the possibility still exists that some common structural feature, such as a shared sequence stretch, might be associable with the capacity to induce a hypersensitivity reaction. Because of its low molecular weight, its importance in clinical pollenosis, and the relationship between responsiveness to Ra5 and the possession of a group of HL-A antigens, studies were undertaken to determine its complete amino acid sequence. A preliminary account of the findings has been presented (Mole, *et al.,* 1974).

Materials and Methods

Isolation of Ra5. Ra5 was isolated by the method of Lapkoff and Goodfriend (1974). The entire amino acid sequence, including the isolation of the tryptic and chymotrptic peptides, was accomplished on 1.4 μmol of starting material (Figure 1).

Enzyme Digests. Tryptic and chymotryptic digests were prepared from fully reduced and ^{14}C-carboxymethylated Ra5 (O'Donnell *et al.,* 1970) by methods previously described by Press *et al.* (1966). Tryptic digestion was per-

* Recipient of National Institutes of Health Career Development Award 6-K4-GM-35,190. Present address: Department of Microbiology, University of Texas Southwestern Medical School, Dallas, Texas 75235.

† From the Department of Microbiology, Mount Sinai School of Medicine of the City University of New York, New York, New York 10029, and the Department of Experimental Medicine, McGill University, Montreal, Canada. *Received August 12, 1974.* This work was supported in part by grants from the National Science Foundation (GB 17046), the U.S. Public Health Service (AI 09810), The Medical Research Council of Canada (MT 2010), and a Grant-in-Aid from the New York Heart Association.

‡ Member of the British Medical Research Council presently on sabbatical leave at Mount Sinai School of Medicine, New York.
§ Established Investigator of the American Heart Association.
# Associate of the Medical Research Council of Canada.

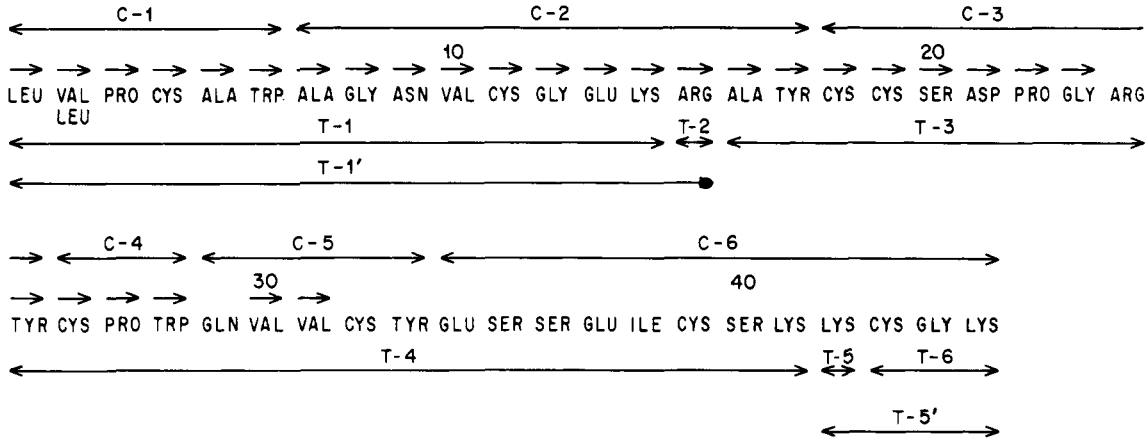

FIGURE 1: The complete amino acid sequence of ragweed allergen Ra5. Arrows to the right above each residue indicate the sequence obtained on the whole protein. The location of each tryptic (T) and chymotryptic (C) peptide is indicated as well.

formed in 1% NH_4HCO_3 buffer (pH 8.1) at 37° for 1 hr at an enzyme to substrate ratio of 1:100. Chymotryptic digestion was performed under similar conditions, but for 4 hr.

Analytical Methods. High voltage paper electrophoresis was carried out as described by Crumpton and Wilkinson (1965). Peptides were stained for arginine and tryptophan by the Sakaguchi and Ehrlich reagents, respectively (Easley, 1965). The peptides containing radioactive *S*-carboxymethylcysteine (CMCys)[1] were detected after overnight radioautography with Kodak BB-54 X-Ray film. Peptides were eluted and hydrolyzed *in vacuo* for 22 hr at 106° in 6.0 N HCl and analyzed on a Beckman Model 121 amino acid analyzer. Whole Ra5 was similarly analyzed and the results presented in Tables II and IV are averages of five analyses.

Whole Ra5 was tested for free sulhydryl groups by the method of Rohrbach *et al.* (1973) and none were detected.

Sequencing Procedures. A. AUTOMATED SEQUENCER. The use of the automated Beckman sequencer has been previously described (Capra, 1971; Capra and Kehoe, 1974). A dimethylallylamine buffer system was employed throughout. The Model 890A sequencer has been updated with an undercut cup and a nitrogen flush system. Certain of the peptides were treated with 4-sulfophenyl isothiocyanate (sPhSCN) (Pierce Chemical) prior to sequencing. Every phenylthiohydantoin (PTH) derivative was studied by gas chromatography (Pisano and Bronzert, 1969), thin-layer chromatography (Summers *et al.*, 1973), and ^{14}C scintillation counting (Capra *et al.*, 1972).

B. END GROUP ANALYSIS. Carboxy-terminal residues were determined by carboxypeptidase digestion as described by Ozols and Strittmatter (1968). Carboxypeptidase A digestion of whole Ra5 released no amino acids while carboxypeptidase B released only lysine. The dansyl-Edman procedure was used to establish the amino terminal residue in certain peptides (Gray, 1967). Dns-amino acids were resolved by thin-layer chromatography on 5 cm × 5 cm polyamide sheets (Woods and Wang, 1967).

Table I: Automated Sequencer Analysis of ^{14}C-Carboxymethylated Ra5.

Position	Gas Chromatography −Silylation	Gas Chromatography +Silylation	Tlc	Cpm	Sequence
1	L/I	L	L/I	225	Leu
2	V and L/I	V and L	L/I and V	121	Val (Leu)
3	P/T		P	351	Pro
4	S/C		C	13,798	Cys
5	A		A	5,149	Ala
6		W	W	1,250	Trp
7	A		A	402	Ala
8	G		G	330	Gly
9			N	206	Asn
10	V		V	283	Val
11	S/C		C	6,488	Cys
12	G		G	4,365	Gly
13		E	E	858	Glu
14			K	720	Lys
15			R	480	Arg
16	A		A	402	Ala
17		Y	Y	745	Tyr
18	S/C		C	8,622	Cys
19	S/C		C	12,617	Cys
20	S/C		S	5,788	Ser
21		D	D	3,020	Asp
22	P/T		P	1,264	Pro
23	G		G	901	Gly
24				690	
25		Y	Y	866	Tyr
26	S/C		C	4,228	Cys
27	P/T		P	3,651	Pro
28		W	W	2,406	Trp
29				1,215	
30	V		V	574	Val
31	V		V	494	Val

Results

Automated Sequencer Analysis of ^{14}C-Carboxymethylated Ra5. The initial sequencer analysis of Ra5 allergen (0.25 μmol) identified 29 of the first 31 residues (Table I). Only positions 24 and 29 did not give an unequivocal identification. A homogeneous sequence was obtained with the exception of position 2 where valine and leucine were found in a molar ratio of 2:1. The amino acid composition of the entire molecule had indicated that both leucine and valine were present in nonintegral amounts, and the tryptic (Table II) and chymotryptic (Table IV) digests confirmed that single peptides containing nonintegral quantities of both of

[1] Abbreviations used are: CMCyS, S-carboxymethylcysteine; sPhSCN, 4-sulfophenyl isothiocyanate.

Table II: Tryptic Peptides of Ra5.

	T-1	T-1′	T-2	T-3	T-5	T-5′	T-6	T-P	Total (1′ + 5′ + P)	Composition (AAA)
Lysine	1.0	1.1			1	1.9	1.0	1.1	4.1	3.7
Arginine		1.0	1	1.0				0.9	1.9	1.9
CMCys	1.5	1.7		1.4		0.6	0.7	2[a]	4.3	5.9
Aspartic acid	1.0	1.0		1.0				1.0	2.0	1.8
Serine				1.1				4.0	4.0	3.9
Glutamic acid	1.1	1.0						3.6	4.6	3.9
Proline	1.0	1.0		1.0				1.9	2.9	3.0
Glycine	1.8	2.0		1.0		1.0	1.2	1.1	4.1	3.8
Alanine	1.8	2.0		1.0				0.9	2.9	2.8
Valine	1.5	1.5						1.7	3.2	3.7
Methionine										
Isoleucine								0.9	0.9	0.9
Leucine	1.2	1.3							1.3	1.5
Tyrosine				0.8				2.5	2.5	2.8
Phenylalanine										
Tryptophan	+	+								
Yield (nmoles)	140	154	127	114	103	186	128	+	2	1.8

[a] Determined by amino acid analysis. Specific activity indicated that this peptide contained five CMCys residues.

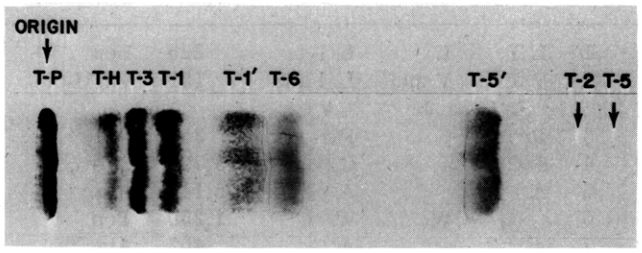

FIGURE 2: Radioautography of the pH 3.5 electrophoretogram of the tryptic digest of Ra5. Peptide T-P was at the origin while all the other tryptic peptides were separated completely in a single dimension. Both free lysine and free arginine were detected by ninhydrin stain (peptides T-5 and T-2). The anode is to the right. T-H proved to be identical in composition to T-3 but present in lower yield.

these amino acids could be isolated. Two additional sequencer runs on 100 nmol of whole Ra5 showed that the heterogeneity at position 2 could be confirmed in two separate allergen preparations.

The Tryptic Peptides of Ra5. After reduction and ^{14}C carboxymethylation, 0.8 μmol of Ra5 was digested with trypsin for 1 hr at 37°. The resulting peptides were separated by high voltage paper electrophoresis at pH 3.5. The radioautograph of the separated peptides is shown in Figure 2, and the composition of the isolated peptides listed in Table II. The finding of free arginine was expected from the sequence at positions 14 and 15. T-1 and T-1′ were identifiable as the N-terminal peptide and the heterogeneity at position two was evident from their compositions. Since the C-terminal residue of intact Ra5 had been shown to be lysine, the finding of free lysine (peptide T-5) implied that the two remaining basic residues were adjacent. Peptide T-P, which remained at the origin, proved to be residues 16–42, and resulted from the incomplete cleavage of the Arg–Tyr bond at positions 24–25.

A precipitate was observed when the lyophilized tryptic digest was dissolved in water. This proved to be mainly the peptide from residues 25–41, with a small amount of contamination by the peptide containing residues 16–41. After

Table III: Automated Sequencer Analysis of ^{14}C-Carboxymethylated, sPhSCN Treated T-4 Peptide of Ra5.

Position	Gas Chromatography −Silylation	Gas Chromatography +Silylation	Tlc	Cpm	Sequence
1				550	
2	S/C		C	12,835	Cys
3	T/P		P	4,295	Pro
4		W	W	3,410	Trp
5		Q	Q	1,620	Gln
6	V		V	1,085	Val
7	V		V	905	Val
8	S/C		C	6,615	Cys
9		Y	Y	3,010	Tyr
10		E	E	1,675	Glu
11	S/C		S	1,075	Ser
12	S/C		S	900	Ser
13		E	E	420	Glu
14	L/I	I	L/I	290	Ile

an additional course of tryptic digestion under the same conditions as above, but with the digestion time extended to 4 hr, the mixture was treated with sPhSCN, and a 100-nmol sample was placed in the sequencer. Since the peptide 16–24 was shorter, present in a lower yield, and contained a C-terminal arginine, it rapidly washed out of the sequencer cup and the sequence of the sPhSCN treated peptide 25–41 (referred to as T-4) could be sequenced for 14 of its 17 positions (Table III).

The sequencer run on the intact protein, plus the analysis of the tryptic peptides mentioned above, located peptides T-5, T-5′, and T-6 at the C terminus of the molecule. Both T-6 and T-5′ were subjected to automated sequence analysis before and after treatment with sPhSCN and their complete sequence was established.

Since the C-terminal residue in the molecule was known

Table IV: Chymotryptic Peptides of Ra5.

Amino Acid	Peptide C-1	C-2	C-3	C-4	C-5	C-6	Total	Composition (AAA)
Lysine		1.0				2.9	3.9	3.7
Arginine		1.0	1.0				2.0	1.9
CMCys	0.8	1.0	1.9	1.1	1.0	2.0	7.8	5.9
Aspartic acid		1.1	1.0				2.1	1.8
Serine			1.0			2.8	3.8	3.9
Glutamic acid		1.1			1.1	2.0	4.2	3.9
Proline	1.0		1.0				3.0	3.0
Glycine		2.1	1.0			1.1	4.2	3.8
Alanine	1.1	2.0					3.1	2.8
Valine	0.7	1.0			1.6		3.3	3.7
Isoleucine						1.0	1.0	0.9
Leucine	1.4						1.4	1.5
Tyrosine		1.0	1.0		1.0		3	2.8
Tryptophan	+			+			2	1.8
Mobility pH 6.5 Asp = 1, Ser = 0)	0.29	0	0.60	0.47	0.37	0.21		
Yield (nmol)	115	128	135	128	103	115		
Total residues	6	11	8	3	5	12		

Table V: Automated Sequencer Analysis of ^{14}C-Carboxymethylated, sPhSCN Treated C-6 Peptide of Ra5.

Position	Gas Chromatography −Silylation	Gas Chromatography +Silylation	Tlc	Cpm	Sequence
1				586	
2	S/C		S	146	Ser
3	S/C		S	103	Ser
4		E	E	283	Glu
5	L/I	I	L/I	458	Ile
6	S/C		C	11,126	Cys
7	S/C		S	1,461	Ser
8				380	
9				289	
10	S/C		C	7,200	Cys
11			G	1,212	Gly

to be lysine, with the exception of the order of the Cys-Ser in position 15 and 16 of peptide T-4 (positions 39–40 in the protein), the entire molecule had been sequenced with less than 1.1 μmol of material.

The Chymotryptic Peptides of Ra5. Reduced and ^{14}C-carboxymethylated Ra5 (0.25 μmol) was subjected to digestion with chymotrypsin and the resulting peptides separated by high voltage paper electrophoresis in one dimension at pH 6.5. The composition of the eluted peptides is shown in Table IV. The compositions agreed with the established sequence and all the peptides could be provisionally placed. Peptide C-6, the 12 residue C-terminal peptide, contained the only unestablished portion of the sequence so this peptide was subjected to automated sequence analysis after treatment with sPhSCN. Since this peptide contained three lysine residues which could be derivatized, the yields at each step in the sequence analysis were especially high and even the penultimate residue was identifiable. As previously, the N-terminal residue was not identified because of the derivatization. Also, positions 8 and 9 were not identified since the derivatized lysine residues occupying these positions are not detected in our identification systems. The radioactivity obtained at positions 6 and 10 of this peptide indicates the extremely high repetitive yield of the sequencer run on this short peptide.

The heterogeneity at position 2 is further evident from the composition of peptide C-D where 0.7 mol of valine and 1.4 mol of leucine were found. No significant deviation from integral values was found for any other chymotryptic peptides, with the exception of peptide C–C which contained the hydrolysis resistant Val–Val sequence.

Discussion

This paper presents the first complete amino acid sequence of a pollen allergen. The derivation of the total sequence of ragweed pollen allergen Ra5 on less than 1.4 μmol of material indicates that structural analysis is increasingly possible on relatively small amounts of highly purified material when a combination of automated and manual sequencing techniques and highly sensitive detection systems are employed.

The structural analysis reported here establishes the homogeneity of Ra5 since a single amino acid was found for every position with the exception of position 2. In three separate sequencer runs (of 250, 100, and 100 nmol) approximately ⅔ valine and ⅓ leucine was found at position 2. The latter two studies were done on separate preparations of whole Ra5. In addition, in both the tryptic and chymotryptic peptide analysis, the N terminal peptide contained nonintegral values for these two amino acids at this position. The significance of this observation is unclear at present, but in a study of the fragments of this allergen, it will be important to establish if both forms have similar biological activities. It is noteworthy that the alternative forms of the protein had not been detected previous to the sequence analysis. Neither forms of the N terminal tryptic nor chymotryptic peptides could be separated by the techniques

used in this study although this is admittedly not surprising in view of the similar size and charge properties of the alternative amino acids involved.

The establishment of the complete amino acid sequence of Ra5 should provide a basis for many structure–function correlative experiments in the field of immediate hypersensitivity. A long series of experiments by King *et al.* (1964, 1967, 1974) has established the conditions under which antigen E either retains or loses biological activity with various chemical modifications (acetylation, succinylation, glycinamide, taurine modification, butyrimidation, etc.). Similar experiments with Ra5 would have the advantage that the precise chemical structures being modified would be identifiable. Removal of the C terminal lysine residue alone may be easily accomplished with carboxypeptidase B. Study of the protein before and after such treatment might be of interest in establishing the function of the C terminal basic sequences in the molecule. The presence of these three basic residues in the C terminal five amino acids may, for example, be essential to the membrane fixation and transport of the Ra5 molecule to the submucosal lymphoid cells where sensitization might occur in a genetically predisposed individual. The 26 amino acid sequence of melittin contains two lysine and two arginine residues in its C terminal 6 amino acid residues (Habermann, 1972) and many other biologically active molecules contain similar C terminal basic sequences which are essential to their function (Goldstein, 1974).

In addition to the C terminal concentration of basic amino acids, the Ra5 sequence contains certain other unusual features. The presence of 8 half-cystines in a molecule of only 45 amino acid residues is exceptional. The occurrence of five aromatic amino acids within the same structure would indicate that a high degree of covalent and noncovalent cross-linking may result in a rather rigid three-dimensional structure which might enhance the immunogenicity of the molecule. The relatively small size of this allergen would suggest that it contains few immunodominant groups. In consequence, fragmentation studies might separate the carrier active from allergenically active fragments.

Amino acid sequence analysis of protein allergens has been only fragmentary to date. Elsayed *et al.* (1973) recently reported the N-terminal 17 amino acid residues of a 50–53 residue fragment of a major protein allergen derived from cod muscle. This fragment was obtained by trypsin digestion of the lysine modified protein and had been previously shown to be capable of mediating much of the biological activity of the native molecule. They noted the presence of the sequence Arg-Ala or Lys-Ala in three locations within the fragment, and they speculated that these unusual and repetitive sequence doublets might represent the allergenically active regions of the molecule. It is of interest that Ra5 contains an Arg-Ala doublet at positions 15–16. No other similarities could be found between the 17 residues reported for the cod allergen and the amino acid sequence of Ra5.

The amino acid composition of several naturally occurring allergens is presently available (reviewed in Marsh, 1974). No consistent structural/compositional feature is obvious among the allergens of rye (*Lolium perenne*), short ragweed (*Ambrosia elatior*) (including antigens E, Ra3 and Ra5), cod, or honey bee, except relatively high content of half-cystines. Only complete amino acid sequence analysis of several of these distinct allergens will establish whether they possess common structural features.

Acknowledgments

We thank Ms. Donna Atherton for expert technical assistance and Dr. Pierre Querinjean for assistance with the thin-layer chromatography system. We also thank Drs. David Marsh and Lawrence Lichtenstein who provided many helpful discussions and critically read the manuscript.

References

Capra, J. D. (1971), *Nature (London), New Biol. 230,* 61.

Capra, J. D., and Kehoe, J. M. (1974), *Proc. Nat. Acad. Sci. U.S. 71,* 845.

Capra, J. D., Kehoe, J. M., Kotelchuck, D., Walter, R., and Breslow, E. (1972), *Proc. Nat. Acad. Sci. U.S. 69,* 431.

Crumpton, M. J., and Wilkinson, J. M. (1965), *Biochem. J. 94,* 545.

Easley, C. W. (1965), *Biochim. Biophys. Acta 107,* 386.

Elsayed, S., Sletten, K., and Aas, K. (1973), *Immunochemistry 10,* 701.

Goldstein, G. (1974), personal communication.

Gray, W. R. (1967), *Methods Enzymol. 2,*

Griffiths, B. W., and Brunet, R. (1971), *Can. J. Biochem. 49,* 396.

Habermann, E. (1972), *Science 177,* 314.

King, T. P. (1974), *Proc. Int. Congr. Allergology, 8th* (in press).

King, T. P., Norman, P. S., and Connell, J. J. (1964), *Biochemistry 3,* 458.

King, T. P., Norman, P. S., and Lichtenstein, L. M. (1967), *Biochemistry 6,* 1992.

Lapkoff, C. B., and Goodfriend, L. (1971), *Proc. Can. Fed. Biol. Soc. 14,* 36.

Lapkoff, C. B., and Goodfriend, L. (1974), *Int. Arch. Allergy Appl. Immunol. 46,* 215.

Lichtenstein, L. M., Roebber, M., and Goodfriend, L. (1973), *J. Allergy Clin. Immunol. 51,* 285.

Marsh, D. G. (1974), in The Antigens, Vol. III, Sela, M., Ed., New York, N.Y., Academic Press.

Marsh, D. G., Bias, W. B., Goodfriend, L., and Ishizaka, K. (1973b), *Fed. Proc., Fed. Amer. Soc. Exp. Biol. 32,* 1000.

Marsh, D. G., Bias, W. B., Hsu, S. H., and Goodfriend, L. (1973a), *Science 179,* 691.

Mole, L. E. Goodfriend, L., Lapkoff, C. B., Kehoe, J. M., and Capra, J. D. (1974), *Fed. Proc., Fed. Amer. Soc. Exp. Biol. 33,* 751.

O'Donnell, I. J., Frangione, B., and Porter, R. R. (1970), *Biochem. J. 116,* 261.

Ozols, J., and Strittmatter, P. (1968), *J. Biol. Chem. 243,* 3367.

Pisano, J. J., and Bronzert, T. J. (1969), *J. Biol. Chem. 244,* 5597.

Press, E. M., Piggot, P. J., and Porter, R. R. (1966), *Biochem. J. 99,* 356.

Rohrbach, M. S., Humphries, B. A., Yost, Jr., F. J., Rhodes, W. G., Boatman, S., Hiskey, R. G., and Harrison, J. H. (1973), *Anal. Biochem. 52,* 127.

Sherman, W. B. (1968), Hypersensitivity. Mechanisms and Management, Philadelphia, Pa., W. B. Saunders.

Summers, M. R., Smythers, G. W., and Oroszlan, S. (1973), *Anal. Biochem. 53,* 624.

Underdown, B. J., and Goodfriend, L. (1969), *Biochemistry 8,* 980.

Wodehouse, R. P. (1971), Hayfever Plants, 2nd ed, New York, N.Y., Hafner Publishing Co.

Woods, R. R., and Wang, K. T. (1967), *Biochim. Biophys. Acta 133,* 369.