Sainfoin lectin : molecular characterization and amino acid sequence
by Rozenn Nicole Kouchalakos

A thesis submitted in partial fulfillment of the requirements for the degree of DOCTOR OF PHILOSOPHY in Biochemistry
Montana State University
© Copyright by Rozenn Nicole Kouchalakos (1982)

Abstract:
Sainfoin lectin, a glycoprotein from the legume Onobrychis viciifolia, Scop., has been characterized according to its complete amino acid sequence, carbohydrate binding, metal content and molecular stability. The protein subunit was a single polypeptide chain containing 236 amino acid residues. Nitrogen and carboxyl termini were alanine and threonine, respectively, and the single site of carbohydrate attachment was asparagine 118. Nb isolectin forms were detected. The sequence showed extensive homology with other leguminous lectins. Dissociation constant with D-glucose was 3.52 mM by equilibrium dialysis, and similar values were obtained by affinity electrophoresis. One carbohydrate binding site was determined per protein subunit. Sainfoin lectin contained one Ca++ and lesser amounts of Mn++ and Mg++ per subunit. The protein was normally heat and trypsin resistant. Succinylation of lysine residues and carboxymethylation of the single cysteine resulted in a modified protein that, upon trypsin hydrolysis, yielded manageable peptide fragments.

to my husband David

with love

SAINFOIN LECTIN: MOLECULAR CHARACTERIZATION AND AMINO ACID SEQUENCE

by

ROZENN NICOLE KOUCHALAKOS

A thesis submitted in partial fulfillment
of the requirements for the degree

of

DOCTOR OF PHILOSOPHY

in

Biochemistry

Approved:

Kenneth D. Hapner
Chairperson, Graduate Committee

Edwin H. Abbott
Head, Major Department

Michael Malone
Graduate Dean

MONTANA STATE UNIVERSITY
Bozeman, Montana

August, 1982

ACKNOWLEDGMENT

I would like to express my gratitude to Dr. K. D. Hapner for his constant help throughout this study. His assistance and advice were most valuable, and his friendly moral support was appreciated.

I wish to acknowledge a fruitful collaboration with Dr. R. A. Bradshaw who provided funds, expertise and technical assistance for the sequence analyses.

I am thankful to Dr. J. E. Robbins for his interest in my research project and numerous helpful discussions.

During the course of this study, I received two travel grants from the Dean of the Graduate School of Montana State University and financial support from Montana Agricultural Experiment Station and Montana State University Department of Chemistry.

TABLE OF CONTENTS

Page

LIST OF TABLES x

LIST OF FIGURES. xi

ABSTRACT xiv

CHAPTER ONE. Introduction 1

CHAPTER TWO. Goals and Objectives 16

CHAPTER THREE. Experimental Procedure 18

Materials 18
Purification of Sainfoin Lectin (SL). 19
Gel Filtration of SL. 20
Affinity Chromatography of
SL at Various pH Values. 21
Anion Exchange Chromatography of SL 21
Immuno Double Diffusion Agar Gels 22
Metal Analysis. 22
Equilibrium Dialysis. 23
Agglutination Assay 24
Standard Electrophoretic Procedures 24
Synthesis of Allyl Alpha-Glycosides 25
Synthesis of Glycosyl Copolymers. 26
Affinity Electrophoresis. 27
Performic Acid Oxidation. 28
Carboxymethylation. 29
Reaction with
5,5'-Dithiobis-(2-nitrobenzoic Acid) 29
Succinylation 30
Trinitrobenzene Sulfonic Acid (TNBS) Assay. 30
Cyanylation 31
Synthesis of
2-Nitro-5-thio[^{14}C]cyano Benzoic Acid (NTCB) 31
Cleavage at Cysteine. 32
Gel Filtration of SL Cysteine Cleavage Product. 32

(continued)

TABLE OF CONTENTS (continued)

Page

Gel Filtration of Succinylated SL Cleaved at Cysteine . . 33
Partial Acid Hydrolysis 33
Digestion with Trypsin. 34
Digestion with Thermolysin. 34
Digestion with Carboxypeptidases. 35
Fractionation of SL Tryptic Digest at Low pH. 35
Gel Filtration of the pH 3.5 Soluble
SL Tryptic Peptides. 36
Repurification of T2 and T10 on LH20. 36
Cation Exchange Chromatography
of the pH 3.5 Soluble SL Tryptic Peptides. 37
Gel Filtration of the pH 3.5 Insoluble
SL Tryptic Peptides. 38
Repurification of T4 and T7 by
High Performance Liquid Chromatography (HPLC). 38
Digestion of T6 with Beta Xylosidase. 39
Boron Chromatography of T6. 39
Chromatography of T6 on WGA-Sepharose 39
Carbohydrate Analysis 39
Gel Filtration of the Thermolytic Peptides. 40
Cation Exchange Chromatography
of the Thermolytic Peptides. 40
Arginine Detection. 41
Ninhydrin Analysis. 41
Amino Acid Analysis 42
End Group Analysis. 42
Sequence Analysis 43

CHAPTER FOUR. Results 46

Section 1. Sainfoin Lectin Stability 46

Gel Filtration of SL 46
Solubility of SL in Aqueous Solution 49
Affinity Chromatography of SL
at Various pH Values. 50
Trypsin Digestion of SL. 52
Characteristics of Succinylated SL 54

(continued)

TABLE OF CONTENTS (continued)

Page

Section 2. Studies on Apparent Heterogeneity of Sainfoin Lectin. 62

Ion Exchange Chromatography of SL. 62
Analyses of SL Ionic Species 65
Affinity Electrophoresis of SL at pH 4.0 69
Affinity Electrophoresis of SL at pH 9.5 72

Section 3. Ligand Interaction Studies. 75

Metal Analyses 75
Estimation of the Dissociation Constant for SL-Glucose Complex by Affinity Electrophoresis . . . 77
Equilibrium Dialysis 82

Section 4. Primary Structure Determination 86

Partial Acid Hydrolysis. 86
Cleavage at Cysteine 90
Trypsin Cleavage at Arginine 99
Purification of the pH 3.5 Soluble SL Tryptic Peptides 104
Purification of the pH 3.5 Insoluble SL Tryptic Peptides 111
Analyses of SL Tryptic Peptides. 116

Peptide T1. 116
Peptide T2. 117
Peptide T3. 118
Peptide T4. 119
Peptide T5. 119
Peptide T6. 119
Assignment of Carbohydrate Position 119
Other Characteristics of SL Tryptic Glycopeptide 120
Peptide T7. 121
Peptide T8. 122
Peptide T9. 122

(continued)

TABLE OF CONTENTS (continued)

Page

Peptide T10 122
Peptide T11 123
Peptide T12 123

Cleavage of SL with Thermolysin. 123

Analyses of Arginine-Containing
SL Thermolytic Peptides 127

Peptide Th1 127
Peptide Th2 128
Peptide Th3 128
Peptide Th4 129
Peptide Th5 129
Peptide Th6 129
Peptide Th7 130

Alignment of SL Tryptic Peptides 130

CHAPTER FIVE. Discussion. 135

Sequence Strategy--Plan of Attack. 135
Production and Purification
of Tryptic Peptides 136
Sequence Analyses. 137
Alignment of Tryptic Peptides. 139
Sainfoin Lectin Amino Acid Sequence. 142
Comparison of SL, Concanavalin A, Favin,
Lentil Lectin and Soybean Agglutinin. 144
Structure Conservation 151

Influence of Circular Permutation
and Subunit Structure. 151
Secondary Structures. 153
Metal Binding Sites 154
Carbohydrate Binding Site 155
Hydrophobic Cavity. 156
Monomer-Monomer and Dimer-Dimer Contacts. . . .157

(continued)

TABLE OF CONTENTS (continued)

Page

Carbohydrate Attachment Site. 159
Structure Conservation, Summation 160

Other Leguminous Lectins 161

CHAPTER 6. Summary and Conclusions. 164

REFERENCES 170

APPENDIX. Amino Acid Symbols. 177

LIST OF TABLES

Table		Page
1.	Behavior of sainfoin lectin on Ultrogel AcA44	48
2.	Amino acid composition of sainfoin lectin ionic species	67
3.	Metal analyses of sainfoin lectin by atomic absorption	75
4.	K_d values for SL-glucose complexes determined by electrophoresis at pH 4.0	79
5.	Amino acid composition of sainfoin lectin	143
6.	Identities among ConA, SL, SBA, favin and LL	148

LIST OF FIGURES

Figure		Page
1.	Calibration of Ultrogel AcA44 column.	47
2.	Affinity chromatography of SL on mannose-Sepharose at different pH values.	51
3.	Affinity chromatography of 90% succinylated SL on mannose-Sepharose	55
4.	Gel filtration of 90% succinylated SL on Ultrogel AcA44	57
5.	Affinity chromatography of 100% succinylated SL on mannose-Sepharose	59
6.	Gel filtration of 100% succinylated SL on Ultrogel AcA44	60
7.	Gel filtration of preheated 100% succinylated SL on Ultrogel AcA44.	60
8.	Diethylaminoethylcellulose chromatography of SL	63
9.	Diethylaminoethylcellulose chromatography of fraction I of Figure 8.	64
10.	Immuno/double diffusion of SL ionic species against antiserum to native SL.	66
11.	Affinity electrophoresis at pH 4.0 of SL.	70
12.	Affinity electrophoresis at pH 9.5 of SL.	73
13.	Plot (L/Lo-L) *vs.* C derived from the affinity electrophoresis of Figure 11.	78
14.	Determination of the dissociation constant of glucose to SL by equilibrium dialysis	83

(continued)

LIST OF FIGURES (continued)

Figure		Page
15.	Sodium dodecylsulfate electrophoresis of SL partial acid hydrolysate.	87
16.	Gel filtration of SL partial acid hydrolysate on Biogel A5m.	88
17.	Sodium dodecylsulfate electrophoresis of SL cleaved at cysteine	91
18.	Gel filtration of SL cleaved at cysteine on Sephadex G50.	92
19.	Gel filtration of succinylated SL cleaved at cysteine on Sephadex G25 in 0.02 M NH_4HCO_3, pH 6.3. . . .	94
20.	Gel filtration of succinylated SL cleaved at cysteine on Sephadex G25 in 3M guanidinium chloride, pH 7.0.	95
21.	Production and purification of sainfoin lectin tryptic peptides.	100
22.	Solubility of SL tryptic peptides as a function of pH.	102
23.	Gel filtration of pH 3.5 soluble SL tryptic peptides on Sephadex G25.	105
24.	Repurification of peptide T10 of Figure 23 on Sephadex LH20	107
25.	Gel filtration of pH 3.5 soluble SL tryptic peptides on Sephadex G25 in 0.01 M acetic acid	108
26.	Repurification of peptide T2 from Figure 25 on Sephadex LH20	109

(continued)

LIST OF FIGURES (continued)

Figure — Page

27. Cation exchange chromatography of pH 3.5 soluble SL tryptic peptides 110

28. Gel filtration of pH 3.5 insoluble SL tryptic peptides on Sephadex G50. 112

29. Repurification of peptide T7 by reversed-phase HPLC. 113

30. Repurification of peptide T4 by reversed-phase HPLC. 115

31. Production and purification of thermolytic overlaps. 125

32. Gel filtration of SL thermolytic peptides on Sephadex G25. 126

33. Reconstruction of seven blocks of continuous sequence from results obtained on SL component peptides. 131

34. Alignment of SL peptides using favin as a template. . . .134

35. Comparison of the amino acid sequences of ConA, SL, SBA, favin and LL 145

36. Schematic alignment of SL and favin with ConA 147

ABSTRACT

Sainfoin lectin, a glycoprotein from the legume Onobrychis viciifolia, Scop., has been characterized according to its complete amino acid sequence, carbohydrate binding, metal content and molecular stability. The protein subunit was a single polypeptide chain containing 236 amino acid residues. Nitrogen and carboxyl termini were alanine and threonine, respectively, and the single site of carbohydrate attachment was asparagine 118. No isolectin forms were detected. The sequence showed extensive homology with other leguminous lectins. Dissociation constant with D-glucose was 3.52 mM by equilibrium dialysis, and similar values were obtained by affinity electrophoresis. One carbohydrate binding site was determined per protein subunit. Sainfoin lectin contained one Ca^{++} and lesser amounts of Mn^{++} and Mg^{++} per subunit. The protein was normally heat and trypsin resistant. Succinylation of lysine residues and carboxymethylation of the single cysteine resulted in a modified protein that, upon trypsin hydrolysis, yielded manageable peptide fragments.

CHAPTER ONE

INTRODUCTION

Lectins are multivalent proteins and glycoproteins with carbohydrate binding abilities (Goldstein & Hayes, 1978). This family of molecules does not include antibodies against carbohydrate structures, glycosidases or monovalent carbohydrate binding proteins. No lectin has been convincingly described that exhibits enzymatic activity.

Lectins were originally detected in plants. A review on the history of lectin discovery and uses is available by Coulet (1979). Lectins are now thought to be ubiquitous. They are found in plants, microorganisms, and animals, including both vertebrates (Simpson et al., 1978) and invertebrates (Cohen, 1974). Several recent reviews dealing with biochemical properties of purified lectins, their interaction with cell surfaces and possible functions have appeared (Barondes, 1981; Goldstein & Hayes, 1979; Brown & Hunt, 1978; Lis & Sharon, 1977; Liener, 1976).

Lectins bind carbohydrate-containing receptors of cell surfaces. Their multivalency allows cross linking of several cells, resulting in agglutination. Although agglutination per se is not well understood (Brown & Hunt, 1978), cell agglutination provides an

easy detection assay for lectin activity. Erythrocytes are generally employed as test cells because they are readily available. The assay is semiquantitative. Erythrocytes are added to a serial dilution of agglutinin and the agglutination pattern is scored visually, usually after one hour. The reciprocal of the highest dilution showing agglutination is referred to as hemagglutination titer.

Lectins bind specific carbohydrate structures, hence the term lectin derived from the Latin legere (to choose). The carbohydrate specificity of a lectin is usually determined by measuring the capacity of monosaccharides or oligosaccharides to inhibit lectin-induced hemagglutination. Specific inhibition by a particular monosaccharide is one basis for lectin classification (Goldstein & Hayes, 1978). There are five families of lectins which show binding specificity: D-mannose (D-glucose), N-acetyl glucosamine, N-acetyl galactosamine, D-galactose and L-fucose. This classification is an oversimplification of lectin binding specificity, but it permits a comparison of the properties of lectins within a group. The specificity of lectin binding is illustrated by the range of erythrocytes agglutinated by a given lectin. Some lectins will agglutinate erythrocytes of a single species exclusively. Some lectins that interact with human erythrocytes show blood group and

blood subgroup specificity (Brown & Hunt, 1978). Specific oligosaccharides interacting with given lectins were compiled by Uhlenbruck (1981). Detailed studies of lectin carbohydrate binding sites and lectin receptors are discussed in the Goldstein & Hayes review (1978).

Lectins are purified by affinity chromatography, that takes advantage of their specific saccharide binding. Most lectins purified to date are derived from plant seeds. The seeds are ground, extracted with a saline buffer and the lectin adsorbed to its specific carbohydrate immobilized on a support such as Sepharose. After washing out all the contaminants the lectin is eluted with free carbohydrate and is thereby recovered in a highly pure state in a single step. Purification schemes for 20 lectins can be found elsewhere (Brown & Hunt, 1978).

The binding of lectins to cell surface receptors induces various biological activities such as cell agglutination, mitogenic stimulation of lymphocytes, inhibition of phagocytosis, toxicity, inhibition of growth of tumor cells, inhibition of fertilization of ovum by sperm, insulin-like effects on fat cells, induction of platelet release reaction, and others (Nicolson, 1974). Because plant lectins are readily available and easily purified they have

become important medical and research tools in immunology and cell biology.

Some plant lectins bind the surface of quiescent lymphocytes triggering growth, proliferation, and blast cell formation. The blast cells synthesize immunoglobulins as if they had been stimulated by antigens. The morphological and biochemical changes occuring in lectin-stimulated lymphocytes provide a useful model for antigen-induced immune reactions that take place in vivo. Lectins will also induce the differentiation of lymphocytes into cytotoxic T cells with the ability to lyse certain target cells. These cells can be obtained from nonimmune animals and are of great interest for the study of the mechanisms of cellular toxicity. The different steps involved in lymphocyte activation by lectins, the factors affecting this activation and what is known of the activation mechanism has recently been reviewed by Lis & Sharon (1977).

The agglutinability of many cell types by lectin is used to probe molecular events taking place at the cell surface. One most important use of lectins is the study of changes in agglutinability accompanying neoplastic transformation. Transformed cells are often found to be agglutinated at much lower concentration of lectin than is required for the corresponding normal cell type. Although the correlation between increased agglutinability and transformation

does not always hold, these studies indicate that in some cases transformation is accompanied by changes in the cell surface which result in facilitated agglutination. The understanding of selective agglutinability of transformed cells may lead to increased understanding of the role of the cell surface in uncontrolled growth, metastasis and resistance to immunological attack. The importance of the use of lectins in this field therefore cannot be overemphasized. A review of increased lectin agglutinability of transformed cells *in vitro* and *in vivo* and of the state of our knowledge of the mechanism of cell agglutination has appeared recently (Brown & Hunt, 1978). Lectins are also used to probe cell surface changes associated with differentiation and development of cells and tissues (Brown & Hunt, 1978).

Besides their use in investigating fundamental problems, certain lectins are utilized routinely in clinical laboratories. Lectins that are specific for human erythrocytes of a given blood group or blood subgroup are used for blood type screening and secretor diagnosis (Brown & Hunt, 1978). Lectins thatare mitogenic for lymphocytes are useful for the detection of congenital and acquired immunologic deficiencies, the monitoring of the effect of various immunosuppressive and immunotherapeutic manipulations and

the detection of genetic abnormalities in chromosome analysis (Lis & Sharon, 1977).

Lectins can be coupled to solid supports such as agarose beads and serve for affinity chromatography of glycoproteins and polysaccharides. Several immobilized lectins are available commercially to this end. The property of certain lectins to bind specific cells is also exploited for specific fractionation of cells (Lis & Sharon, 1977).

In spite of the extensive use of lectins, their physiological roles in the organisms from which they are derived remains hypothetical. Several lines of evidence indicate that lectins may function in recognition processes, and these have been reviewed by Barondes (1981). Lectins were suggested to play a role in the aggregation of slime molds (Barondes, 1981). Slime molds are eukaryotes that change from discrete amoebae to aggregating colonies upon starvation. The development of cell-cell adhesiveness is correlated with the synthesis of lectins detectable on the cell surface, hence the hypothesis that lectins mediate cell cohesion. A thorough review of slime mold lectin properties and of the evidence bearing on their biological role was done by Bartles et al. (1982).

Lectins that may be involved in serum glycoprotein clearance have been found at the cell surface of mammalian hepatocytes

(Ashwell & Harford, 1982). The most studied is a galactose binding lectin that binds asialoglycoproteins. When serum glycoproteins are desialylated, thereby exposing their galactose residues, their survival time in the circulation is drastically reduced. The desialylated glycoproteins presumably bind the hepatic galactose-specific receptor, resulting in internalization of ligand-receptor by endocytosis and degradation of the glycoproteins in the lysosomes. Biologically reactive glycoproteins may be cleared from the serum by this mechanism after they serve a particular physiological function (Ashwell & Harford, 1982). A receptor for D-mannose/N-acetylglucosamine and one for L-fucose are also found at the surface of some hepatocytes. They may also play a role in glycoprotein clearance.

A phosphomannosyl receptor is found in liver, spleen, lung, kidney, and testes. It is believed to play a role in the uptake of lysosomal hydrolases and is suggested to direct newly synthesized acid hydrolases to the lysosomes (Ashwell & Harford, 1982). Cells do not transfer proteins from the endoplasmic reticulum (ER) to lysosomes <u>via</u> an exclusively intracellular route. Newly synthesized lysosomal hydrolases are transported from the ER to the Golgi, packaged in secretion granules and secreted. They are then internalized by endocytosis and directed to the lysosomes (Herzog,

1981). It has been suggested that the phosphomannosyl receptor binds the secreted hydrolases, thereby initiating their endocytosis. Furthermore, intracellular phosphomannosyl receptors would guide the hydrolases to the lysosomes (Ashwell & Harford, 1982).

Several vertebrate lectins have been identified that become prominent at a specific stage in the development of individual tissues (Barondes, 1981). Chick embryonic muscle cells show a large increase in lectin levels correlating temporally with fusion of the cells. This suggests the lectin to mediate myoblast fusion. Another chick embryo lectin was speculated to be involved in neuronal maturation and a third one was prominent in embryonic liver and kidney. The three lectins and their potential role in the differentiative process have been reviewed by Barondes (1981). Another important finding is that, although the lectins are developmentally regulated in some tissues, they are especially prominent in others in adulthood. The same lectins may therefore be utilized in different ways (Barondes, 1981).

Plant lectins have also been suggested to be involved in recognition processes (Barondes, 1981). Lectins derived from leguminous plants may play a role in the rhizobium-legume symbiosis. The lectin-recognition hypothesis states that recognition at infection sites involves the binding of specific legume lectins to

unique carbohydrates found on the surface of the appropriate rhizobial symbiont. Evidence supporting the hypothesis has been reviewed by Dazzo (1980) and Barondes (1981). Models of the contribution of lectins to the infection process can be found in Bauer (1981). There is much support and no decisive negative evidence for the role of soybean lectin and clover lectin in binding specific rhizobium to root hairs. A lectin necessary for specific binding to strains nodulating alfalfa has also been described (Paau et al., 1981). The major exception is associated with the tropical legumes (jack bean, peanut, etc.). Tropical legumes undergo infection by a mechanism different from the root hair infection of temperate legumes (Bauer, 1981). The rhizobia intrude into the legume roots through voids created by lateral root emergence. The recognition processes in these legumes are less specific and have broad boundaries. The recognition hypothesis would not hold for those legumes whose lectins can bind indiscriminately to many rhizobia (Dazzo, 1980).

Legume lectins have been postulated by Hankins & Shannon (1978) to be plant enzymes based on their finding of a lectin with alpha galactosidase activity in mung bean. Later screening of a large number of legume species revealed the presence of alpha-galactosidases without hemagglutination activity in certain legumes

having well known lectins (Hankins et al., 1980). Hemagglutination is therefore a transient property of some legume alpha galactosidases, and lectins are not enzymes. These conclusions are supported by the recent purification from soybean of an alpha-galactosidase that has hemagglutinin properties and is different from soybean lectin (Del Campillo & Shannon, 1982).

Several other roles for plant lectins have been suggested for which there is much less experimental evidence. Leguminous lectins were suggested to act as storage proteins in seeds (Millerd, 1975). This hypothesis stemmed from the fact that legume lectins are present in large amount in the seeds and their level decreases dramatically upon germination. There is, however, no experimental data to support this role and the major storage proteins in seeds do not have lectin activity.

Lectins were suggested to protect legume seeds from attack by insect seed predators (Janzen et al., 1976). This hypothesis was based on the sole observation that addition of black bean lectin to the diet of bruchid beetle results in death of the bruchid larvae.

Lectins were speculated to function in transport of carbohydrates, particularly during extension growth of plant cells, a process requiring deposition of sugars for cell wall formation (Kauss & Glaser, 1974). A lectin in potato tubers has a

glycopeptide domain that bears close resemblance to the glycopeptides derived from the plant cell wall, suggesting that the lectin may be a precursor of the insoluble cell wall material (Ashford & Neuberger, 1980). None of these potential lectin involvements in cell wall growth was demonstrated experimentally.

Lectins were postulated to act as receptors for pathogens and to inhibit the growth of microorganisms (Sequeira, 1978). Wheat germ agglutinin (WGA) inhibits fungal growth (Mirelman et al., 1975). The finding that WGA is localized precisely at potential infection sites of germinating wheat (Mishkind et al., 1982) is consistent with a protective function of the lectin. A protective role against fungal infection was also suggeted by the experiments of Etzler and colleagues who found a lectin in the cell wall of leaves and stems of the legume Dolichos biflorus. The lectin had biochemical properties indicating a possible precursor relationship to the seed lectin (Talbot & Etzler, 1978). Its level in the plant cell walls increased during fungal infection or upon wounding of the plant, suggesting an antibody-like function (Etzler et al., 1979; Gibson & Etzler, 1979).

The paucity of data supporting possible roles of plant lectins, except for the legume-rhizobium recognition hypothesis, contrasts

with the extensive use of these lectins in cell biology and immunology.

Interest in the determination of structure of lectins, in order to better understand their interaction with cell surfaces, is growing. Goldstein & Hayes (1978) recently reviewed the biochemical characterization of the best known lectins. Although no generalization can be made about the structure of lectins, those derived from leguminous plants show certain common characteristics. There are two types of leguminous lectins, those containing a single type of subunit and those containing two types of subunits. The single chain lectins are usually homologous tetramers. A notorious exception is the lectin from the red kidney bean which is a tetramer of two different kinds of subunits, one mitogenic for lymphocyte, the other one with hemagglutination properties (Miller et al., 1975). The subunit of the single chain lectins has a molecular weight (MW) of approximately 26,000. The amino acid compositions of the lectins show a high content of acidic and hydroxyamino acids and a low content of sulfur-containing amino acids (Goldstein & Hayes, 1978). The two chain lectins have so far all been found in a single tribe of leguminous lectins, VICIEAE. They are composed of a small alpha-subunit, about 6,000 MW, and a large beta-subunit, 14,000 to 20,000 MW. They have the subunit structure alpha-2-beta-2 and

their amino acid composition is similar to that of single chain lectins.

Four leguminous lectins have had their amino acid sequence established: concanavalin A (ConA) from the jack bean (Wang et al., 1975; Cunningham et al., 1975); lentil lectin (LL) (Foriers et al., 1981); favin from the fava bean (Cunningham et al., 1979; Hopp et al., 1982) and soybean agglutinin (SBA) (Hemperly et al., 1981). Only one of them, ConA, had its crystalline tertiary structure established (Becker et al., 1975; Reeke et al., 1975). While ConA and SBA are tetramers of a unique polypeptide chain, LL and favin have the subunit structure alpha-2-beta-2. The four lectins were found homologous in their sequence of amino acids. The two chain lectins resemble SBA when the beta-chain and alpha-chain are aligned as though they were a single polypeptide chain beta-alpha (Hemperly et al., 1981). A circular permutation was observed between ConA and all the other lectins. Favin beta-alpha, LL beta-alpha and SBA are each homologous to a segment of ConA beginning at residue 120. The homology extends to the carboxyl terminus of ConA and continues, without interruption, through its amino-terminal 119 residues (Cunninghamm et al., 1979; Foriers et al., 1981; Hemperly et al., 1981).

Favin, LL and ConA belong to the D-mannose (D-glucose) binding family; SBA binds N-acetylgalactosamine (Golstein & Hayes, 1978). In spite of their differences in sugar specificity these lectins appear to have a common evolutionary origin.

Sainfoin lectin (SL) is extracted from the seeds of the leguminous forage crop sainfoin (Onobrychis viciifolia, Scop.). It is a dimer of two identical or very similar subunits of 26,000 MW (Hapner & Robbins, 1979). Its amino acid composition is typical of leguminous lectins. The lectin exhibits charge heterogeneity by electrophoretic criteria, interpreted as reflecting the possible presence of allelomorphic or isolectin forms (Hapner & Robbins, 1979). Each subunit bears a short carbohydrate chain linked to an asparagine residue (Namen & Hapner, 1979).

A lectin was isolated from sainfoin roots with similar molecular characteristics to the seed lectin (Hapner & Robbins, 1979) indicating that conclusions from studies performed on the seed lectin should apply to the root lectin. Localization of SL on the legume root by fluorescence microscopy was consistent with a putative role in rhizobium recognition (Hapner & Hapner, 1978). Sainfoin lectin belongs to D-mannose (D-glucose) binding family and it agglutinates cat erythrocytes specifically (Hapner & Robbins, 1979). The lectin is mitogenic for mice lymphocytes (Hapner,

unpublished observations). Because of its narrow specificity the lectin may be very useful as a probe for carbohydrate structure and for isolation of glycoproteins and glycolipids.

CHAPTER TWO

GOALS AND OBJECTIVES

The principal objectives of this research involve the biochemical and structural characterization of sainfoin lectin, a glycoprotein that is isolated from seeds of the leguminous plant Onobrychis viciifolia, Scop.. The sainfoin lectin molecule is known to be a 52,000 molecular weight dimer containing two identical polypeptide chains of 236 amino acid residues each. Specific research objectives are listed below.

1. Determination of the complete primary structure (amino acid sequence) of sainfoin lectin.

    a. This objective requires a collaborative arrangement with Dr. R. A. Bradshaw, Washington University, St. Louis, Missouri. The actual protein sequencing will be performed in Dr. Bradshaw's laboratory. Protein and peptide preparation will be done at Montana State University, Bozeman, Montana.

    b. The established sequence will be compared with that of other leguminous lectins which have been studied.

2. Elaboration of successful biochemical methods to produce peptide fragments from purified sainfoin lectin that are suitable for sequence analysis.

3. Determination of carbohydrate binding, metalloprotein and molecular stability characteristics of sainfoin lectin.

CHAPTER THREE

EXPERIMENTAL PROCEDURES

Materials

Seeds of sainfoin (Onobrychis viciifolia. Scop., variety Eski) were supplied by the Plant and Soil Science Department, Montana State University. Sephadex G-types resins, Sepharose 4B, Sephadex LH20, WGA-Sepharose, and ConA-Sepharose were obtained from Pharmacia. Biogel A5m was purchased from Biorad, DEAE cellulose from Whatman and Ultrogel AcA44 from LKB. DC-X8-11 was from Dionex Chemical. Ultrafiltration membranes and phenylboronate agarose PBA-30 were obtained from Amicon, equilibrium dialysis membranes from Hoeffer. Cat erythrocytes were purchased from Colorado Serum. Biofluor, D-[^{14}C(U)]-glucose, iodo[1-^{14}C]acetic acid and K^{14}CN were from New England Nuclear. Acrylamide, phenantrenequinone, iodoacetic acid, succinic anhydride, guanidine HCl, beta xylosidase, and thermolysin (protease type X) were obtained from Sigma. The iodoacetic acid was recrystallized from petroleum ether before use. Ninhydrin, methylcellosolve, dansylchloride, trinitrobenzene sulfonic acid, and sodium dodecylsulfate were from Pierce. Bisacrylamide, beta mercaptoethanol and 5,5'-dithiobis-(2-nitrobenzoic acid) were purchased from Aldrich. Cyanogen bromide,

pyridine, acetonitrile and HPLC water were obtained from Baker. Pyridine was distilled (1 g ninhydrin/liter) before use. Dithiothreitol was from Calbiochem, 2-nitro-5-thiocyanobenzoic acid from Eastman. Standard dansyl amino acids were purchased from Seikagaku Kogyo Company. Polyamide sheets for thin layer chromatography of the dansyl amino acids were purchased from Gallard Schlesinger Chemical Company. Trypsin treated with N-tosylphenylalanine chloromethyl ketone was obtained from Millipore Company. Carboxypeptidase B was from Worthington. Sequence grade phenylisothiocyanate, heptane, heptafluorobutyric acid, benzene and ethyl acetate were purchased from Beckman. Polybrene was a gift from Abbott Laboratories. Silica thin layer plates with fluorescent indicator were from Analtech. Noble agar was purchased from Difco. All other chemicals were of reagent grade. Distilled water was used throughout.

Purification of Sainfoin Lectin

Sainfoin lectin was purified from the seeds of sainfoin (Onobrychis viciifolia, Scop.), variety Eski, by a modification of the method of Hapner & Robbins (1979). Dehulled and finely ground (200 g) seeds were stirred overnight with one liter PBS (0.01 M sodium phosphate, 0.15 M sodium chloride) containing 0.1 M glucose, 0.01 M ascorbic acid, 3 M sodium azide, pH 7.0. The insoluble

material was removed by squeezing through cheesecloth and subsequent centrifugation at 10,000 X g for 20 min. The soluble extract was dialyzed 48 h against 10 liters PBS with two changes of buffer. Precipitate from the dialyzate was removed by centrifugation at 13,000 X g for 30 min.

The supernatant was submitted to affinity chromatography in a batchwise fashion. Mannose-Sepharose beads (50 ml) prepared as previously described (Fornsted & Porath, 1975) were added to the supernatant solution and the mixture was stirred 3 to 6 h. The beads were collected on a Buchner, washed extensively with PBS and packed into a 2.5 X 10 cm column. The lectin was eluted with 0.5 M glucose in PBS. The lectin was dialyzed extensively against PBS and repurified by affinity chromatography on a 2.5 X 5 cm column of fresh mannose-Sepharose.

Gel Filtration of SL

Sainfoin lectin (1 mg/ml) was chromatographed on a 1.5 X 115 cm Ultrogel AcA44 column. Elution was performed with three different buffers: PBS, pH 7.0; 0.01 M sodium acetate, 0.15 M sodium chloride, pH 4.5; 0.01 M sodium acetate, 0.15 M sodium chloride, pH 4.0 (with and without 0.1 M glucose). Flow rate was 22 ml/h and absorbance at 280 nm was recorded. Standards used to calibrate the

column were bovine serum albumin (MW 68,000), ovalbumin (MW 43,000), pepsin (MW 35,000) and chymotrypsinogen (MW 25,700).

Affinity Chromatography of SL at Various pH Values

Sainfoin lectin (1 mg/ml) was chromatographed on a mannose-Sepharose column 0.9 X 12 cm. Three buffers were used for the elution: PBS, pH 7.0; 0.01 M sodium acetate, 0.15 M sodium chloride, pH 4.5; 0.01 M sodium acetate, 0.15 M sodium chloride, pH 4.0. After washing with 20 ml buffer, elution was performed at 23 ml/h with a linear glucose gradient, 0-0.1 M glucose in 100 ml. The effluent was monitored at 280 nm and 1 ml fractions were collected.

Anion Exchange Chromatography of SL

Sainfoin lectin, 32 mg in 4 ml 5 mM sodium phosphate, 0.1 M sucrose, pH 8.3 was chromatographed on a diethylaminoethyl (DEAE) cellulose column 2.6 X 95 cm at 22 ml/h. Elution was performed with 0.01 M sodium phosphate, 0.1 M sucrose, 3 mM sodium azide, pH 7.9.

Fractions (2 ml) were collected and their absorbance at 280 nm was measured on a spectrophotometer. Fractions that eluted at the column void volume were pooled, concentrated to 8 mg/ml by ultrafiltration over a UM 10 membrane, and adjusted to pH 8.5. The concentrate was chromatographed again on the DEAE cellulose column

in 0.01 M Tris, 0.1 M sucrose, 3 mM sodium azide, pH 8.5, at 14 ml/h; 1 ml fractions were obtained.

Immuno Double Diffusion Agar Gels

Rabbit antiserum to sainfoin lectin previously produced (Hapner & Hapner, 1978) was utilized. Double diffusion tests were done on 5 cm 0.8% Noble agar plates in 0.05 M sodium barbital, pH 8.4 according to Ouchterlony (1948). Antiserum (50 µl) was placed in the center well and SL fractions (25 µl) were placed in the outer wells. The plates were incubated at 22°C for 24 h.

Metal Analysis

Sainfoin lectin was extensively dialyzed against 0.01 M glucose in metal free water. It was then prepared for metal analysis by one of two procedures:

1. 10 mg lectin in 10 ml solution was made 1 mM in ethylenediamine tetraacetate (EDTA), adjusted to pH 7.0 and incubated 15 min at 80°C followed by centrifugation. The supernatant was evaporated to dryness, 1 N nitric acid was added and the sample was aspirated into the atomic absorbtion instrument.
2. 20 mg water-dialyzed SL was lyophilized and dissolved in 5 ml concentrated nitric acid with heating. Heating was

continued at a temperature just below the boiling point of nitric acid until 1 ml nitric acid was left. After cooling, 2.5 ml of 20,000 ppm potassium nitrate was added and the final volume adjusted to 25 ml with metal free water. The solution was aspirated into the atomic absorbtion instrument.

In both procedures protein-free blanks were prepared and analyzed. The metals measured were Ca^{++}, Mn^{++} and Mg^{++}.

Equilibrium Dialysis

Equilibrium dialysis was performed in 0.5 ml dialysis cells in PBS that contained 3 mM sodium azide, pH 7.0. Dialysis membranes were washed exhaustively with distilled water before use. Solutions of ^{14}C-glucose (0.1-10 mM, 2.6-8.7 μCi) and protein (1.58 mg/ml), 0.5 ml each, were placed in adjacent compartments of the dialysis cells. The unit, containing eight cells, was rotated for 72 h at room temperature to establish equilibrium. Two 50 μl samples were then removed from each compartment and mixed with 4 ml scintillation mixture (Biofluor). The radioactivity in each sample was determined in a Tricarb model 3320 (Packard) liquid scintillation counter.

Agglutination Assay

Agglutination tests were performed in PBS in Microtiter plates by adding 25 µl 2% cat erythrocytes in PBS, pH 7.0 to serial twofold dilutions of lection (25 µl). When inhibition by sugars was examined, carbohydrate was included in the diluent PBS. Agglutination was evaluated visually after 30 min to 1 h at 22°C.

Standard Electrophoretic Procedures

Sodium dodecyl sulfate (SDS) electrophoresis was performed in 10% polyacrylamide gels using a phosphate buffer system according to Weber and Osborn (1975). The gels were stained with 0.25% Coomassie blue R-250 in destaining solution (methanol:acetic acid:water, 227:46:227). Protein and peptide samples, 10-50 µg in sample buffer containing 1% SDS and 1% beta-mercaptoethanol, were heated 5 min on a boiling water bath prior to electrophoresis. Mobilities were calculated relative to bromophenol blue tracking dye.

Disc electrophoresis in 7% polyacrylamide gels was performed at pH 9.5 in Tris, glycine buffer and at pH 3.8 in beta-alanine, acetic acid buffer using procedures and chemicals provided by Canalco. Gels were stained with Coomassie blue R-250 or, alternatively, with 0.04% Coomassie brilliant blue G-250 in perchloric acid:water, 3.5:96.5 according to Reisner et al. (1975).

Gels for isoelectric focusing each contained protein sample (10-50 µl) in 1.5 ml water, 0.56 ml of 30% (w/v) acrylamide, 1% (w/v) bisacrylamide in water, 0.056 ml of 40% (w/v) Ampholine Carrier ampholytes, pH 3-10 range (LKB), and 0.13 ml potassium persulfate 1% (w/v) in water. After the gels were polymerized they were focused at 1 mA/tube for 3 to 5 h. Tank electrode solutions were 0.2% sulfuric acid on the anodic side and 0.4% ethylenediamine on the cathodic side. After completion of focusing, the ampholytes were allowed to diffuse out of the gels by soaking the gels 30 min in a solution consisting of methanol:water, 30:70 containing 3.45% (w/v) sulfosalicylic acid and 11.5% (w/v) trichloracetic acid. The gels were then stained in Coomassie blue R-250 and destained in ethanol:acetic acid:water, 250:80:670.

Synthesis of Allyl Alpha-Glycosides

Allyl alpha-glucoside and allyl alpha-galactoside were synthesized by the method of Horejsi (1974): 100 g D-glucose or D-galactose was suspended in 200 ml allyl alcohol containing 6 g dry HCl. The mixture was stirred 4 h at 80°C. After cooling, the solution was neutralized with concentrated ammonia and the allyl alcohol was removed by rotary evaporation at 70°C. The resulting brown syrup was extracted eight times with 100 ml dry acetone. The extracts were combined and evaporated. The solid residue was

dissolved in hot ethanol. In the case of allyl alpha-galactoside, it crystallized readily upon cooling in ice. The ethanol solution containing allyl alpha-glucoside was treated with activated vegetable carbon and concentrated. Anhydrous ether was added to cloudiness and the solution was incubated at -10°C where it crystallized. The product was recrystallized twice from ethanol:ether, 70:30. Both allyl alpha-glycosides showed a single spot of Rf 0.37 on TLC in butanone:ethanol:water, 75:8:3. The alpha nature of the linkage was confirmed by carbon 13NMR through comparison of the C-1 chemical shift with standard compounds.

Synthesis of Glycosyl Copolymers

Glucosyl and galactosyl copolymers were prepared by the method of Horejsi (1977): 500 mg allyl alpha-glycoside, 1 g acrylamide and 20 mg ammonium persulfate were dissolved in 10 ml water. The solution was boiled on an oil bath for 5 to 9 min. It was then diluted with 20 ml water and extensively dialyzed against water. The carbohydrate content of the glycosyl copolymer was determined spectrophotometrically (Dubois, see carbohydrate analysis) with reference to the corresponding free sugar.

Affinity Electrophoresis

Solutions used for the polymerization of affinity gels at pH 4.5 were as follows:

A: 60 ml 1 N potassium hydroxide, 21.5 ml acetic acid, 5 ml N,N,N'N'-tetramethyl ethylene diamine, 13.5 ml water, pH 4.5.

B: 37.5% (w/v) acrylamide, 1% (w/v) bisacrylamide in water.

C: 0.35% (w/v) ammonium persulfate in water.

The polyacrylamide affinity gels (7%) were made by mixing 1 ml A with 2 ml B and 4 ml C; glucosyl copolymer and solid free glucose were added to the mixture in amounts needed for the desired final concentrations of immobilized and free carbohydrate; the mixture was then adjusted to 10 ml total volume with water. Sample size was 20 µl protein applied to each gel. Disc electrophoresis was performed using the beta alanine: acetic acid buffer system of Canalco, at 5 mA/gel, 10°C for 3½ h. Gels were stained with Coomassie blue R-250 and destained in methanol:acetic acid:water, 227:46:227.

Two affinity gels were electrophoresed for each combination of immobilized glucose and free glucose concentrations. Control affinity gels were polymerized with the amount of galactosyl copolymer required to yield an immobilized galactose concentration

equivalent to the immobilized glucose concentration of the experimental affinity gels.

Solutions used for the polymerization of affinity gels at pH 9.0 were as follows:

D: 60 ml 1 N HCl, 45.25 g Tris base, 0.3 ml N,N,N',N'-tetramethyl ethylene diamine, H_2O up to 200 ml, pH 9.0.

E: 35% (w/v) acrylamide, 0.92% (w/v) bisacrylamide in water.

C: as for pH 4.5 affinity gels.

Polyacrylamide affinity gels (7%) were made by mixing 2 ml D with 2 ml E and 2 ml C. Glucosyl copolymer and free glucose were added to the mixture in an analogous fashion to the pH 4.0 affinity gels. Disc electrophoresis was performed using the Tris:glycine buffer system of Canalco, at 2 mA/gel for 6½ h. Other procedures were as described for affinity electrophoresis at pH 4.0.

Performic Acid Oxidation

The single cysteine residue of SL (Hapner & Robbins, 1979) was oxidized to cysteic acid with performic acid by the method of Hirs (1967). Performic acid reagent was prepared by mixing 1 ml 30% hydrogen peroxide with 19 ml 99% formic acid and allowing the mixture to stand 2 h at 25°C. One mg protein was dissolved in 1 ml formic acid and incubated at -5°C. Methanol (0.2 ml) and 0.5 ml performic acid were added to the protein solution. The reaction

mixture was incubated at -5°C for 2½ h. It was then diluted with 25 ml water and lyophilized twice. Cysteic acid was measured by amino acid analysis after acid hydrolysis of the modified protein.

Carboxymethylation

Sainfoin lectin was carboxymethylated according to Thomas et al. (1981). Briefly, the lyophilized protein was dissolved in 6 M guanidinium chloride, 0.1 M Tris base, 1 mM bis(aminoethyl)-glycolether-N,N,N',N'-tetraacetic acid (metal chelator), pH 9.5. It was reduced by addition of tenfold protein molar excess dithiothreitol, in the dark, under a N_2 atmosphere, for 2-4 h. The protein was then carboxymethylated at pH 8.0 with threefold molar excess (over total sulfhydryl groups) iodo[1-^{14}C]acetic acid (specific activity 7.47 X 10^8 cpm/mmole). The reaction was stopped after 5 min at 25°C with excess beta-mercaptoethanol. Reagents were removed by dialysis in water and the carboxymethylated protein was lyophilized.

Reaction with 5,5'-Dithiobis-(2-nitrobenzoic acid)

Native or succinylated (see below) protein at 1 mg/ml in 0.01 M Tris, 1 mM EDTA, 0.1% (w/v) octylglucoside was incubated at pH 2 for 30 min. The pH was then adjusted to 8.0, fivefold protein molar excess DTNB was added, and the solution was incubated at 50°C for 20

min. Extent of the reaction was estimated by measuring the liberation of 2-nitro-5-thiobenzoate using the molar extinction coefficient of 13,600 M^{-1} at 412 nm (Ellman, 1959).

Succinylation

Sainfoin lectin was succinylated by a procedure derived from Klotz (1967). The protein (5 mg) was dissolved in 5 ml 0.01 M Tris, 0.1 M NaCl, 0.1 M glucose, pH 8.0. Fiftyfold molar excess (over total protein amino groups) solid succinic anhydride was added in five increments at 10 min intervals. The pH was maintained at 8.0 with 0.1 N NaOH. Excess reagent was removed by extensive dialysis against water. When carboxymethylated protein was succinylated, the succinic anhydride was added to the reaction mixture 10 min after the quenching of the carboxymethylation reaction with beta-mercaptoethanol.

Trinitrobenzene Sulfonic Acid (TNBS) Assay

Free amino groups were estimated by colorimetric titration with TNBS according to Means & Feeney (1971). To 1 mg protein in 1 ml PBS were added 1 ml 4% (w/v) sodium bicarbonate, pH 8.5 and 1 ml TNBS 1% (w/v) in water (freshly prepared). The reaction mixture was incubated at 40°C for 2 h followed by the addition of 1 ml 10% (w/v) SDS and 0.5 ml 1 M HCl. Absorbance at 335 nm was measured against a

protein-free blank. The number of amino groups was calculated using the molar extinction coefficient 14,000 M^{-1} cm^{-1}.

Cyanylation

Five mg native or succinylated protein in 5 ml 6 M guanidine HCl, 0.01 M Tris, 1 mM EDTA, pH 9.0, was cyanylated at 25°C with fivefold protein molar excess NTCB or ^{14}C-NTCB (see below for synthesis). The reaction was measured by the liberation of 2-nitro-5-thiobenzoate, using the molar extinction coefficient of Ellman (1959). When the reaction was complete, the protein solution was acidified to pH 2.0 and filtered through a Sephadex G10 column, 1.5 X 43 cm, in 0.02 M sodium chloride. Elution was monitored by the 280 nm absorbance. The desalted protein was concentrated to 2 ml by ultrafiltration over a UM 10 membrane.

Synthesis of 2-Nitro-5-thio[^{14}C]cyano Benzoic Acid (NTCB)

The hemipotassium salt of [^{14}C]NTCB was synthesized according to Degani & Patchornik (1971): 250 mg potassium bicarbonate and 70 mg potassium cyanide were dissolved in 5 ml water followed by the addition of 0.25 mCi K^{14}CN. The mixture was stirred and 100 mg DTNB was added. It was incubated 30 min at 25°C. A 3% (w/v) cyanogen bromide solution in water was slowly added to the reaction mixture until no more color was discharged. The pH was then decreased to 5

with acetic acid:water, 50:50. Nitrogen was bubbled through the solution for 4 h. The solution was then acidified with 6 N HCl until white crystals appeared. After cooling in ice, the crystals were collected and dried in a dessicator. The product was recrystallized from 95% ethanol. The yield of synthesis was 47%. The synthesized [^{14}C]-NTCB had a melting point of 237°C, a specific activity of 910,295 cpm/mg , showed a single spot of Rf 0.70 on TLC in butanol:acetic acid:water, 25:6:25, and exhibited two UV absorbtion maxima at 291 nm and 202 nm which corresponded with the commercial unlabeled reagent.

Cleavage at Cysteine

Cyanylated SL was cleaved at the cyanylated cysteine according to Jacobson et al. (1973). Five mg derivatized protein in 5 ml 0.01 M sodium borate, 1 mM EDTA, 6 M guanidinium chloride, pH 9.0, was incubated at 37°C for 40 h. The reaction was quenched by acidifying to pH 6 with acetic acid.

Gel Filtration of SL Cysteine Cleavage Product

Five mg peptides derived from the cleavage of cyanylated SL were chromatographed on a Sephadex G50 medium column, 1.5 X 117 cm, at 28 ml/h. The elution buffer was 0.01 M sodium phosphate, 0.1%

SDS, pH 7.0. Fractions (1.1 ml) were collected and their absorbance at 230 nm was measured.

Gel Filtration of Succinylated SL Cleaved at Cysteine

Succinylated protein cyanylated with ^{14}C-NTCB and cleaved at cysteine was gel filtered through a Sephadex G25 fine column, 1.5 X 115 cm. The column was eluted with 0.02 M ammonium bicarbonate, pH 6.3 in one case and 3 M guanidinium chloride, 0.01 M sodium phosphate, pH 7.0 in another case. Flow rate was 29 ml/h.

Fractions of 1.8 ml were collected and their absorbance at 230 nm was measured. The fractions collected from the chromatography in 3 M guanidinium chloride were read at 280 nm because of the high UV absorbing background of the solvent at 230 nm. Aliquots of the fractions (0.4 ml) were mixed with 4 ml scintillation mixture (Biofluor) and counted for 10 min in a Tricarb liquid scintillation counter.

Partial Acid Hydrolysis

Acid sensitive peptide bonds of SL were cleaved as described by (Fraser et al., 1972). Protein (20 mg) in 4 ml 6 M guanidinium chloride, 10% (v/v) acetic acid, pH 2.5, was incubated at 40°C for 120 h. The digest was then adjusted to pH 4.5 and submitted to gel filtration.

The partial acid hydrolyzate was fractionated on a Biogel A5m column, 1.5 X 110 cm, at 5.4 ml/h. The elution buffer was 6 M guanidinium chloride, 0.1 M sodium acetate, pH 4.5. One ml fractions were collected and their absorbance at 230 and 280 nm was recorded.

Digestion with Trypsin

The method used was as follows, unless otherwise indicated: 2.8 μmoles (73 mg) carboxymethylated succinylated SL in 35 ml 1 mM Tris, 1 mM EDTA, pH 8.0 was incubated at 50°C for 30 min. Trypsin (1 mg/ml) was dissolved in 1 mM HCl, 1 mM calcium chloride and kept at 4°C prior to its use. Trypsin (1%, w/w) was added to the protein substrate and the pH was maintained at 8.0 with 0.1 N NaOH. When base uptake ceased, the solution was acidified to pH 2.0 with HCl and incubated at 50°C for 1 h. The pH of the solution was then readjusted to 8.0 and a second trypsin addition was made. The digestion was continued until base uptake stopped. The tryptic digest was held at 4°C until fractionation at pH 3.5.

Digestion with Thermolysin

One μmole carboxymethylated succinylated SL was dissolved in 3 ml 0.12 M ammonium bicarbonate, 1 mM calcium chloride, pH 8.0. Thermolysin (1% w/w) was added to the protein substrate. The

reaction mixture was incubated at 25°C for 1 h and submitted to gel filtration on Sephadex G25 immediately.

Digestion with Carboxypeptidases

Digestion with carboxypeptidase B was as follows. An aliquot containing 20 nmoles trypsin-treated SL in 200 µl 0.2 M Tris, pH 8.0 was incubated with 10 µg carboxypeptidase B (Ambler, 1972) at 37°C for 2 h. The reaction mixture was acidified to pH 2.5 with acetic acid. One ml of 0.2 M sodium citrate, pH 2.2, was added and the solution was centrifuged to remove any precipitate. The supernatant solution was analyzed for its basic amino acid content with the amino acid analyzer. An enzyme blank was prepared and analyzed similarly.

Digestion of SL tryptic peptides with carboxypeptidase Y was performed according to Hayashi (1977).

Fractionation of SL Tryptic Digest at Low pH

In order to choose a pH for the fractionation of the tryptic digest, aliquots of tryptic digest were adjusted to different pH values and centrifuged. Absorbance at 280 nm of the supernatants were recorded.

Trypsinized SL was routinely adjusted to pH 3.5 and incubated at 0°C for 30 min. It was then centrifuged and precipitated and

supernatant solutions were separated. The soluble peptides were adjusted to pH 7.0 and gel filtered on Sephadex G25 or chromatographed on a cation exchanger. The insoluble peptides were dissolved in 0.02 M ammonium bicarbonate, pH 7.9, and submitted to gel filtration on Sephadex G50.

Gel Filtration of the pH 3.5 Soluble SL Tryptic Peptides

One μmole of pH 3.5 soluble tryptic peptides was concentrated to 4 ml by rotary evaporation and chromatographed on a Sephadex G25 fine column, 2.5 X 120 cm, at 35 ml/h. Elution buffer was 0.02 M ammonium bicarbonate adjusted to pH 6.3 with carbon dioxide or, alternatively, 1 mM acetic acid. Fractions (2.5 ml) were collected and their absorbance at 230 nm was measured. When the peptides were derived from the digestion with trypsin of SL previously carboxymethylated with iodo-[^{14}C] acetic acid, 100 μl aliquots of the gel filtration column were mixed with 5 ml scintillation mixture and counted for 10 min in a Tricarb liquid scintillation counter.

Ultraviolet absorbtion spectra of the peptides were obtained on a Varian Techtron 635 recording spectrophotometer.

Repurification of T2 and T10 on LH20

The fraction pool from Sephadex G25 chromatography, containing the peptide of interest, was concentrated to 1 ml by rotary

evaporation and chromatographed on a Sephadex LH20 column, 0.9 X 95 cm, at 4.5 ml/h. The column was eluted with water. Fractions (1 ml) were collected and their absorbance at 230 nm was recorded. In the case of peptide T2, 10 µl aliquots of the fractions were counted 5 min for radioactivity.

Cation Exchange Chromatography of the pH 3.5 Soluble SL Tryptic Peptides

One µmole of pH 3.5 soluble tryptic peptides were lyophilized, dissolved in 1.5 ml 30% (v/v) acetic acid, 0.4% (v/v) pyridine (in water), pH 2.5, and chromatographed on a DC-X8-11 column, 0.9 X 20 cm, at 30 ml/h according to the method of Bradshaw et al. (1980). After loading the sample, elution was performed with sample buffer for 5 min. A four chamber buffer gradient was then started which consisted of: a) 200 ml sample buffer; b) 200 ml 27.8% (v/v) acetic acid, 1.61 % pyridine, pH 3.1; c) 200 ml 20% (v/v) acetic acid, 4.03% (v/v) pyridine, pH 3.7; d) 200 ml 14.3% (v/v) acetic acid, 16.1% (v/v) pyridine, pH 5.0. The chromatography was monitored with ninhydrin with an automated peptide analyzer. Three ml fractions were collected.

Gel Filtration of the pH 3.5 Insoluble SL Tryptic Peptides

One μmole of pH 3.5 insoluble tryptic peptides were dissolved in 3 ml 0.02 M ammonium bicarbonate, pH 7.9, and chromatographed on a Sephadex G50 superfine column, 2.5 X 120 cm, at 16 ml/h. Elution buffer was 0.02 M ammonium bicarbonate adjusted to pH 6.8 with CO_2. Three ml fractions were collected and their absorbance at 230 nm was recorded.

Repurification of T4 and T7 by High Performance Liquid Chromatography (HPLC)

The Sephadex G50 chromatographic pools containing the peptides to be repurified were concentrated to 2 ml by ultrafiltration over a UM 02 membrane and aliquots (200 μl) were chromatographed on a Beckman Ultrasphere ODS HPLC column preequilibrated with 0.01 M ammonium bicarbonate, pH 7.0. Elution was performed at 0.5 ml/min with the following successive buffers: 0.01 M ammonium bicarbonate for 3 min; 20% acetonitrile in 0.01 M ammonium bicarbonate for 3 min; a linear gradient of 20% to 35% acetonitrile in 0.01 M ammonium bicarbonate for 30 min; 35% acetonitrile in 0.01 M ammonium bicarbonate for 5 min. Absorbance at 280 nm was monitored with a Perkin Elmer UV monitor, model LC75. Peaks of 280 nm absorbance were collected manually, lyophilized twice, and analyzed for amino acid content.

Digestion of T6 with Beta Xylosidase

T6 (220 nmoles) in 1 ml 0.2 M sodium acetate, 0.1 M sodium chloride, pH 5.5, was incubated with 0.5 units beta xylosidase at 37°C for 4 h. The reaction was quenched by adjusting the pH to 7.2. The digest was chromatographed on a ConA-Sepharose column, 1.5 X 2.5 cm, in 0.02 M Tris, 0.1 M sodium chloride, 1 mM calcium chloride, pH 7.2, at 2 ml/h. The elution was monitored by the absorbance at 230 nm.

Boron Chromatography of T6

T6 (200 nmoles) was chromatographed on a phenyl boronate agarose PBA-30 column, 1.5 X 1.5 cm, in 0.05 M ammonium carbonate, pH 9.1, at 2 ml/h. Absorbance at 230 nm was recorded.

Chromatography of T6 on WGA-Sepharose

T6 (200 nmoles) was chromatographed on a wheat germ agglutinin (WGA)-Sepharose column, 1.5 X 1.5 cm, in 0.02 M Tris, 0.1 M sodium chloride, 1 mM calcium chloride, 1 mM manganese chloride, pH 7.0, at 2 ml/h. The elution was monitored through the absorbance at 230 nm.

Carbohydrate Analysis

Neutral carbohydrates were measured colorimetrically by the method of Dubois et al. (1956) as modified by Misaki & Goldstein (1977). One ml of an aqueous solution of glycopeptide was mixed

with 0.5 ml 5% (v/v) phenol in water and 3 ml concentrated sulfuric acid. The solution was incubated at 25°C for 1 h and its absorbance at 490 nm was measured. D-Glucose was used to prepare the standard curves.

Glucosamine was identified on a Beckman 120 C amino acid analyzer after hydrolysis of the glycopeptide.

Gel Filtration of the Thermolytic Peptides

The thermolytic digest of SL was chromatographed on a Sephadex G25 fine column, 2.5 X 120 cm, at 29 ml/h. The elution buffer was 0.02 M ammonium bicarbonate, pH 6.4. Fractions (2.8 ml) were collected and their absorbance at 230 nm was recorded. The fractions were pooled appropriately and lyophilized twice.

Cation Exchange Chromatography of the Thermolytic Peptides

The thermolytic peptide fractions were dissolved in 1 ml acetic acid:water, 50:50, and chromatographed on the Beckman AA-15 column, 0.9 X 23 cm, of the amino acid analyzer. Gradient elution was performed at 40 ml/h in exactly the same fashion as the one described in "Cation Exchange Chromatography of the pH 3.5 Soluble Tryptic Peptides." At the end of the gradient, the column was eluted with the final buffer for 1.5 h followed by 6 M pyridine for 0.5 h. Aliquots (0.3 ml) of every third fraction (3.5 ml) were

lyophilized and submitted to an arginine fluorescent spot test. Aliquots of the arginine positive fractions were hydrolyzed in base and analyzed with ninhydrin in order to describe completely the peptide elution pattern.

Arginine Detection

Arginine was detected by the method of Yamada & Itano (1966). The lyophilized fractions to be analyzed were dissolved in 30 µl acetone:water, 50:50, and spotted on 3 MM Whatman paper, 10 µl at a time. The paper was dipped in a freshly made mixture of equal volumes of phenantrenequinone 0.02% (w/v) in ethanol and 10% (w/v) sodium hydroxide in 60% ethanol. It was air dried for 30 min and arginine fluorescence was visualized under a UV lamp.

Ninhydrin Analysis

Fractions to be analyzed were first submitted to alkaline hydrolysis according to Hirs (1967): the fractions were dried in an oven at 110°C, 0.15 ml 13.5 N NaOH was added, and the samples were autoclaved at 121°C for 20 min. Ninhydrin analysis was performed according to Cocking and Yemm (1954): to each fraction were added 0.25 ml acetic acid, 0.5 ml 0.2 M sodium citrate, pH 5.0, 0.2 ml ninhydrin 5% (w/v) in methylcellosolve, and 1 ml 0.01 M sodium cyanide 2% (v/v) in methylcellosolve. The solutions were heated 15

minutes on a boiling water bath. Absorbances (570 nm) were measured after the samples had cooled.

Amino Acid Analysis

Peptides were hydrolyzed with 1 ml 6 N HCl in evacuated glass tubes at 110°C for 18 h according to Spackman et al. (1958). Amino acid analyses were performed on a Beckman 120C or a Durrum D-500 analyzer.

End Group Analysis

Two methods were used: a) 100 nmoles peptide in 0.2 M sodium bicarbonate, pH 8.4, was added forty-twofold molar excess iodoacetic acid over total amino groups, readjusted to pH 8.4 and incubated at 50°C for 19 h. The alkylated sample was evaporated to dryness, hydrolyzed with 2 ml 6 N HCl and analyzed with the amino acid analyzer. The results were compared to those obtained with another sample of the peptide treated in the same fashion but without addition of alkylating agent.

b) Alternatively, the peptide was analyzed with the dansyl chloride method (Gray, 1972): 20 nmoles peptide was dissolved in 50 µl 0.2 M sodium bicarbonate in ammonia free water, 50 µl dansyl (dimethyl-amino-napthalenesulfonyl) chloride, (5 mg/ml in acetone was added and the solution incubated at 37°C for 3 h. The

solution was evaporated to dryness with a stream of filtered air, 150 µl 6 N HCl was added and the sample was hydrolyzed for 16 h at 110°C. It was then dried under vacuum, 25 µl 95% (v/v) ethanol was added, followed by centrifugation. Four µl of the supernatant was spotted in the corner of a 5 X 5 cm polyamide sheet. The sheet was developed with formic acid 1.5% (v/v) in water in the first dimension and with toluene:acetic acid, 9:1:, followed by ethylacetate:methanol:acetic acid, 20:1:1, in the second dimension. The fluorescent dansylated amino acid derivatives were revealed under a UV lamp. Amino acid identifications were done by comparison with standard dansyl amino acids.

Sequence Analysis

All polypeptide chains were sequenced automatically in a Beckman 890 C Sequencer. The analyses were performed according to Edman & Begg (1967) using a 0.33 M Quadrol buffer degradation program (Thomas et al., 1981) modified from that of Brauer et al. (1975). One degradation cycle is briefly described as follows. The sample in the spinning cup was dissolved in 0.33 M Quadrol [N,N,N',N' tetrakis-(2 hydroxypropyl)-ethylene diamine]; benzene precipitated the peptide and carried off the excess phenylisothiocyanate; ethylacetate washed out the Quadrol and the

breakdown products of phenylisothiocyanate; anhydrous heptafluorobutyric acid cleaved off the amino terminal amino acid as the anilino thiazolinone derivative of the amino acid; butylchloride extracted the cleaved product that was transferred to a fraction collector. About 3 mg polybrene was included with the peptide to diminish its washout from the spinning cup.

The anilinothiazolinones from each cycle were converted to the corresponding phenylthiohydantoins in the following way: the samples in butylchloride were dried at 50°C with a nitrogen stream, 200 µl 1 N HCl with 1 µl/ml beta mercaptoethanol was added and conversion was done at 80°C for 10 min.

The phenylthiohydantoins were extracted twice with 750 µl ethyl acetate. The aqueous phase was saved. The extracts were combined, dried under nitrogen, and dissolved in 50 µl ethylacetate. Most residues were identified by a combination of one-dimensional thin layer (Jacobs & Niall, 1975) and gas chromatography (Pisano & Bronzert, 1969). A portion (20-30%) of the ethylacetate phase was spotted in the corner of a 20 X 20 cm silica plate. The plate was developed with methylene chloride:acetic acid, 30:7, dried and visualized under a UV lamp. Derivatives were identified by comparison with standards run in parallel with the unknown on the thin layer plate.

The phenylthiohydantoin of S-[^{14}C]-carboxymethylcysteine was identified by either liquid scintillation counting or autoradiography of the thin layer plate, or both. The phenylthiohydantoins of histidine and arginine were identified by spot tests of the aqueous phases (Easley, 1965; Yamada & Itano, 1966).

In the later stages of the work, phenylthiohydantoins were identified by high performance liquid chromatography using a 4.6 mm X 25 cm Altex Ultrasphere ODS 5 μm column, 15 mM sodium acetate (pH 5.5) as the aqueous phase, methanol:acetonitrile (17:3) as the mobile phase and detection at 269 nm (Gordon et al., 1982).

CHAPTER FOUR

RESULTS

Section 1. Sainfoin Lectin Stability

Gel Filtration of SL

A column of Ultrogel AcA44, 1.5 X 115 cm, was calibrated with standard proteins (Figure 1). Regression analysis of the plot log MW vs. VE yielded a correlation coefficient of -0.98. Sainfoin lectin was gel filtered on this column at several pH values. In all cases the lectin eluted as a single peak. The elution volume of the peak was correlated to a molecular weight with the plot of Figure 1. Table 1 shows the apparent molecular weight and the recovery of SL when it was gel filtered at pH 7.0, 4.5, 4.0 and 3.0. The molecular weight of the protein decreased from 51,700 at pH 7.0 to 32,400 at pH 4.0. Recovery of the protein decreased when the pH of the elution buffer was decreased. Recovery at pH 4.0 depended upon the presence of interacting carbohydrate. The lectin was completely lost at pH 4.0 in absence of glucose and partially recovered in presence of 0.1 M glucose. However, it was lost at pH 3.0 independent of the presence of glucose.

The elution volume of gel filtered SL depended upon the proportion of monomer and dimer species present under the

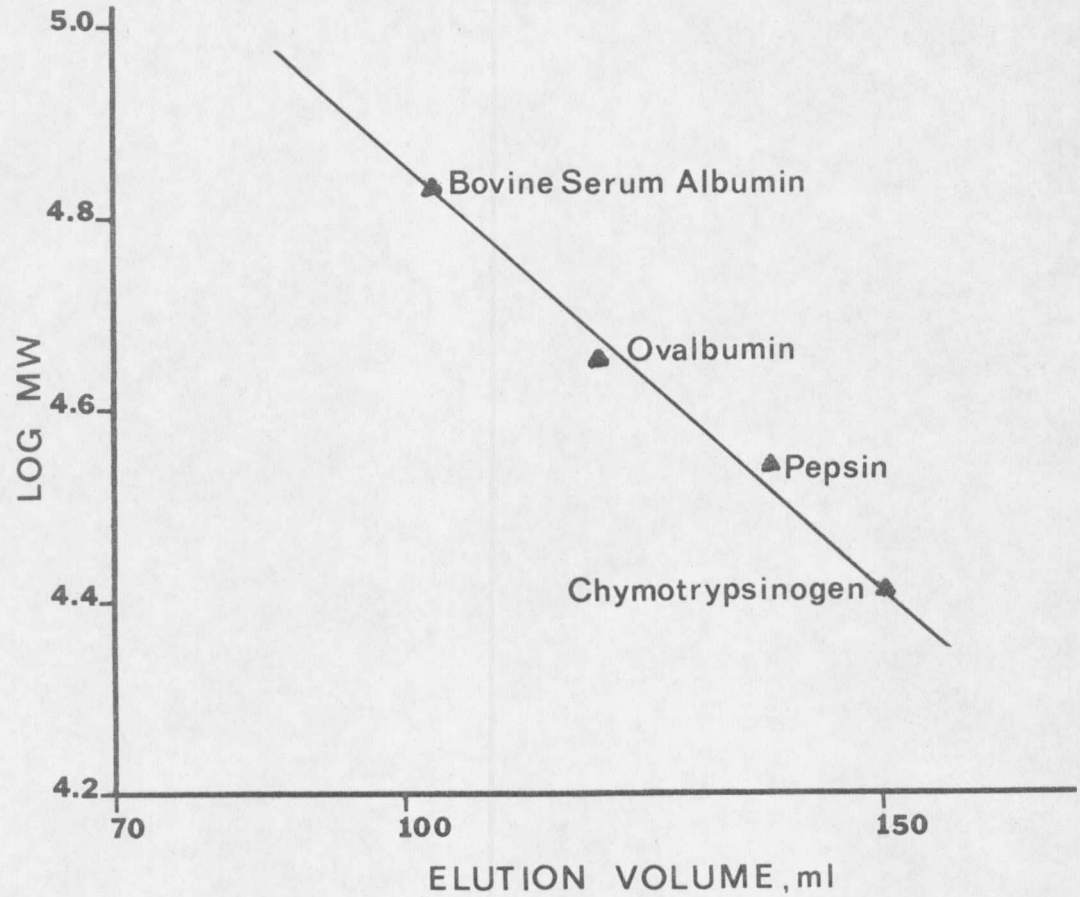


Figure 1. Calibration of Ultrogel AcA44 column, 1.5 X 115 cm

Table 1. Behavior of SL on Ultrogel AcA44

pH		elution volume	apparent MW	recovery
7.0		115 ml	51,678	98%
4.5		135 ml	34,993	76%
4.0	a.	∞ in absence of glucose		0%
	b.	139 ml in presence of 0.1 M glucose	32,368	54%
3.0		∞ in presence of glucose		0%

chromatographic conditions. Hapner & Robbins (1979) previously demonstrated SL monomerization at pH 4.0. In the experiments reported here the lectin exhibited an apparent molecular weight of 32,400 at pH 4.0, as opposed to 26,000 for the monomer species. The lectin was presumably incompletely monomerized at pH 4.0. The discrepancy in the results may have been due to the different gel filtration media employed in the two studies. Quaternary changes associated with variations in pH are reported in the literature for concanavalin A (ConA) (McKenzie et al., 1972), wheat germ agglutinin (WGA) (Nagata & Burger, 1974), peanut agglutinin (PNA) (Fish et al., 1979) and mung bean lectin (del Campillo et al., 1981). It is possible that these lectins, including SL, may undergo changes of

quaternary structure *in vivo*. These changes would be accompanied by alterations in valency that could result in alterations in activity and could be of importance to the physiological role of the lectins. The complete loss of SL during the gel filtration at low pH observed in the experiments reported here was interpreted as denaturation of the protein. The denatured protein presumably precipitated in the Ultrogel matrix or eluted in such a large volume that it could not be detected. Hapner & Robbins (1979) previously obtained evidence for SL denaturation at low pH. The protein cannot be electrophoresed on polyacrylamide gels at pH 2.0 and 2.3; at pH 2.0 it precipitates at the top of the gel. At pH 2.3 it yields a diffuse pattern. Sainfoin lectin is therefore unstable at pH 4.0 and below. In addition, the experiments reported here demonstrated that the lectin could be partially protected from denaturation at pH 4.0 by the addition of glucose. Binding to the carbohydrate site may therefore stabilize the lectin toward low pH denaturation.

Solubility of SL in Aqueous Solution

When SL was extensively dialyzed against water or PBS, the maximum concentration of soluble protein achieved at 25°C was about 1 mg/ml. However, in the presence of 0.1 M glucose or sucrose the protein was soluble at concentrations of 10 mg/ml and above.

The increased solubility of SL in the presence of carbohydrate, together with the increased stability of the protein at pH 4.0 in the presence of carbohydrate, supported the possibility of a stabilizing conformational change induced by sugar binding to the carbohydrate site. Such conformational changes are reported in the literature for other lectins and are reflected, among other effects, by altered ultraviolet absorbtion spectra in the absence and presence of interacting sugars (Matsumoto et al., 1980).

Affinity Chromatography of SL at Various pH Values

Sainfoin lectin was chromatographed on mannose-Sepharose at pH 7.0, 4.5 and 4.0. The lectin was eluted from the affinity matrix with a shallow gradient of glucose. Figure 2 shows that SL was completely retained by the mannose-Sepharose at pH 7.0 and 4.5, while only 13% was retained at pH 4.0. The lectin peak eluted with the glucose gradient was much broader in the case of the affinity chromatography at pH 4.5 than at pH 7.0. Total protein recoveries from the affinity chromatographies at pH 7.0, 4.5 and 4.0 were 92%, 71% and 66%, respectively.

The difference of SL behavior on mannose-Sepharose at pH 4.5 and 4.0 was interpreted to be due to protein denaturation at pH 4.0. The monomer species of SL that is the predominant form of the protein at pH 4.5 (see gel filtration of SL above) would bind

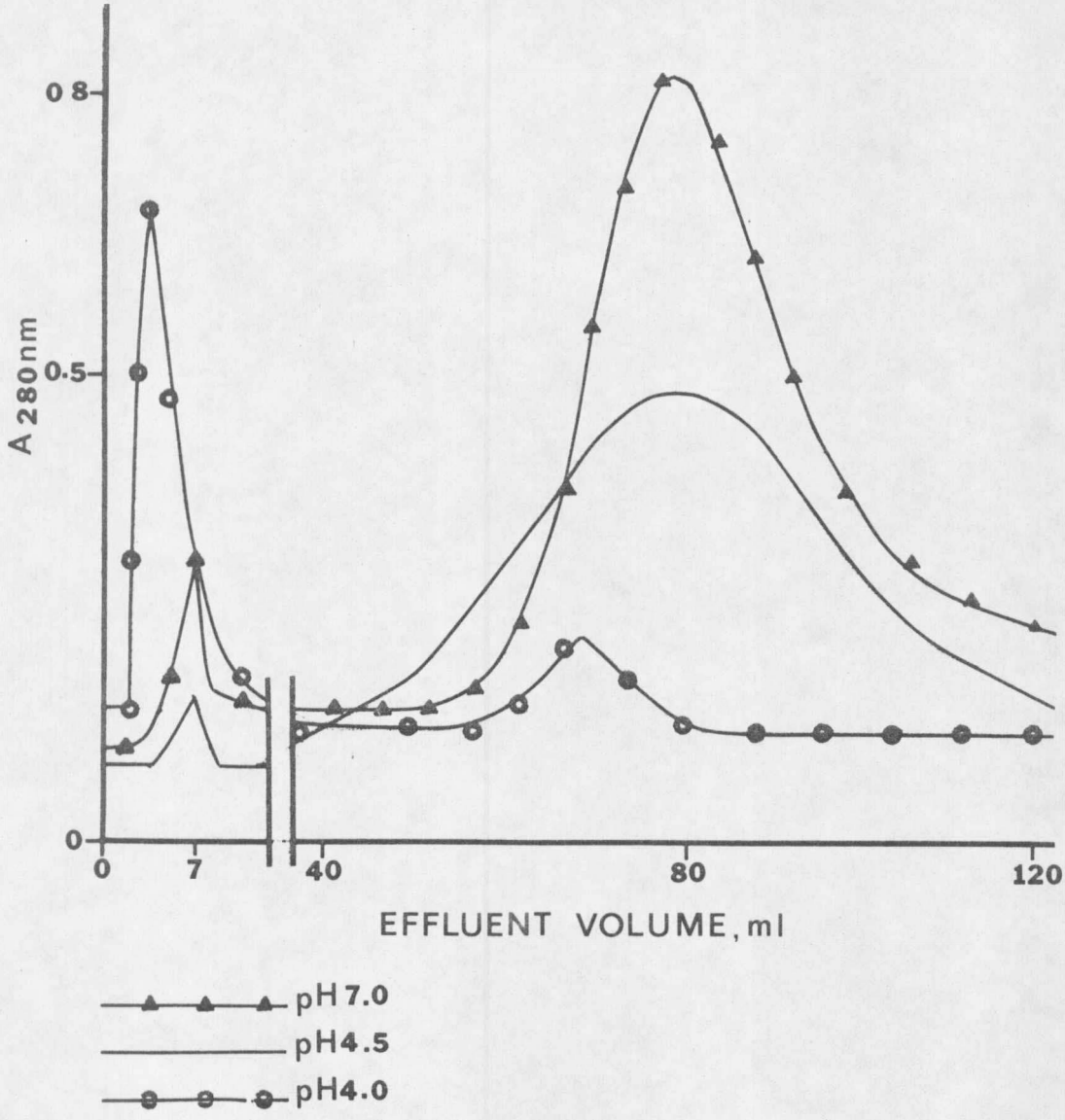


Figure 2. Affinity chromatography of SL on mannose-Sepharose at different pH values

mannose-Sepharose. At pH 4.0, increased denaturation of the monomer would suppress its carbohydrate binding ability. The broadness of the SL peak eluted with the glucose gradient from the affinity matrix at pH 4.5 suggested the presence of different species with similar but non-identical affinities toward the immobilized mannose or the free glucose, or both. These species could be the monomeric and dimeric forms of the protein. The results of affinity electrophoresis presented below support this interpretation.

Trypsin Digestion of SL

Sainfoin lectin was resistant to trypsin at pH 8.0, 25°C using a 10% weight ratio of enzyme to protein substrate and a 5 h incubation time. The absence of tryptic cleavage was demonstrated by the absence of base uptake during the incubation with the enzyme and the persistence of intact protein as shown by SDS electrophoresis. Addition of glucose to the reaction mixture did not induce trypsin digestion. Preincubation of SL at pH 2.5, 40°C for 30 min or at pH 11.0, 37°C for 15 min did not make the protein sensitive to trypsin digestion at 40°C.

Sainfoin lectin therefore appeared totally resistant to trypsin in its native conformation in the presence and absence of interacting carbohydrate. Denaturation of the protein at low or high pH did not expose any tryptic sites. It is possible that the

denatured molecules formed a trypsin resistant aggregate or that they renatured quickly when the pH was adjusted to 8.0. The extreme trypsin resistance of SL lead to the hypothesis that the protein could be a trypsin inhibitor. Sainfoin lectin did not affect hydrolysis of the synthetic substrate p-toluenesulfonyl-L-arginine methyl ester by trypsin, indicating that the trypsin active site was not blocked. Trypsin resistance of SL was therefore due to the unavailability of susceptible peptide bonds rather than possible inhibition of the trypsin activity.

Partial digestion of SL with 1% trypsin was observed at pH 8.0, 40°C after the protein was preincubated at pH 11.0, 60°C for 20 min. Digestion of the trypsin treated protein with carboxypeptidase B liberated 4 out of 11 lysine residues and 4.6 out of 12 arginine per monomer lectin. Approximately 36% of lysine and arginine bonds were cleaved by trypsin. When SDS electrophoresis was performed on the tryptic digest, only intact lectin was visible on the gels after the staining-destaining procedure, indicating the absence of large tryptic peptides. It was found that alkylation with iodoacetic acid or disulfide interchange with DTNB of the single cysteine in SL induced at least partial trypsin sensitivity. Complete digestion resulted after incubation at 50°C, pH 2.0 for 1 h prior to a second

trypsin addition. Details of the conditions for complete trypsin digestion are described in Experimental Procedures.

The results presented here demonstrated that partial trypsin digestion of SL did not result in preferential cleavage of a few peptide bonds. Instead a population of denatured molecules was completely digested while the others remained intact. The absence of trypsin sensitive peptide bond suggested that SL did not have a protease sensitive site analogous to the natural cleavage site of ConA, Asn118-Ser119 (Wang et al., 1971). The above results indicated that derivatization of SL cysteine sensitized the protein toward trypsin digestion and further incubation at low pH and high temperature completely denatured the protein. Derivatization of SL was done in the presence of metal chelator and strong denaturant. It is probable that sequestration of metals was at least partially responsible for SL denaturation.

Characteristics of Succinylated SL

Sainfoin lectin was succinylated by the method described in the Experimental Procedures. Reaction of an aliquot of the succinylated protein with the amino group reagent TNBS revealed the presence of one underivatized amino group per lectin monomer. The lectin was therefore 90% succinylated. Figure 3 shows the behavior of succinylated SL on mannose-Sepharose at pH 7.0. One-half of the

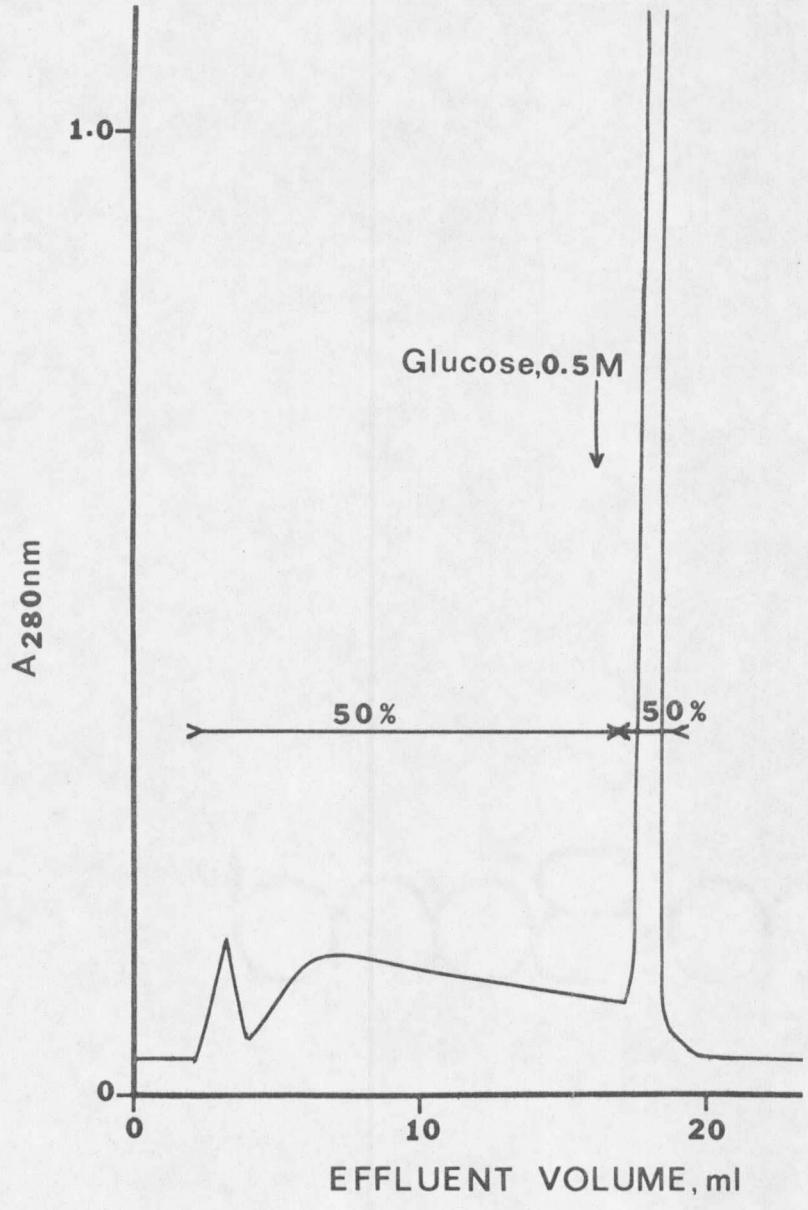


Figure 3. Affinity chromatography of 90% succinylated SL on mannose-Sepharose

protein was only retarded by the matrix and eluted with carbohydrate free PBS. The other 50% was retained and eluted with 0.1 M glucose, suggesting that 90% succinylated SL retained some binding activity. Figure 4 shows the elution profile of the modified protein on the previously calibrated Ultrogel AcA44 column (Figure 1). Two species eluted, of apparent molecular weights 59,000 and 48,000, possibly corresponding to succinylated dimer and succinylated monomer. The succinylated protein was incubated with 1% trypsin at 25°C, pH 8.0 for 2 h. No decrease in intact lectin was apparent by SDS electrophoresis and no tryptic fragments were observed.

The experiments reported above suggest that 11 out of the 12 free amino groups of SL could be succinylated without prior denaturation of the protein. The unmodified amino group could have been any one of the 11 lysine residues or the protein alpha-amino group. This specific residue would be protected from acylation by its local environment in the protein native conformation. Alternatively, the 10% unmodified amino groups could be randomly distributed. The succinylated protein contained two species of 59,000 and 48,000 MW. These MW values differed from those of SL dimer and monomer. Succinylation may have induced a conformational change in the protein resulting in a change in the Stokes radius of the molecule and its apparent molecular weight. It is conceivable

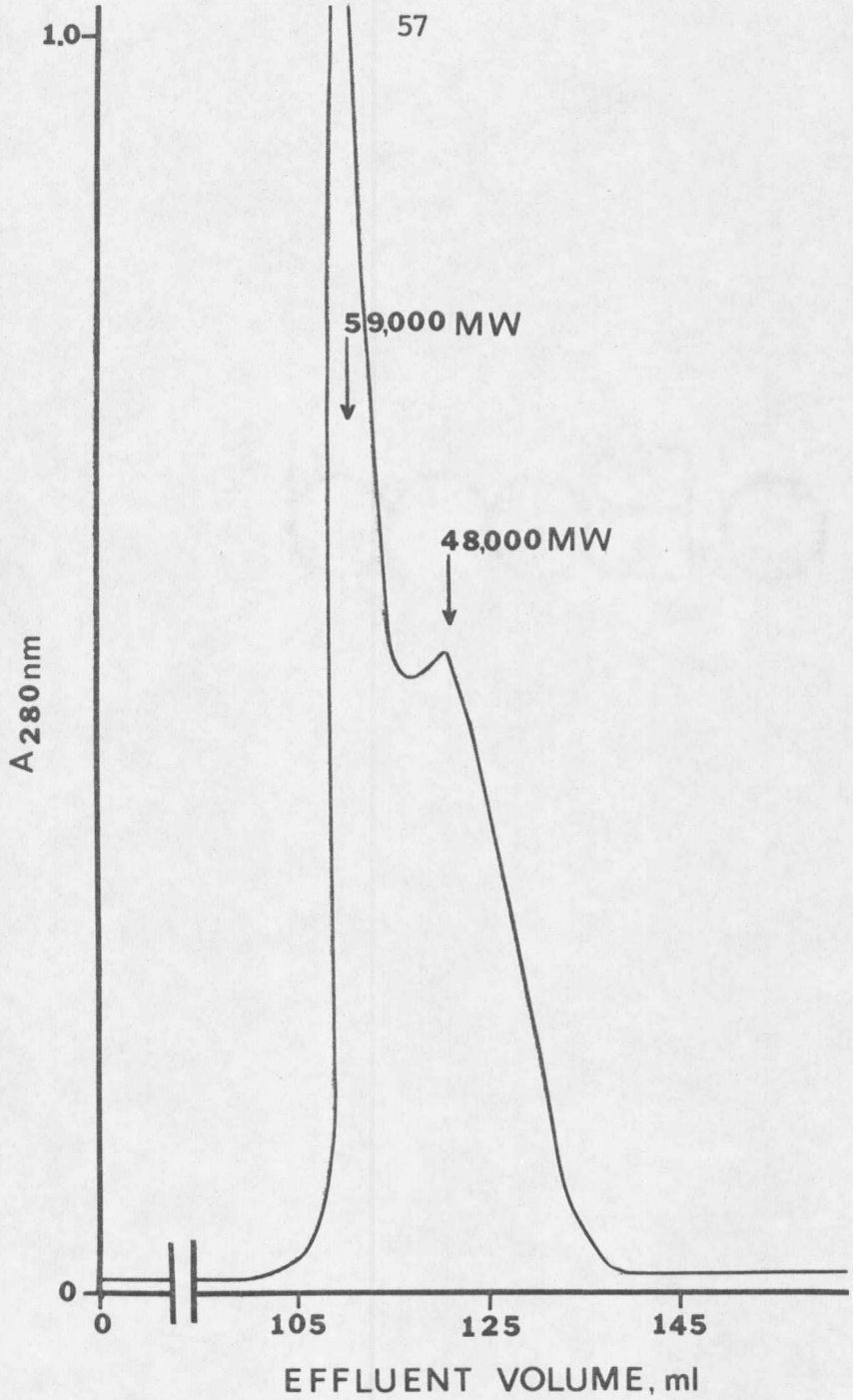


Figure 4. Gel filtration of 90% succinylated SL on Ultrogel AcA44

that the two species were succinylated dimer and succinylated monomer. Predominance of the largest species suggest that 90% succinylated lectin was mostly in the dimer state. The modified protein retained its trypsin resistance. No arginine sites were therefore exposed as a result of succinylation.

The 90% acylated SL was completely succinylated after incubatioon of the protein at pH 11.0 for 30 min. No free amino group could be detected with TNBS after this additional reaction. Eighty three percent of the completely succinylated SL eluted from mannose Sepharose in absence of free interacting carbohydrate (Figure 5), suggesting that the protein had lost its binding activity. Figure 6 shows the behavior of the completely modified lectin on Ultrogel AcA44. Two species eluted with apparent MW values of 63,100 and 49,000 probably corresponding to 100% succinylated dimer and monomer, respectively. When the modified protein was heated at 60°C for 30 min prior to gel filtration, it eluted mostly as a single species of 39,000 MW (Figure 7). The twice succinylated protein was still trypsin resistant at 25°C. Conditions for the successful trypsin digestion of the succinylated protein were the same as for the native protein, i.e., derivatization of the unique cysteine and denaturation by low pH and heat.

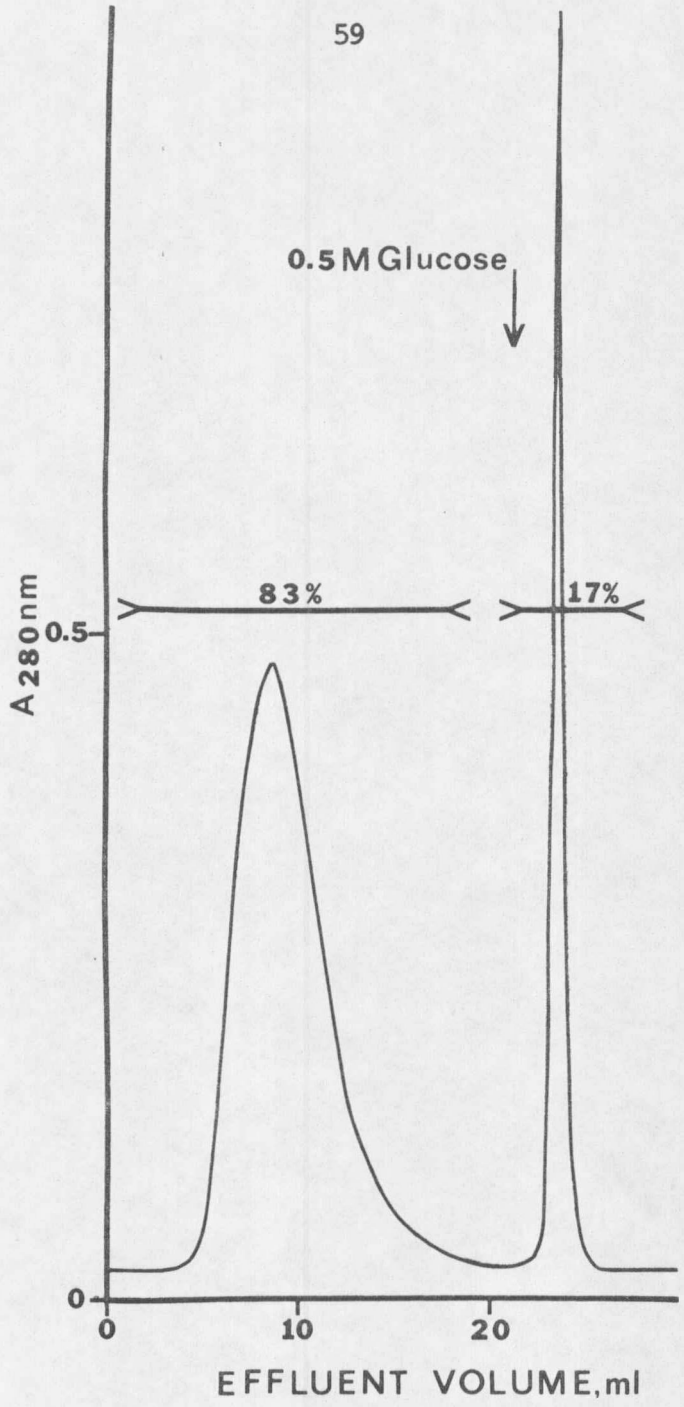


Figure 5. Affinity chromatography of 100% succinylated S1 on mannose-Sepharose

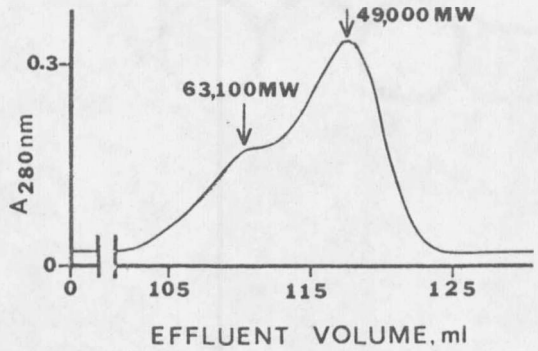


Figure 6. Gel filtration of 100% succinylated SL on Ultrogel AcA44

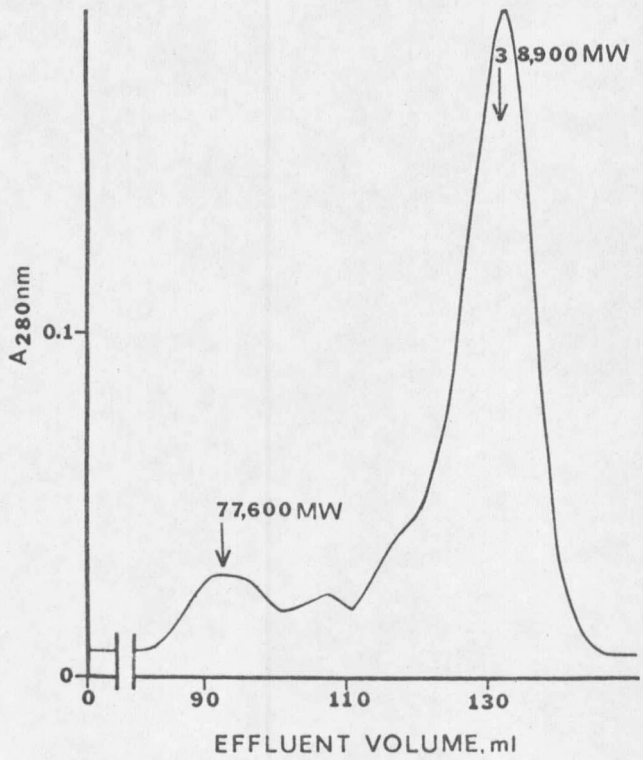


Figure 7. Gel filtration of preheated 100% succinylated SL on Ultrogel AcA44

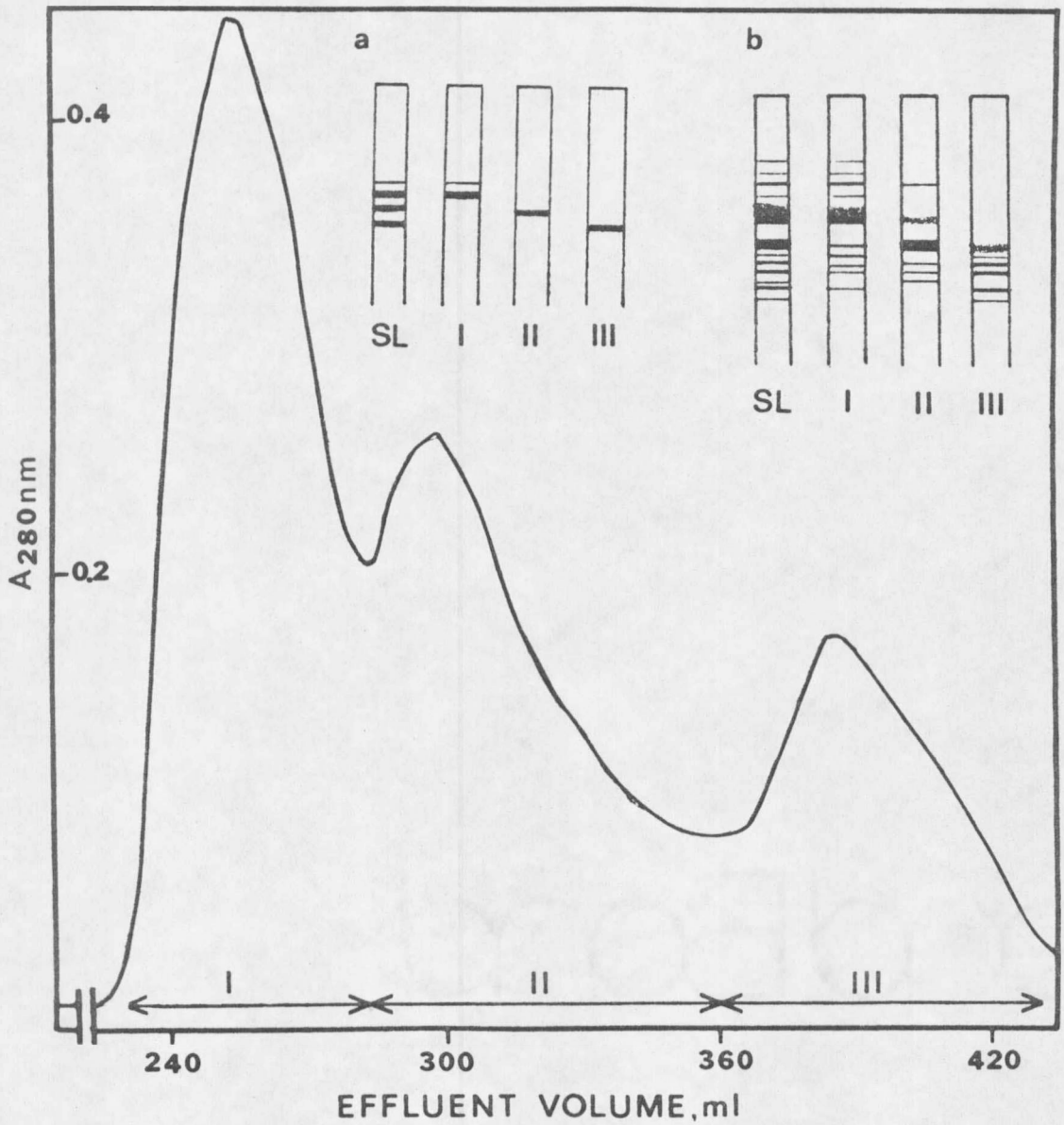
Denaturation of 90% succinylated SL allowed acylation of the previously underivatized amino groups. The 100% succinylated protein contained a large predominance of the 49,000 MW species that was transformed into a 39,000 MW species by heating. It can be speculated that totally succinylated SL was mostly in the monomeric conformation and heating converted it completely to that form. In view of the facts that 90% succinylated SL was a dimer that exhibited an average of one free amino group per monomer, it can be speculated that a single amino group was responsible for the dimer maintenance. The dimer maintenance could be due to formation of a specific salt bridge. An analogous situation stabilizes the tetrameric structure of ConA. In ConA, two lysines of one dimer form salt links with a glutamic acid residue of the other dimer (Reeke et al., 1975); consequently, acylation of ConA with maleic anhydride results in dimerization of the molecule (Gunther et al., 1973). There is, however, no additional evidence to support this idea with regard to SL. In the case of SL, exhaustive succinylation also induced the loss of carbohydrate binding activity indicating extensive conformational changes in the saccharide binding region. Persistent trypsin resistance of succinylated SL, however, emphasized the extreme stability of the molecule with regard to tryptic cleavage.

Section 2. Studies on Apparent Heterogeneity of Sainfoin Lectin

Ion Exchange Chromatography of SL

Figure 8 shows the behavior of SL on DEAE cellulose at pH 7.9. The protein was resolved into three fractions. Aliquots of these fractions were submitted to polyacrylamide gel electrophoresis (PAGE) at pH 9.5. Insert Figure 8a shows the results of the electrophoresis. The first fraction to emerge from the DEAE cellulose column (fraction I) contained the two most basic species of native SL. The other two fractions each showed a single electrophoretic band. The DEAE cellulose fractions were also analyzed by isoelectricfocusing as shown in the insert Figure 8b. The individual fractions exhibited multiple banding patterns that corresponded with a range or group of isofocusing bands on unpurified protein. The fraction I peak was asymmetrical (Figure 8) and was rechromatographed on the same column at pH 8.5. The elution pattern is shown in Figure 9. The peak was further resolved into two components.

The above results show that SL could be fractionated into four ionic species—IA, IB, II and III--by anion exchange chromatography. Hapner & Robbins (1979) previously reported that SL was homogeneous according to SDS electrophoresis, sedimentation velocity and PAGE at pH 4.0. In contrast, the protein was heterogeneous on PAGE at pH



Insert a. Polyacrylamide gel electrophoresis pH 9.5 of SL and fractions I, II & III

Insert b. Polyacrylamide gel isoelectricfocusing of SL and fractions I, II & III

Figure 8. Diethylaminoethylcellulose chromatography of SL

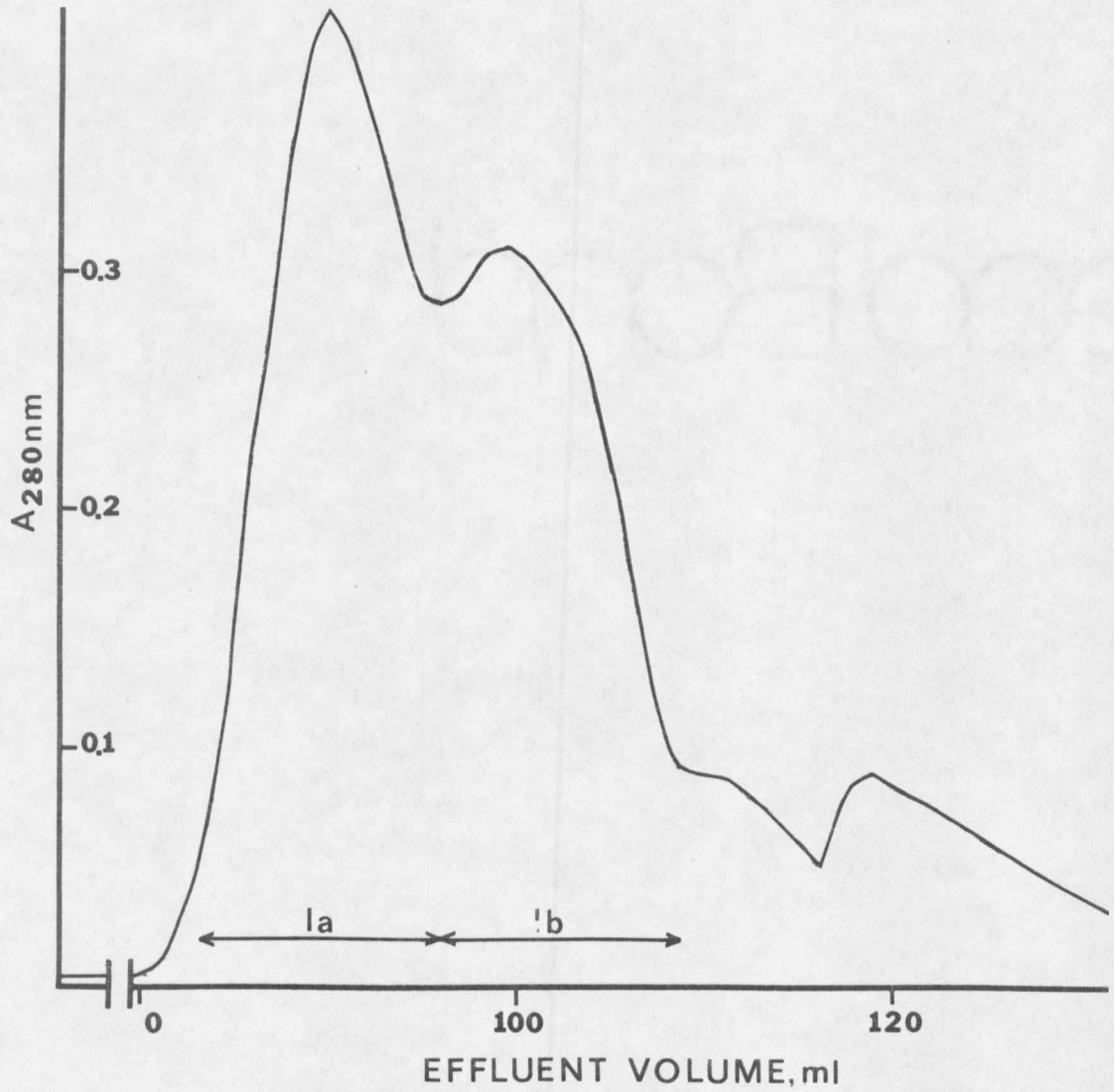
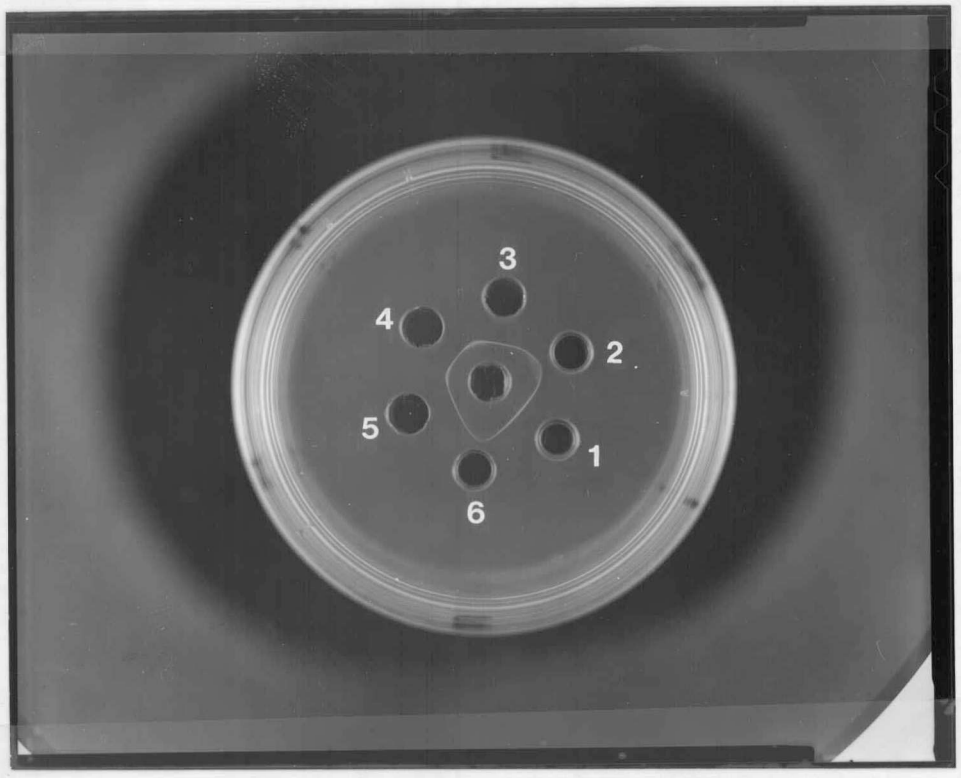


Figure 9. Diethylaminoethylcellulose chromatography of fraction I of Figure 8

9.5 and isoelectricfocusing. The authors speculated that the heterogeneity observed under certain conditions may have reflected the presence of isolectins or allelomorphic forms of the protein. Results from this work confirm SL heterogeneity through isolation of several molecular species. The basis of the heterogeneity was not as speculated above, but rather was derived from charge differences resulting from microheterogeneity among the protein amide groups (see below).

Analyses of SL Ionic Species

The four DEAE fractions all showed hemagglutination activity towards cat erythrocytes when tested at a concentration of 200 µg/ml. The four species were therefore lectins that expressed biological activity. Three out of four species exhibited identical immunodiffusion patterns against antiserum to native SL as shown in Figure 10. Species IA showed a slightly different immunoprecipitation pattern. Amino acid compositions of the lectin species are presented in Table 2. The compositions were extremely similar with only minor differences in lysine content. These results must be interpreted in light of the fact that subsequent work showed unfractionated SL to have a unique amino acid sequence (see Section 4 of Results). The similar amino acid compositions of SL ionic species showed them not to be true isolectins, i.e., different gene



1. unfractionated SL
2. ionic species Ia
3. ionic species Ib
4. ionic species II
5. ionic species III
6. conA

Figure 10. Immuno/double diffusion of SL ionic species against antiserum to native SL

Table 2. Amino acid composition of SL ionic species

	Ia	Ib	II	III
Lys	12.6	11.4	11.4	10.6
His	3.7	3.4	3.3	3.3
Arg	11.9	12.1	12.2	12.0
Asx	35.9	33.4	34.3	34.8
Thr	20.4	20.3	19.8	19.4
Ser	23.4	23.9	24.6	24.7
Glx	19.0	18.6	18.9	19.0
Pro	10.3	10.4	9.8	10.7
Gly	16.3	16.8	16.4	16.3
Ala	12.0	12.6	12.3	12.3
Val	14.8	17.3	16.9	17.2
Ile	12.8	12.8	12.6	12.8
Leu	20.5	19.5	20.0	19.8
Tyr	7.1	7.2	7.2	7.1
Phe	16.2	16.0	16.0	15.8

products. The charge differences allowing fractionation of the species were the result of differences in amide content of the proteins that would be undetectable by amino acid analysis as well as by immunoprecipitation and PAGE at pH 4.0. Charge

microheterogeneity was indicated by the presence of multiple isoelectric focusing bands. The correspondence between isoelectric focusing bands of purified fractions and grouped bands in unpurified protein further indicated that the DEAE column resolved groups or classes of charged species, rather than individual molecular forms. Sequence analysis would not detect amide heterogeneity due to partial conversion of the amide to acid form normally observed during Edman degradation (Pisano & Bronzert, 1969). For the reasons stated above and the improbability of isolectins differing only in aspartic-asparagine and glutamic-glutamine content, SL is concluded to be a single protein that shows charge heterogeneity due to variations in amide content. This variation may be naturally present or it might be an artifact of isolation.

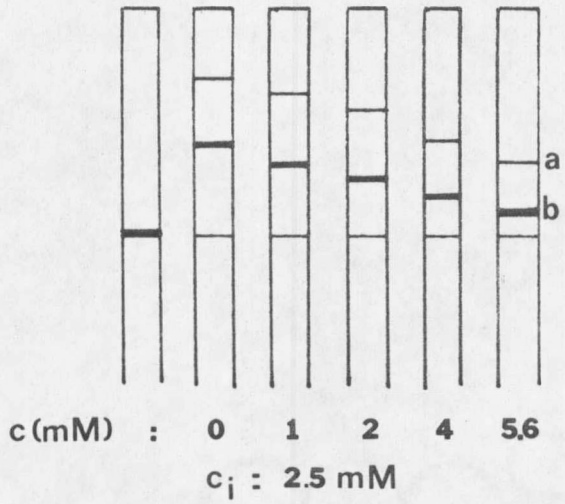
True isolectins have been observed for several plant lectins (Goldstein & Hayes, 1978). Examples of such multiple lectin forms are found in WGA, Bandeiraea simplicifolia I and phytohemagglutinin. The presence of multiple forms, as seen here, is not always related to different gene products. Soybean contains four ionic species separable on DEAE cellulose that have very similar amino acid compositions (Lis *et al.*, 1966). Lentil and pea each contain two ionic species that have almost identical amino acid compositions (Goldstein & Hayes, 1978). Peanut lectin can be resolved into eight

ionic species by isoelectric focusing (Miller, 1982); however, no amino acid composition data were shown. It is possible that the type of heterogeneity shown by SL, i.e., amide differences, is common among plant lectins.

Affinity Electrophoresis of SL at pH 4.0

Apparent heterogeneous forms of SL differing in carbohydrate binding strength was observed by this technique. Sainfoin lectin was submitted to affinity electrophoresis according to the following. Allyl alpha-glucoside was synthesized and copolymerized with acrylamide to form water soluble copolymers. These copolymers were added to the usual polymerization mixture resulting in glucose entrapment in the polyacrylamide gels. Sainfoin lectin was electrophoresed on these gels using the standard discontinuous buffer system of pH 4.0. A pH value of 4.0 was chosen because SL shows one electrophoretic band at this pH (Hapner & Robbins, 1979). Figure 11 shows the results of a typical affinity electrophoresis run. The first gel on the left was a control gel in which galactose was the immobilized sugar. The other gels contained immobilized glucose and various amounts of added free glucose. Three bands were visible in the affinity gels of SL. One of them showed a migration unaffected by the presence of immobilized glucose. The other two bands had retarded mobilities in the affinity gels that became less

retarded with added concentrations of free glucose. Binding constants toward free and immobilized glucose were calculated from these data and are fully presented in Section 3 of the Results.



Note: The gel at the far left contains immobilized galactose. The other gels contain immobilized glucose in concentration ci and free glucose in concentration c.

Figure 11. Affinity electrophoresis at pH 4.0 of SL

Affinity electrophoresis is a new technique in which proteins can be separated on the combined basis of their specific affinity to immobilized ligands and their ionic charge character (Horejsi, 1981). The presence of immobilized ligand decreases the mobility of the interacting protein. The mobility is restored by the addition

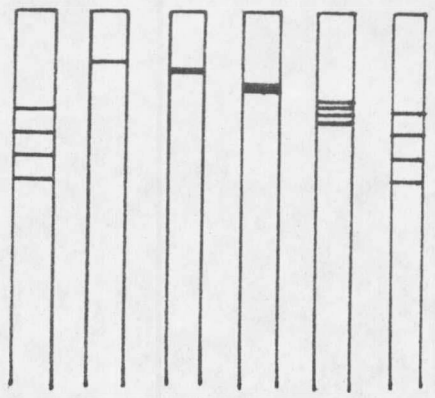
of free ligand to the affinity gels. These results showed that SL was resolved into three components, and since the lectin shows only one band on PAGE at pH 4.0 (Hapner & Robbins, 1979), the components therefore differed by their carbohydrate binding affinity. Entrapment of a non-interacting sugar (galactose) in the control gel did not affect SL mobility, indicating that the alteration of protein mobility in the glucose containing gels was not due to a non-specific phenomenon. The SL species whose migration was independent of the presence of glucose was presumably denatured SL. The other two components resolved on the affinity gels possibly corresponded to SL monomer and dimer species that differed in their interaction with the immobilized glucose. The K_d values for free glucose calculated for the two species were very similar, whereas the dissociation binding constants for immobilized glucose were different (see Section 3 of Results). It is conceivable that the dimer would interact more strongly with the entrapped glucose because of its greater valency and would be more retarded on the affinity gels. In agreement with this hypothesis was the observation that the fastest component was in larger amount than the slowest one, as would be expected from the predominance of monomers at pH 4.0 (see Section 1 of Results). Alternatively, dimer and monomer may differ in their access to the immobilized sugar for

steric reasons, or the sugar may have been chemically immobilized in different conformations resulting in different binding constants. In any case, there is no other evidence that suggests SL may have different carbohydrate binding strengths and the apparently contrary results here must be ascribed to experimental conditions. Similarly, the apparent conflict with the gel filtration experiments at pH 4.0 which showed little carbohydrate binding (Section 1 of Results) is likely due to an elevated pH in the affinity electrophoresis gel.

Affinity Electrophoresis of SL at pH 9.5

Figure 12 illustrates the behavior of SL on affinity electrophoresis at pH 9.5. The control gel on the left with immobilized galactose exhibited the usual four bands patterns observed in native SL (Hapner & Robbins, 1979). Only one band was present on the affinity gels in the absence of free glucose. When glucose was added the single band progressively diffused into the four-banded pattern characteristic of SL.

Because of the limited migration of the protein in the presence of little or no free glucose, one possible interpretation was that each gel in fact exhibited several bands, but it was not possible to distinguish them. If this hypothesis was correct, the K_d for



Note: The gel at the far left contains immobilized galactose. The other gels contain immobilized glucose in concentration ci and free glucose in concentration c.

Figure 12. Affinity electrophoresis at pH 9.5 of SL

the ionic species toward immobilized glucose would all be different and of such values as to produce exactly the same protein mobility for the four ionic species in absence of free glucose. An alternative interpretation was that the presence of immobilized ligand eliminated the heterogeneity of SL visible on alkaline electrophoresis in a fashion analogous to the elimination of concanavalin A heterogeneity on isoelectric focusing in presence of sugar (Lis & Sharon, 1973). However, it is unclear why free glucose would not act in the same way as entrapped glucose. Published data show that soybean and lentil lectin ionic species are present on affinity electrophoresis and do not differ in their binding

parameters (Hauzer et al., 1979). This odd behavior of SL in affinity electrophoresis at pH 9.5 remains unexplained.

Section 3. Ligand Interactions Studies

Metal Analyses

The results of atomic absorption metal analyses of SL are shown in Table 3.

Table 3. Metal analyses of sainfoin lectin by atomic absorption

Experiment	Ca^{++}/monomer	Mg^{++}/monomer	Mn^{++}/monomer
1	1.03	0.54	0.38
2	1.17	0.03	0.26
3	1.12	0.17	0.65
Average	1.11	0.25	0.43

In the first two experiments of Table 3 an EDTA extract of SL was submitted to atomic absorption. In the third, the whole protein was digested with nitric acid and aspirated into the instrument. Sainfoin lectin had one Ca^{++} per monomer and less than stoichoimetric amounts of Mg^{++} and Mn^{++} with a predominance of Mn^{++}. An hemagglutination assay of SL was performed in the presence of 0.1 M EDTA. The lectin was found to be fully active. EDTA does, however, sensitize the molecule toward heat denaturation and loss of activity (K. D. Hapner, personal communication). Most plant lectins

whose metal content is known possess stoichiometric amounts of Ca^{++} and Mn^{++} (Goldstein & Hayes, 1978). Three leguminous lectins contain Mg^{++} instead of Mn^{++}: peanut lectin (Neurohr et al., 1980), Bandeiraea simplicifolia I (Hayes & Goldsein, 1974), and a lectin from Phaseolus vulgaris (Itoh et al., 1980). Concanavalin A which contains one Ca^{++} and one Mn^{++} per monomer, does not require its transition metal for biological activity (Christie et al., 1978) and retains less than stoichiometric amounts of Mn^{++} when dialyzed extensively. Furthermore, the transition metal site of demetallized ConA can be filled with metals other than Mn^{++}, such as Co^{++}, Ni^{++}, Zn^{++} and Cd^{++} (Shoham et al., 1973).

It is likely that SL which shows strong homologies with ConA and near total conservation of the metal binding sites (see Discussion), possesses one Ca^{++} and one Mn^{++} binding site. The lectin could lose some of its transition metal during the dialysis stage of the purification procedure and the vacant sites could then be filled by other metals such as Mg^{++} if present in the environment. It would explain the variations in content of Mn^{++} and Mg^{++} of the three experiments reported here. Sainfoin lectin was found biologically active in the presence of EDTA, suggesting that the metal(s) needed for maintenance of biological activity (by analogy with ConA) were tightly bound to the molecule. In view of

the variability in transition metal content of SL, it is reasonable to conclude that the Ca^{++} site is of high affinity and is responsible for stabilization of the active molecule whereas the transition metal site is of lower affinity and specificity.

Estimation of K_d for SL-Glucose Complex by Affinity Electrophoresis

Sainfoin lectin was submitted to affinity electrophoresis at pH 4.0 as shown in Figure 11. The affinity gels contained a constant concentration of immobilized glucoside (Ci). The concentration of free glucose (C) was varied. A plot of L/Lo-L *vs.* C (where L = protein mobility in affinity gels, Lo = protein mobility in control gels) is shown in Figure 13. Line a was obtained with the slowest lectin component on the affinity gels. Line b corresponded to the fastest component. The K values for the lectin-free glucose complexes were obtained from the x-axis intercepts, K = -C. The Ki values for the lectin-immobilized complexes were calculated from the y-axis intercepts, Ki = (L/Lo-L) X Ci (Horejsi *et al.*, 1977). Three independent affinity electrophoreses were performed. In all cases the straight lines computed by regression analysis for the plots L/Lo-L *vs.* C corresponded to correlation coefficients equal or superior to 0.990. The results of the three experiments are summarized in Table 4.

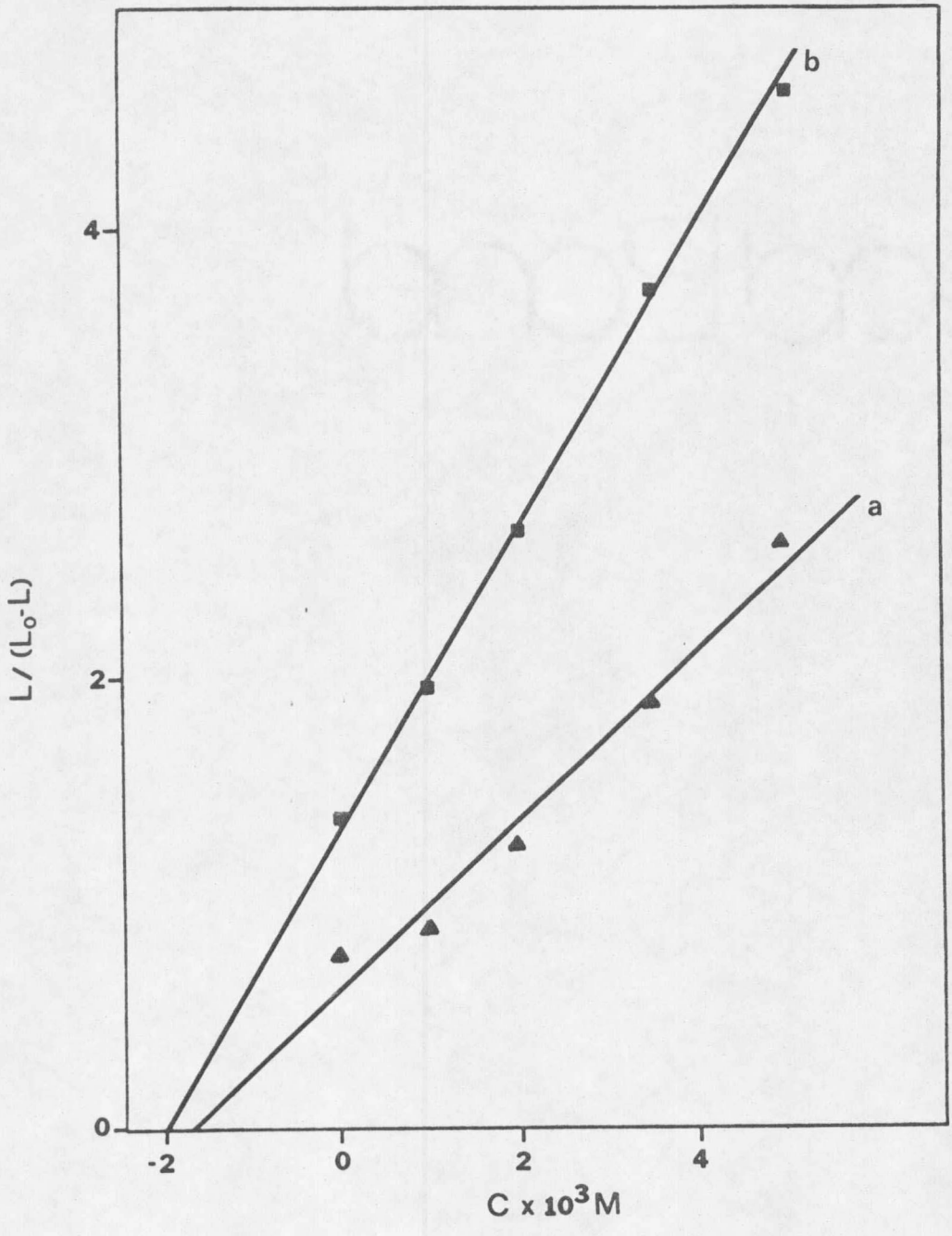


Figure 13. Plot (L/Lo-L) *vs.* C derived from the affinity electrophoresis of Figure 11

Table 4. K_d values for SL-glucose complexes determined by affinity electrophoresis at pH 4.0

	Experiment Number 1	2	3	Average Values
Ka(M)	1.79×10^{-3}	1.40×10^{-3}	1.58×10^{-3}	$(1.59 \pm 0.19) \times 10^{-3}$
Kb(M)	1.93×10^{-3}	1.96×10^{-3}	2.02×10^{-3}	$(1.97 \pm 0.04) \times 10^{-3}$
Ki_a(M)	0.57×10^{-3}	0.97×10^{-3}	0.90×10^{-3}	$(0.81 \pm 0.21) \times 10^{-3}$
Ki_b(M)	2.05×10^{-3}	3.63×10^{-3}	4.03×10^{-3}	$(3.29 \pm 1.05) \times 10^{-3}$

Dissociation constants for the lectin-free glucose complexes were found to be 1.59×10^{-3} M and 1.97×10^{-3} M for the slowest and fastest lectin species, respectively. The Ki values of the two lectin species differed from each other by a factor of 3.6 to 4.5, indicating that the differences in affinity of the two species toward immobilized glucose were responsible for their separation on the affinity gels.

Because of the peculiar behavior of SL on affinity electrophoresis at pH 9.5 (Figure 12), four sets of K and Ki could be calculated with L/Lo-L *vs.* C plots. The K values varied between 2.5×10^{-3} M (for the slowest component) and 5.2×10^{-3} M (for the fastest component). The Ki values varied between 4.2×10^{-3} M (slowest species) and 1.9×10^{-3} M (fastest species).

Dissociation constants of Sl-glucose complexes were determined from the following equation (Horejsi et al., 1977):

$$\frac{L}{Lo-L} = \frac{Ki}{Ci}\left(1 + \frac{C}{K}\right)$$

Evaluation of Ki values with this equation is valid only under the assumption that all molecules of immobilized ligand are accessible to interaction with the protein (Horejsi, 1981). In the affinity electrophoreses at pH 4.0 reported here, the affinity gels for experiments #2 and #3 were polymerized in the presence of the same batch of glucosyl copolymer while a different batch was used for experiment #1. It is likely that the accessibility of immobilized glucose varied with the batch of glucosyl copolymer used. It would explain the variations in Ki values between the first experiment and the others. The two SL species separable on affinity electrophoresis pH 4.0 had similar affinity toward free glucose. This observation was compatible with the previously made hypothesis that the two species were the dimer and the monomer forms of SL (see Section 2 of Results). The K values for SL species were comparable to those of ConA-glucose (4.3×10^{-3} M) and lentil lectin-glucose (6.4×10^{-3} M) complexes as determined by affinity electrophoresis by Horejsi et al. (1977). The SL component that was slowest on the

affinity gels at pH 4.0 exhibited a Ki value lower than its K. On the contrary, the fastest component had a Ki larger than its K. Horejsi et al. (1977) reported an increased affinity toward immobilized ligands relative to free ligands for all the lectins they tested on affinity electrophoresis. These authors explained the phenomenon by the influence of the aglycon and the multiplicity of sugar binding sites in lectin molecules. The observations reported here therefore suggested that SL slowest species was its dimer. Theoretical calculations predict that multiple interactions of a protein molecule with immobilized ligand result in curvilinear plots (Horejsi, 1979). The plots are in some cases experimentally indistinguishable from straight lines, but their use for the evaluation of Ki and K leads to values lower than true Ki and K. The Ki and K determined in the experiments reported here for the component believed to be the dimer of SL may therefore have been incorrect. The importance of the K and Ki values calculated for SL ionic species on affinity electrophoresis at pH 9.5 was certainly questionable because of the odd behavior of the ionic species in these conditions. However, it was probably significant that the K values calculated for the ionic species differed at most by a factor of three from the K values calculated by affinity electrophoresis at

pH 4.0. It suggested that the affinity of SL toward glucose did not vary to any great extent between pH 4.0 and pH 9.5.

Equilibrium Dialysis

Sainfoin lectin was slightly succinylated by a single addition of succinic anhydride. The succinylated protein emerged from an Ultrogel AcA44 column in the same elution volume as native SL, indicating that the succinylation had not modified the protein molecular size and hence the valency. The slightly succinylated protein also had the same hemagglutination titer as native SL. The binding ability of the lectin was therefore not affected by the modification. Succinylated SL was soluble at 1.58 mg/ml after exhaustive dialysis to remove carbohydrate and was used for equilibrium dialysis. The ligand chosen was uniformly labeled ^{14}C-glucose. A Scatchard plot (1949) of the results is shown in Figure 14. Regression analysis yielded a straight line of slope -284.32 M^{-1}, x-intercept 6.28 X 10^{-5} M and correlation coefficient -0.954. Dissociation constant and number of binding sites (n) were calculatd from the slope and the x-intercept, respectively. Slope = $-1/K_d$; x-intercept = n X total protein concentration. A K_d of 3.52 X 10^{-3} M was determined. Using the concentration of lectin monomer, n = 1.03 was calculated. Sainfoin lectin therefore contained one binding

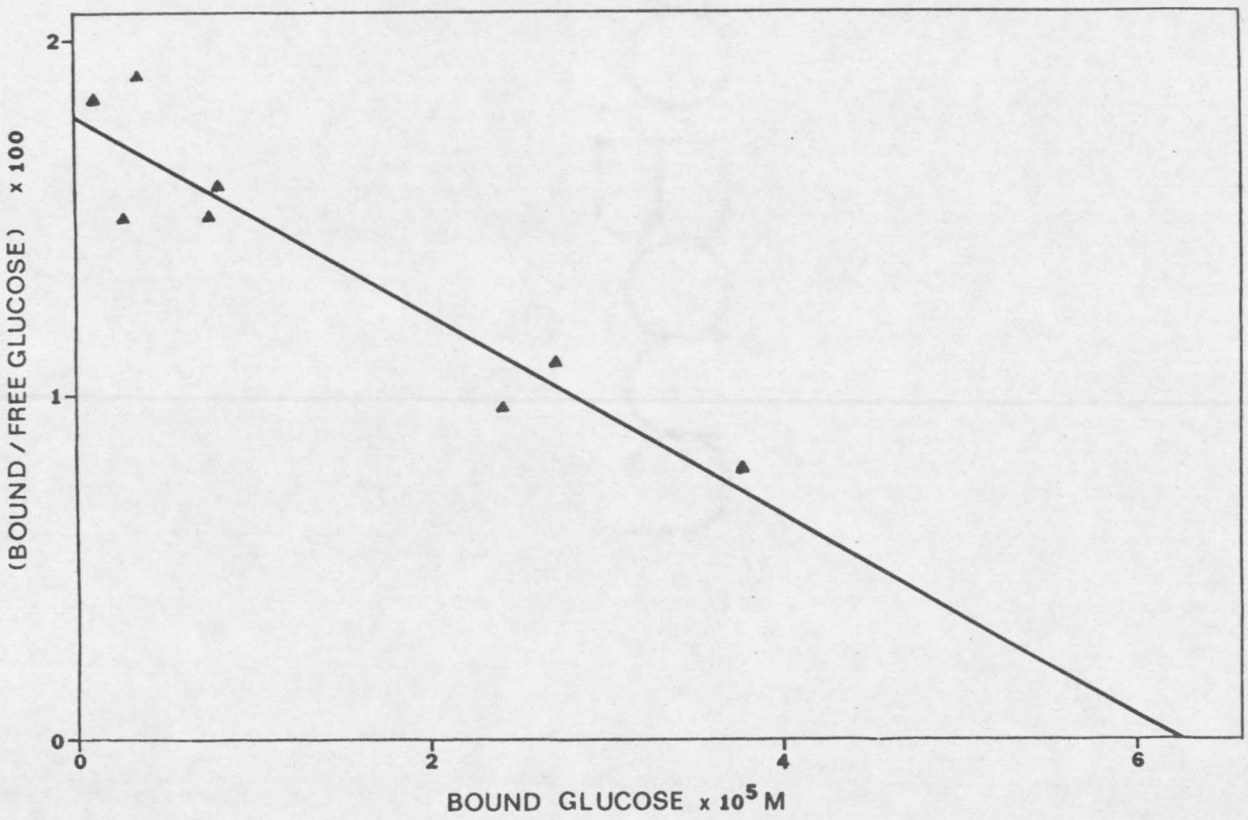


Figure 14. Determination of the dissociation constant of glucose to SL by equilibrium dialysis

site per subunit. The linearity of the Scatchard plot indicated that the two binding sites of SL dimer were equivalent.

Equilibrium dialysis of SL was performed on slightly succinylated protein in order to increase the solubility of the protein. Previous experiments done with native SL yielded questionable data because of the limited solubility of the protein and the relatively weak affinity of the lectin for glucose. These factors had resulted in small amounts of bound carbohydrate and very small differences in the carbohydrate concentrations of the protein and protein-free compartments. Large errors had therefore characterized the data. In the experiments reported here slight succinylation of SL did not affect its valency or its hemagglutination activity. Consequently, the K_d calculated by equilibrium dialysis of slightly succinylated SL was also valid for native SL. The accuracy of the experiment was also enhanced by using [^{14}C-U] glucose of highest possible specific activity in order to bind a number of counts in large excess to the counting error. The K_d calculated by equilibrium dialysis was in good agreement with those calculated from the affinity electrophoresis experiments. The K_d was also similar to the concentration of glucose (3.1×10^{-3} M) resulting in 50% inhibition of SL agglutination of cat erythrocytes (Kouchalakos & Hapner, manuscript in preparation). Inhibition of

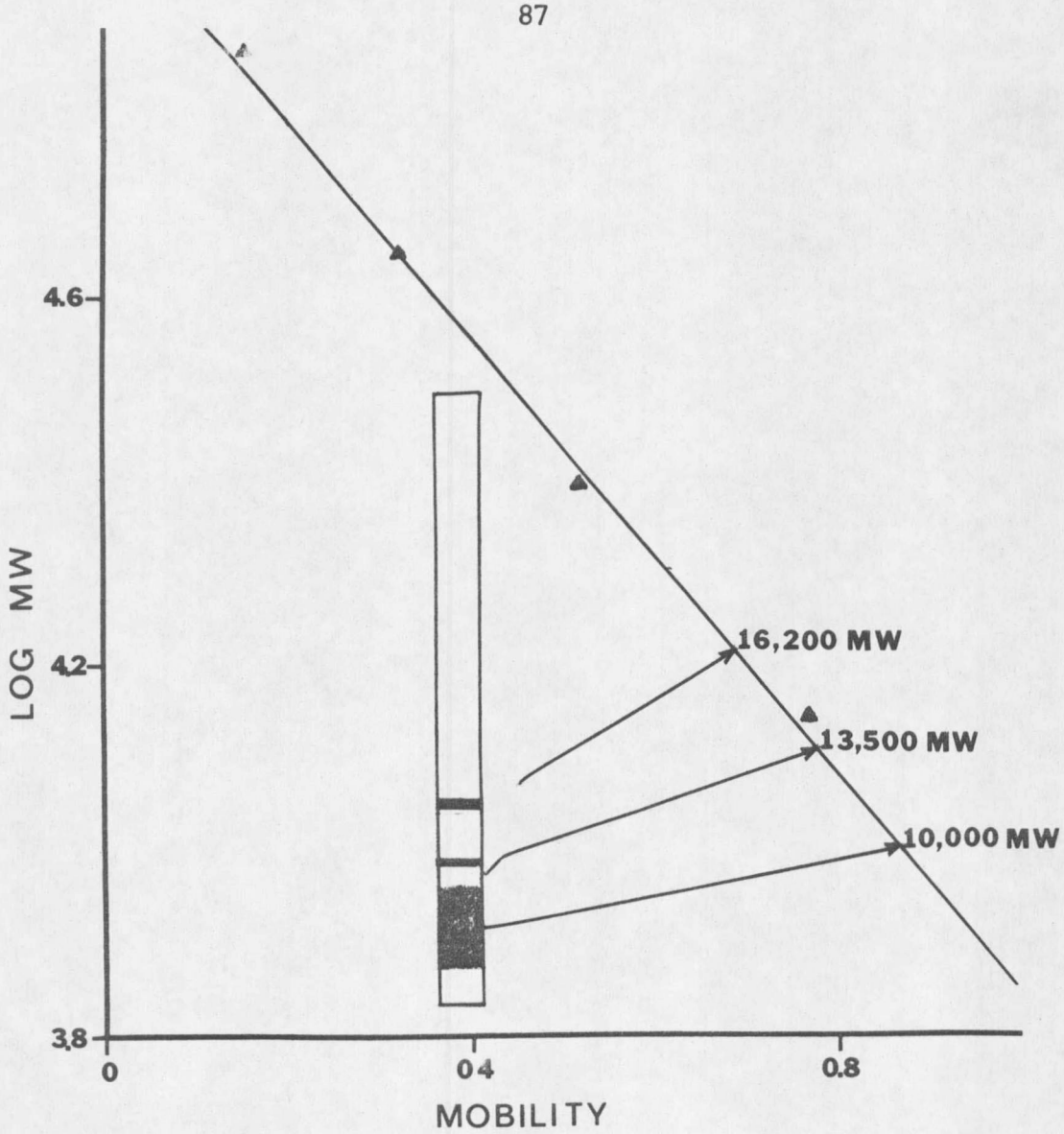
lectin induced agglutination by carbohydrates has been interpreted in terms of dissociation constants (Poretz & Goldstein, 1970). The sugar concentration that produced 50% inhibition of SL hemagglutination cannot be considered as true dissociation constant, as the phenomenon observed was one of competition between the monosaccharide and the carbohydrates of the erythrocytes surface for the lectin binding site. However, it can be regarded as a K_d estimate. The concentration of mannose needed to inhibit 50% of SL agglutination of cat erythrocytes is 0.4×10^{-3} M (Kouchalakos & Hapner, manuscript in preparation). Sainfoin lectin has therefore more affinity toward D-mannose than toward D-glucose. The same observation is true of all lectins from the D-mannose (D-glucose) binding family (Allen et al., 1976).

Section 4. Primary Structure Determination

Partial Acid Hydrolysis

Sainfoin lectin was partially cleaved by incubation in 10% acetic acid containing 6 M guanidine HCl at 40°C, pH 2.5. An aliquot of the hydrolysate was submitted to SDS electrophoresis as shown in Figure 15. In addition to a trace of remaining intact lectin, three other bands were visible which corresponded to cleavage products. The MW values of the fragments were estimated from a plot, log MW *vs.* protein mobility, established with standard proteins (Figure 15). The three bands corresponded to 16,200, 13,500, and 10,000 MW. The band that migrated furthest was very diffuse and probably contained more than one fragment. The 10,000 value was an average MW for all the peptides electrophoresed at this location. Figure 16 shows the gel filtration of the cleaved lectin on Biogel A5m in 6 M guanidine HCl. Seven peaks of absorbtion at 230 nm were observed, indicating that at least seven peptides were produced during the lectin partial hydrolysis. No hydrolysis of SL occurred when it was incubated in absence of guanidine HCl at pH 2.5.

The conditions under which SL was hydrolyzed have been reported to induce cleavage at aspartyl-prolyl bonds (London, 1977). Later sequence analysis demonstrated the presence of one such bond in SL.



Note: The plot mobility vs. log MW was obtained with standard proteins

Figure 15. Sodium dodecylsulfate electrophoresis of SL partial acid hydrolysate

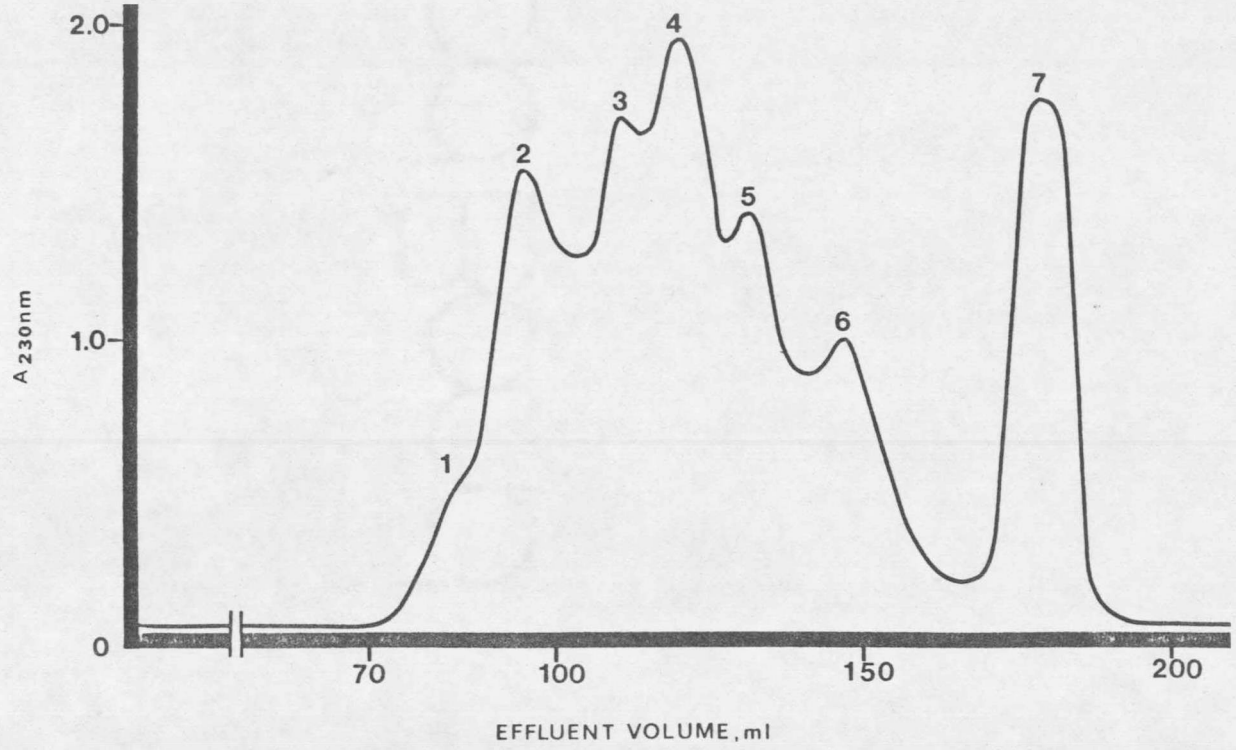


Figure 16. Gel filtration of SL partial acid hydrolysate on Biogel A5m

Partial acid hydrolysis therefore resulted in additional peptide bond cleavages. In view of the observation that no hydrolysis of SL could be achieved in absence of denaturant, it is probable that the aspartyl-prolyl bond of SL was not more acid labile than other peptide bonds and it could not be specifically cleaved. Gel filtration of the peptides required the use of Biogel A5m whose fractionation range was 10,000 to 5 X 10^6 MW. The presence of denaturant during the chromatography presumably denatured the peptides totally. This resulted in very large Stockes radii and the peptides behaved as if they were large proteins. This phenomenon has been reported previously (Fisch et al., 1969).

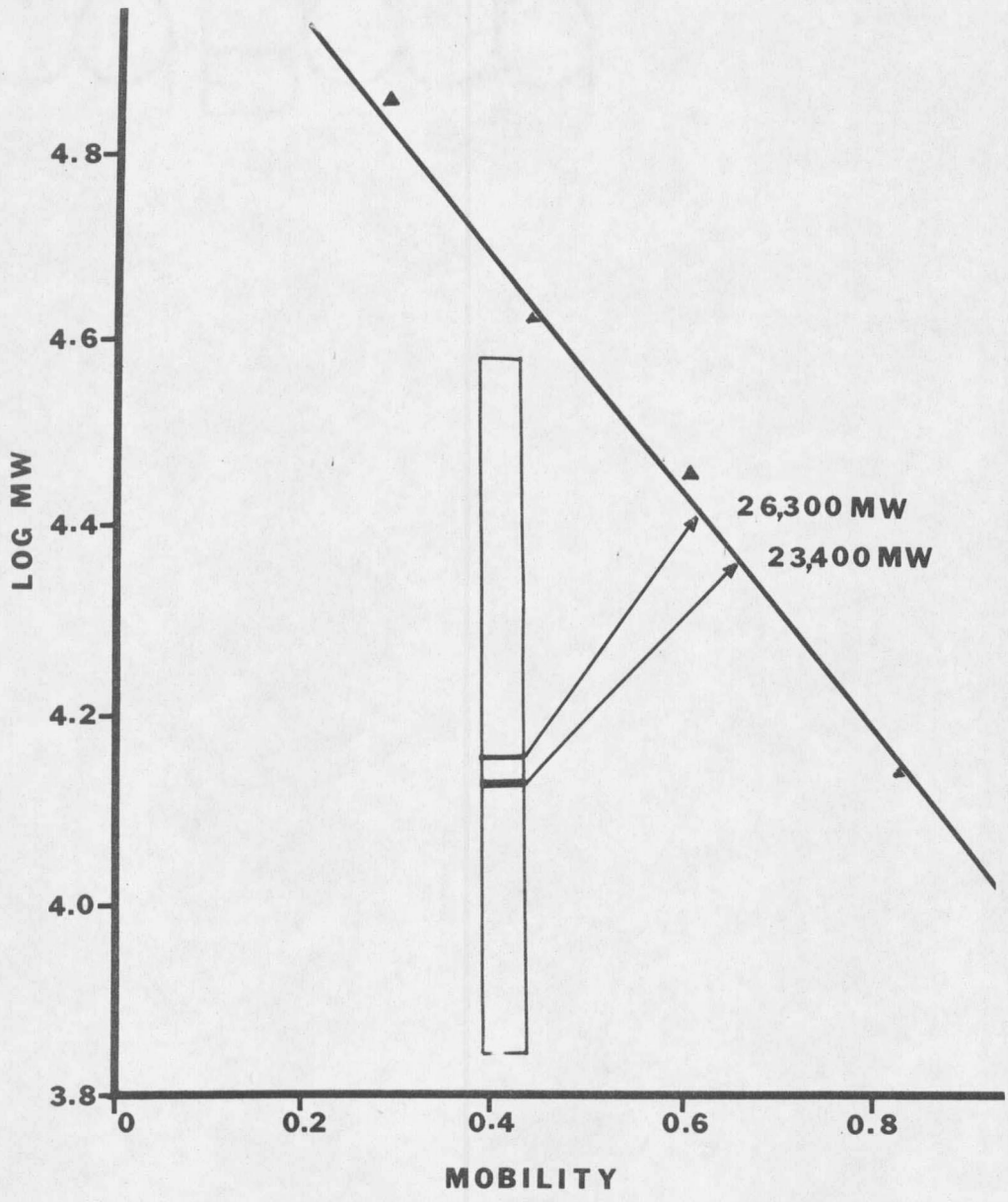
The peptides recovered from the gel filtration on Biogel A5m were lost when desalted in water, 0.01 M acetic acid or 0.01 M ammonium bicarbonate due to precipitation in the desalting column. They could be desalted in 10 to 50% formic acid, although significant losses occurred even in these conditions. After lyophilization, all the fractions were insoluble in aqueous buffers and in pyridine containing buffers. Fractions 2 and 4 were also insoluble in the presence of SDS. The peptides could only be dissolved in 10% formic acid or in 6 M guanidine HCl. Essentially the same results were obtained when SL was hydrolyzed in the absence of denaturant by incubation at 83°C and pH 2.3 for one-half hour.

Peptides were produced with very restricted solubility. Furthermore, they were very poorly resolved by gel filtration in the absence of denaturant, indicating a tendency toward aggregation.

Because of the impurity of the peptides resulting from partial acid hydrolysis of SL and because no purification techniques were available tjat could be performed in organic acids or in strong denaturant, the peptides were not useful for sequence analysis.

Cleavage at Cysteine

Sainfoin lectin was cyanylated with 2-nitro-5-thiocyano-benzoic acid (NTCB). The reaction was monitored by the increase in absorbtion at 412 nm due to the liberation of 2-nitro-5-thio-benzoate. The yield of cyanylation of the one cysteine residue was calculated to be 80%. The cyanylated protein was cleaved by exposure to pH 9.0. An aliquot of the cleaved protein was submitted to SDS electrophoresis as shown in Figure 17. Two bands were visible. One component migrated at the position of intact SL; it exhibited a MW of 26,900 and corresponded to unreacted protein. The other component had a MW of 23,400 and was a product of the cleavage at cysteine. Figure 18 shows the gel filtration of cleaved SL on Sephadex G50. Two peaks of absorbtion at 230 nm were observed in addition to a salt peak. The void volume peak contained intact protein and the 23,440 MW cleavage product according to subsequent



Note: The plot mobility vs. log MW was obtained with standard proteins

Figure 17. Sodium dodecylsulfate electrophoresis of SL cleaved at cysteine

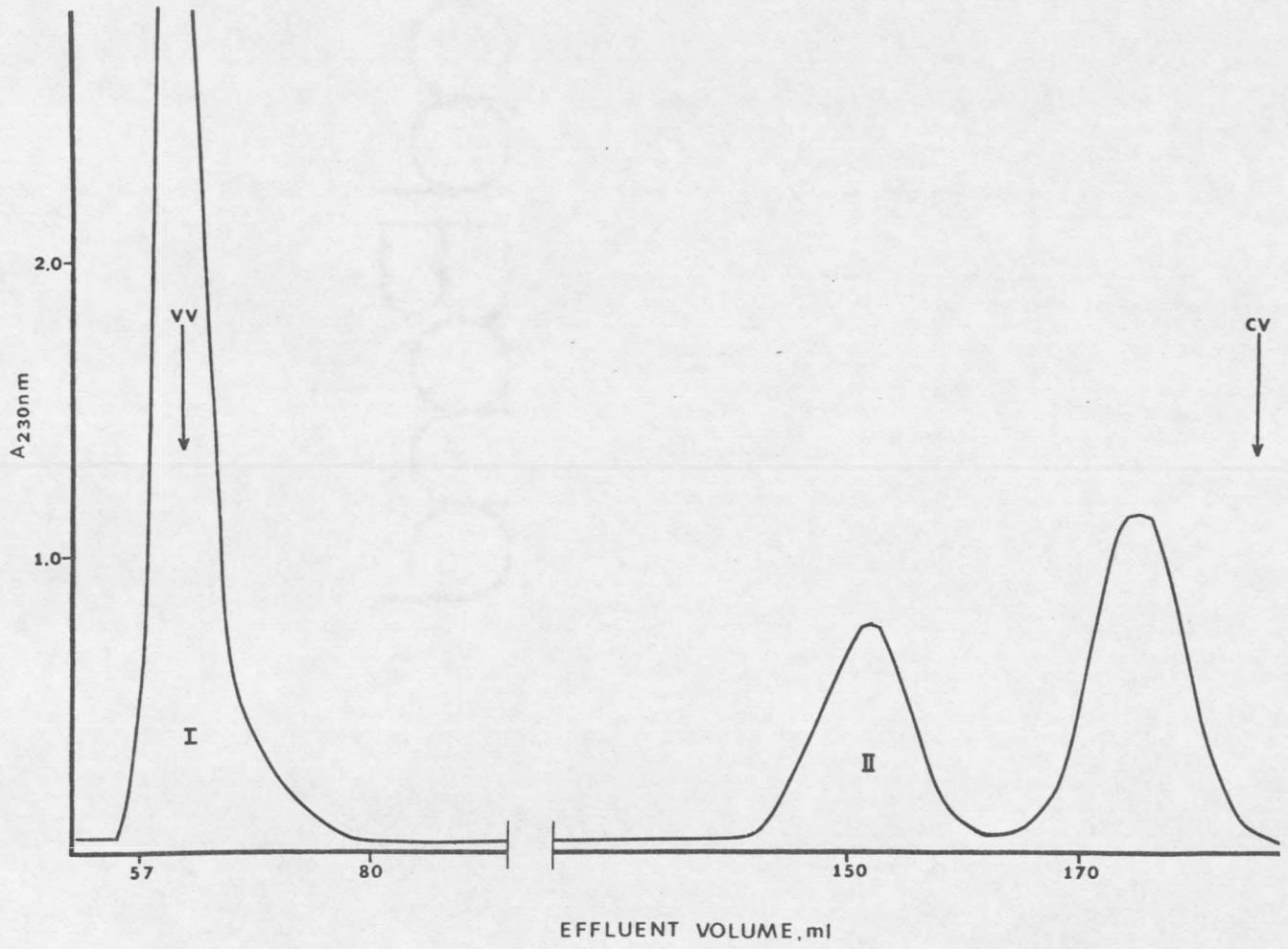


Figure 18. Gel filtration of SL cleaved at cysteine on Sephadex G50

SDS electrophoresis (not shown). A smaller, later peak was observed in front of the inclusion volume and corresponded to the smaller fragment released by cleavage at cysteine. The gel filtration was performed in SDS which could not be successfully removed from the small fragment. Consequently, no reproducible amino acid analysis was obtained for the peptide contained in this fraction.

A second cleavage at cysteine was performed on succinylated SL. Succinylation of the lectin prevented subsequent insolubility of the fragments produced by cysteine cleavage. The succinylated protein was reacted with NTCB labeled with ^{14}C in its cyano group in order later to detect the fragment containing the cysteine derivative. The yield of incorporation of the isotope into the protein was 74%. The succinylated, cyanylated SL was cleaved by exposure to pH 9.0. Half of the reaction mixture was gel filtered on Sephadex G50, as shown in Figure 19. The only peaks visible in the chromatogram were the void volume and salt peaks. All the radioactivity eluted at the void volume. The other half of the cleaved SL was gel filtered in 3M guanidine HCl as shown in Figure 20. The elution was monitored at 280 nm because of strong background absorbtion at 230 nm of the guanidine. Only void volume and salt peaks were visible in the elution profile. All the radioactivity was associated with the void volume peak. The same results were obtained when cleaved SL was

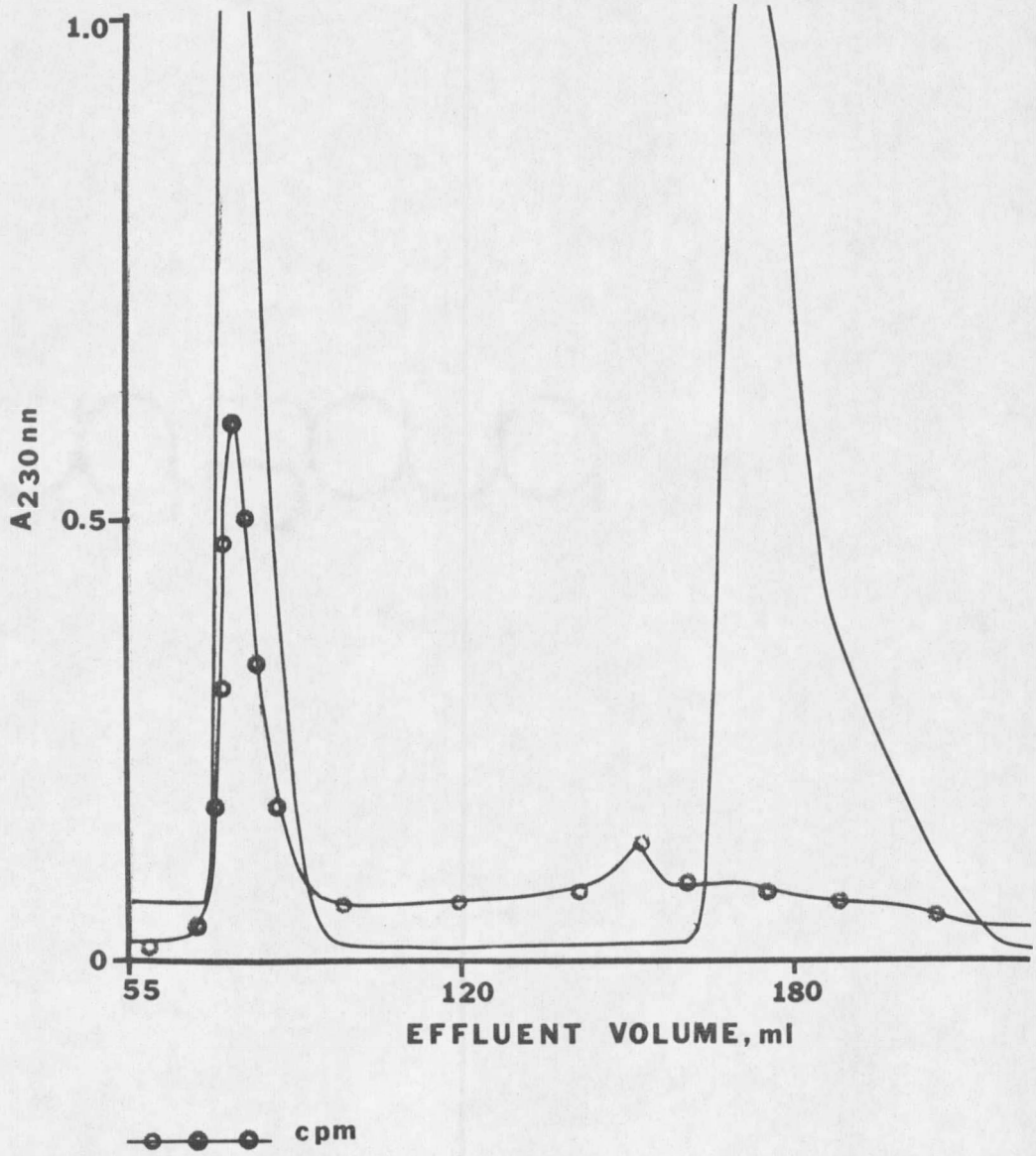


Figure 19. Gel filtration of succinylated SL cleaved at cysteine on Sephadex G25 in 0.02 M NH_4HCO_3, pH 6.3

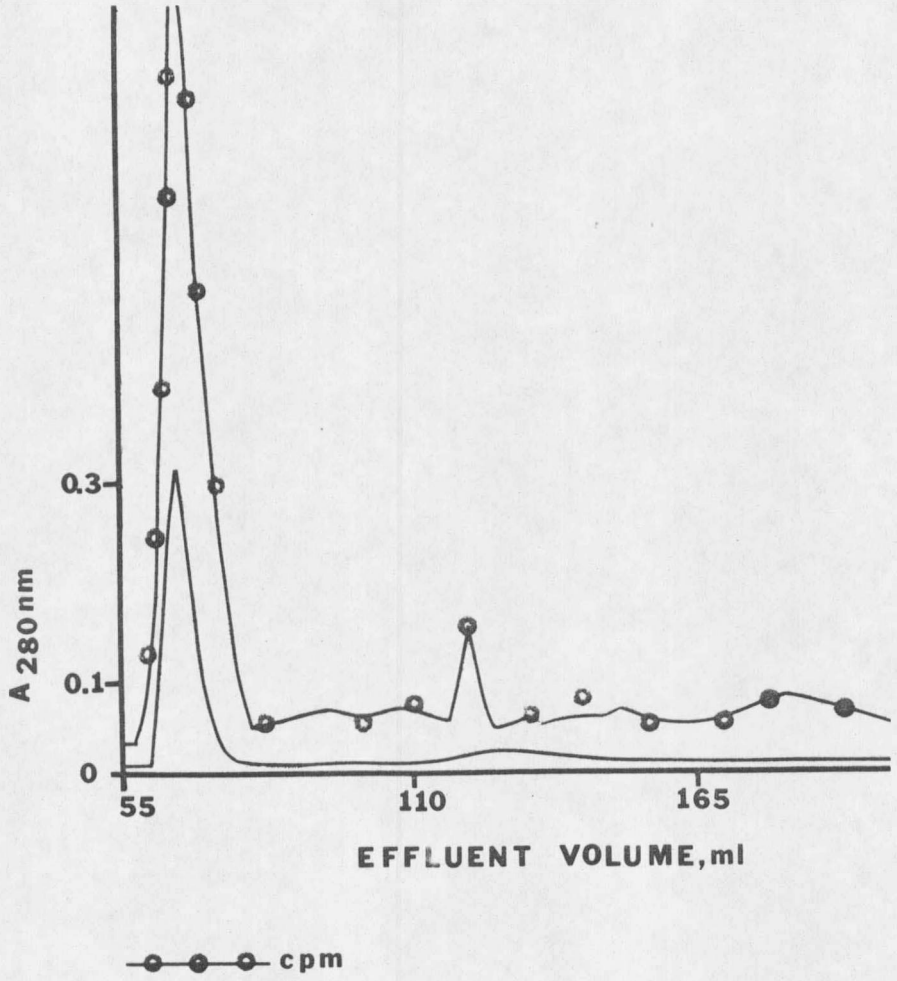
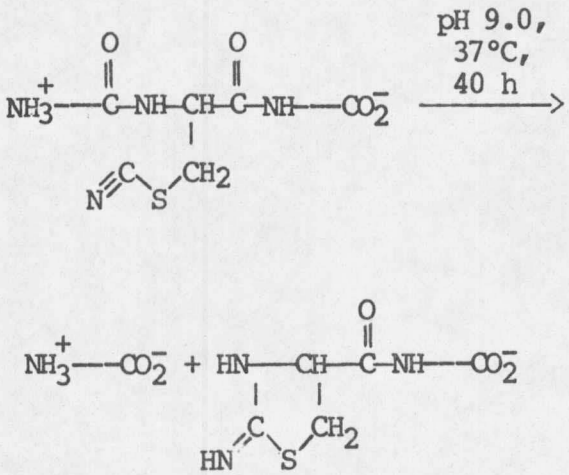


Figure 20. Gel filtration of succinylated SL cleaved at cysteine on Sephadex G25 in 3 M guanidinium chloride, pH 7.0

incubated overnight with beta-mercaptoethanol prior to the gel filtration in 3 M guanidine HCl; i.e., all the radioactivity introduced by ^{14}C-NTCB was recovered at the void volume of the column.

The reagent NTCB cyanylates specifically the sulfhydryl groups of proteins (Stark, 1977). Exposure at high pH of the cyanylated protein results in cleavage at the amino group of cyanylated cysteine by the following reaction (Jacobson et al., 1973):



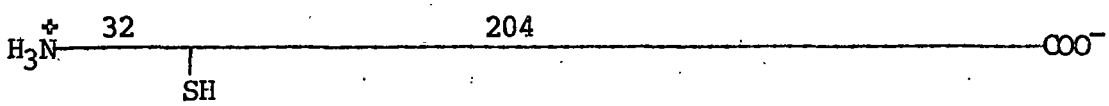
The first set of experiments presented above demonstrated that cleavage at the one cysteine of SL liberated two fragments. One fragment had an approximate MW of 23,400. The other fragment probably had a 3,500 MW in order to account for the whole protein

molecule (26,900 MW according to SDS electrophoresis). The small fragment was not visible on SDS gel because of its small size; it diffused out of the gel during the staining-destaining procedures.

Evidence for the small fragment was obtained by gel filtration of the cleavage products. These experiments demonstrated that the cysteine residue of SL was located approximately 30 amino acids from one terminus of the protein; either the amino or the carboxyl terminus. In the second set of experiments reported above, SL was cyanylated with ^{14}C-NTCB and then cleaved at cysteine. Gel filtration of the cleavage products in absence of denaturant did not allow resolution of the two fragments which probably aggregated to each other. Gel filtration of the cleavage products in the presence of 3 M guanidine HCl did not allow detection of the small fragment. The elution from the gel filtration column was monitored at 280 nm. The small fragment did not absorb at this wavelength and could not be detected. However, in the denaturing conditions of the chromatography, the fragments were probably separated. The recovery of the radioactivity introduced by ^{14}C-NTCB at the void volume of the gel filtration column suggested that the large fragment contained the cysteine derivative. Gel filtration in the presence of 3 M guanidine HCl of the cleavage products preincubated with beta-mercaptoethanol also resulted in an elution profile where all

the radioactivity was associated with the void volume. The reducing agent scavenges the cyano groups from the uncleaved cyanylated protein molecules (Jacobson, 1973). All the radioactivity was therefore associated with the large cleavage product as opposed to uncleaved cyanylated protein molecules. This finding further indicated that the large peptide contained the cysteine derivative. It was therefore derived from the carboxyl terminal portion of SL.

The approximate location of the cysteine residue of SL deduced from these experiments is schematized in the diagram below:



This assignment was later confirmed by sequence analysis which revealed that the cysteine residue was the thirty third residue of the protein. Sequence analysis of SL was not performed on the products of the cleavage at cysteine for two reasons: (1) there was no satisfactory method for removing the iminothiazolidine group from the amino terminus of the large fragment. Automated Edman degradation therefore could not be performed directly on this peptide; (2) the small fragment could be and was sequenced as the amino terminal region of the intact protein thereby rendering the small fragment unnecessary for structure determination.

Trypsin Cleavage at Arginine

The production and purification of tryptic peptides from SL are summarized in the flow chart of Figure 21. The lectin's cysteine residue was carboxymethylated with [^{14}C-1] iodoacetic acid. Incorporation of radioactivity was 0.95 mole ^{14}C iodoacetic acid per mole lectin monomer. Amino acid analysis of the carboxymethylated protein after acid hydrolysis confirmed the presence of 0.9 residue carboxymethylcysteine per monomer. No other residue was modified to any detectable extent by the alkylation reaction. The carboxymethylated protein was then succinylated twice to exhaustively derivatize the lysine amino groups and to protect them from subsequent trypsin cleavage. Carboxymethylated, succinylated SL was digested with trypsin under the conditions described in the Experimental Procedures.

The amount of base consumed during the reaction indicated that the trypsin digestion was complete. An aliquot of the digest was submitted to SDS electrophoresis. Nothing was visible on the gels indicating the absence of intact protein (partially cleaved) or large fragments. The peptides presumably diffused out of the gel during the staining-destaining procedures. An aliquot of the tryptic digest was treated with carboxypeptidase B and analyzed with the amino acid analyzer. The enzyme releases carboxyl terminal arginine

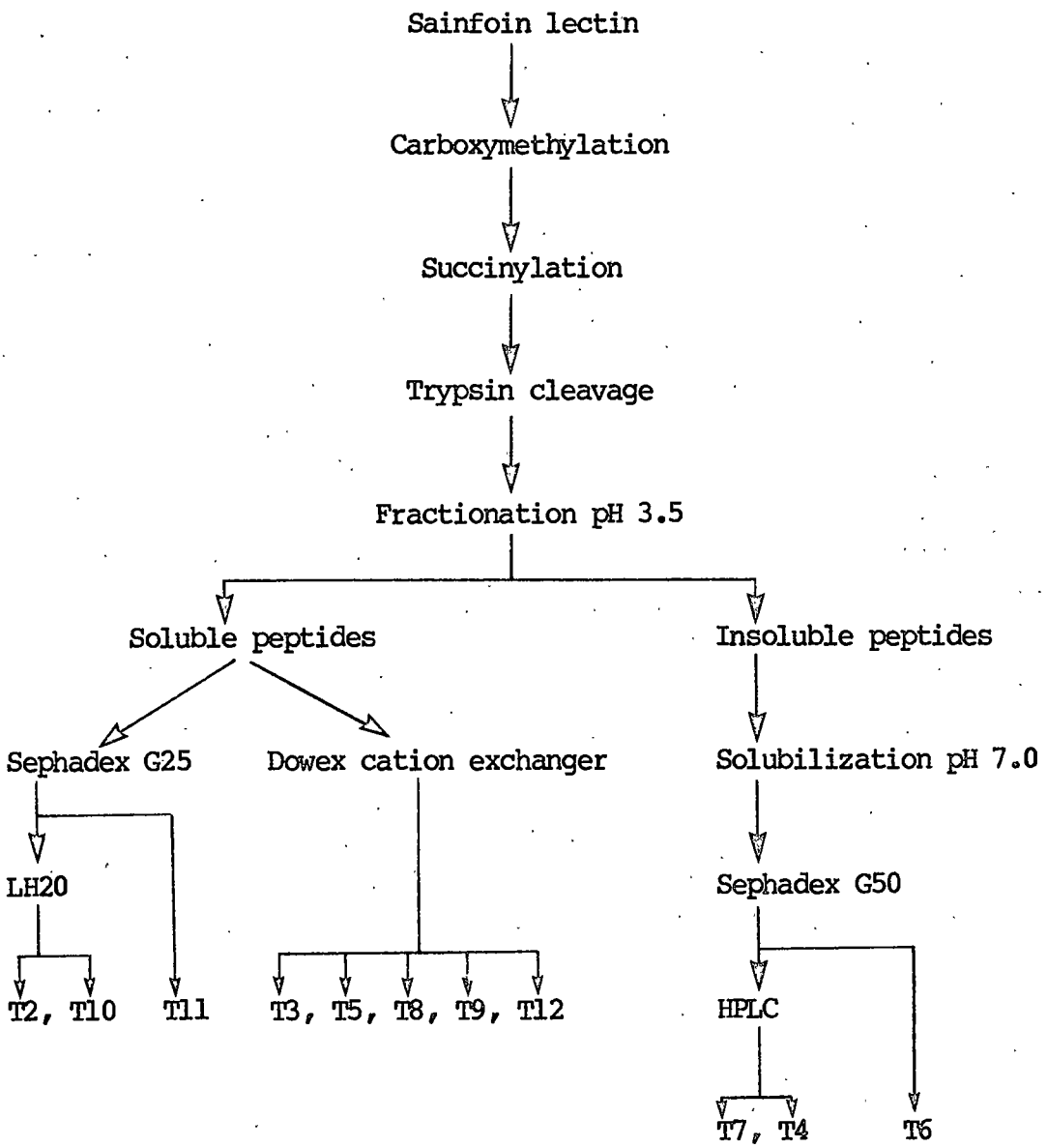


Figure 21. Production and purification of sainfoin lectin tryptic peptides

and lysine. Carboxypeptidase B liberated 9.7 arginine and 0.27 lysine per lectin monomer. Trypsin had therefore cleaved an average of about 10 out of the 12 arginyl peptide bonds of SL.

All the succinylated trypsin peptides were soluble at pH 8.0. Solubility of the peptides varied with pH as illustrated in Figure 22. All the peptides were soluble above pH 5.5. Precipitation increased rapidly between pH 5.5 and 4.3, stabilized between pH 4.3 and 2.0, and increased again below pH 2.0. The tryptic digest was adjusted to pH 3.5 and the soluble and insoluble fractions were separated by centrifugation. Forty percent of the absorbtion at 280 nm of the tryptic digest before the fractionation was associated with the supernatant at pH 3.5, indicating that the soluble peptides contained approximately two of the five tryptophans present in SL. Ninety-five percent of the radioactivity incorporated into SL during carboxymethylation was associated with the pH 3.5 soluble peptides indicating inclusion of the carboxymethyl cysteine peptide. Carbohydrate was detected in the pH 3.5 insoluble peptides showing that they included the peptide bearing the carbohydrate chain of SL.

The procedure worked out for trypsin digestion of SL described above gave the following advantages. Carboxymethylation of the one cysteine residue protected it from possible side reactions and associated poor recovery of the cysteine containing tryptic peptide.

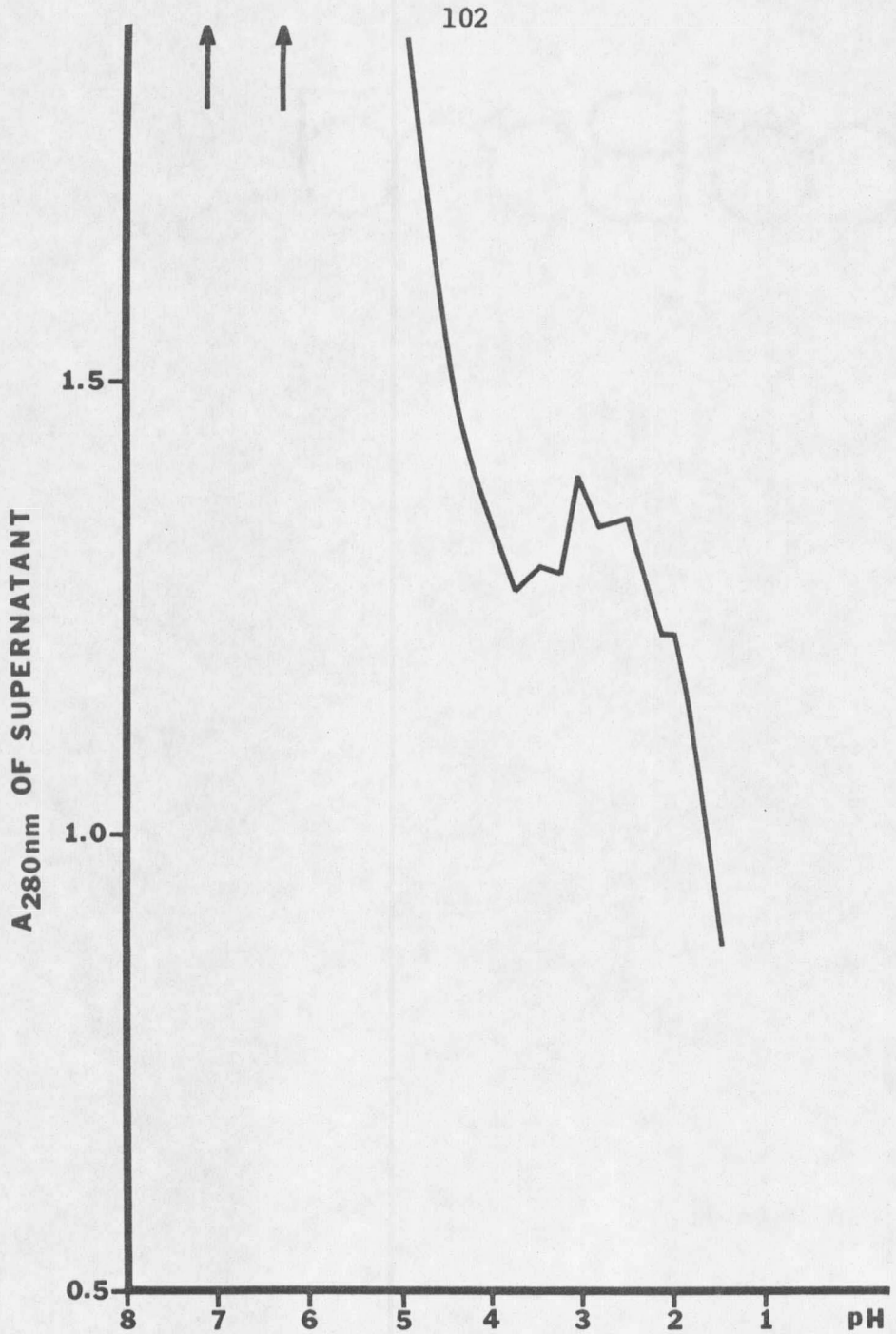


Figure 22. Solubility of SL tryptic peptides as a function of pH

Additionally, modification of the thiol group appeared to sensitize SL to trypsin digestion. Carboxymethylation with [^{14}C-1]-iodoacetic acid introduced a label in the cysteine residue and permitted later identification of the cysteine containing peptide. Succinylation served two purposes: it restricted trypsin cleavage to arginine residues and enhanced the solubility of the resulting peptide fragments and thereby assisted their purification.

Sainfoin lectin monomer contains 12 arginines and 11 lysines (Hapner & Robbins, 1979). If all tryptic sites were cleaved, 24 peptides would result. Succinylation of the substrate restricted trypsin cleavage and produced a lesser number of peptides that were later easier to align. The enhanced solubility of the peptides produced from succinylated SL was responsible for the large choice of purification techniques that could be utilized to fractionate these peptides. Trypsin digestion of carboxymethylated succinylated SL liberated 10 of the 12 arginines and 0.27 lysine residue. Later sequence analysis demonstrated that SL possessed one arginyl-prolyl bond. This type of bond is trypsin resistant (Smyth, 1967). Consequently a maximum of 11 bonds could have been cleaved by trypsin. The yield of trypsin cleavage was therefore 90%. Liberation of 0.27 lysine by carboxypeptidase B indicated the presence of some unsuccinylated lysine residues that exhibited

trypsin sensitivity. Any peptide contaminants derived from trypsin cleavage at these sites were removed during purification. The tryptic peptides were initially fractionated at pH 3.5 on the basis of their solubility. The procedure separated the large (insoluble) tryptic peptides from the small (soluble) ones. This facile fractionation simplified the subsequent purification procedures described below.

Purification of the pH 3.5 Soluble SL Tryptic Peptides

Purification of soluble tryptic peptides is illustrated in the flow chart of Figure 21. The peptides were fractionated by two alternative techniques: gel filtration or ion exchange chromatography. Figure 23 shows the gel filtration of the pH 3.5 soluble peptides on Sephadex G25 in 0.02 M ammonium bicarbonate, pH 6.5. The peptides contained in each 230 nm absorbtion peak are indicated on the elution profile. They were numbered according to their position in the lectin polypeptide chain. The last peak of the elution profile contained only salt. Peptide T11 was pure enough to be utilized for sequence analysis and was submitted to Edman degradation. All the other peptides isolated from the gel filtration needed repurification. The peak containing T10 and T8

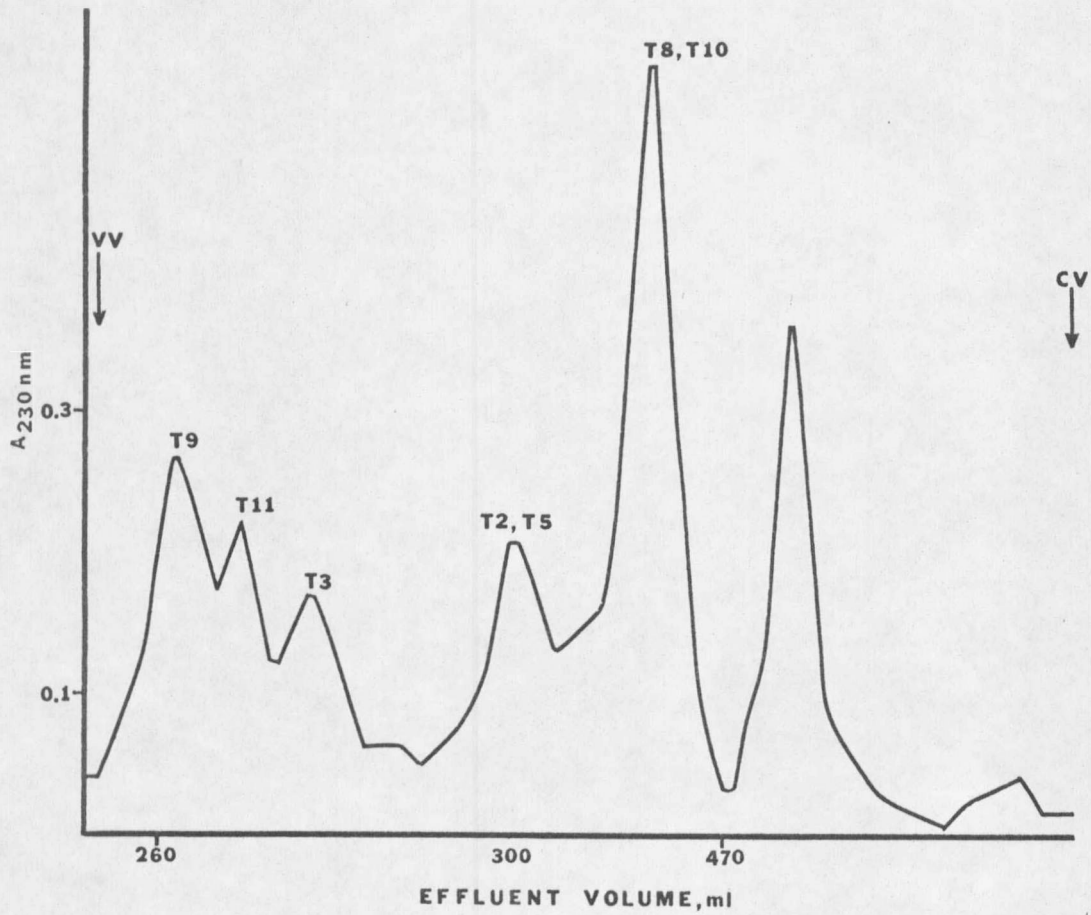


Figure 23. Gel filtration of pH 3.5 soluble SL tryptic peptides on Sephadex G25, in 0.02 M NH_4HCO_3

were concentrated by rotary evaporation and gel filtered on LH20 as shown in Figure 24. Peptide 10 was purified to homogeneity by this technique and was submitted to sequence analysis. The very large 230 nm absorbtion peak eluted at the inclusion volume of the column contained buffer salts. Peptide T8 was not recovered from this chromatography, possibly due to its low absorbancy or precipitation in the column.

The radioactive carboxymethylcysteine-containing peptide, T2, was isolated from the pH 3.5 soluble peptides by a combination of two gel filtration runs. In the first gel filtration, the soluble peptides were run on Sephadex G25 in 0.01 M acetic acid, as shown in Figure 25. The fraction containing the radioactivity was then repurified on LH20, as shown in Figure 26. Peptide T2 thereby purified by this procedure was suitable for sequence analysis.

The second method of purification of the soluble peptides was ion exchange chromatography. Figure 27 shows the fractionation of the peptides on DC-X8-11. Seven major peaks of ninhydrin color were visible on the chromatogram. Each peak contained a single peptide indicated on the elution profile. The minor peaks corresponded to traces of peptidic contaminants. The soluble peptides were completely resolved by this technique. Peptides T3, T5, T8, T9 and T12 thereby purified were utilized for subsequent sequence analysis.

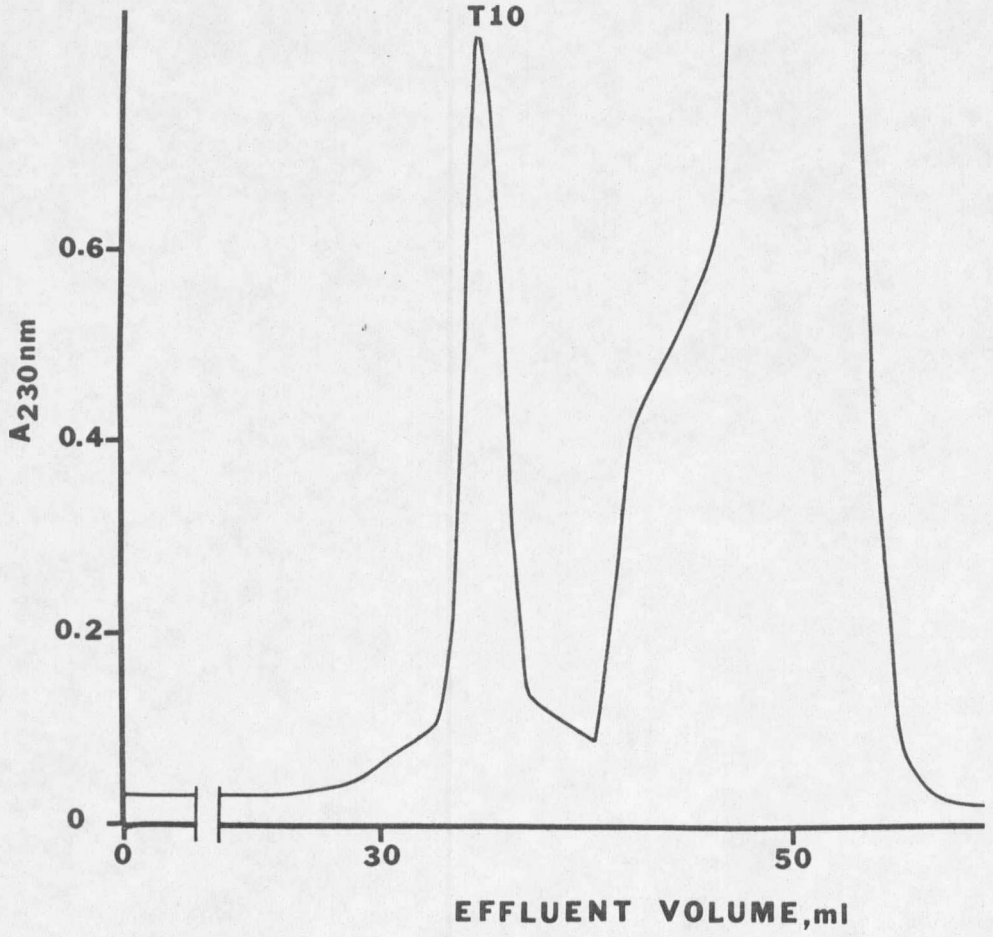


Figure 24. Repurification of peptide T10 of Figure 23 on Sephadex LH20

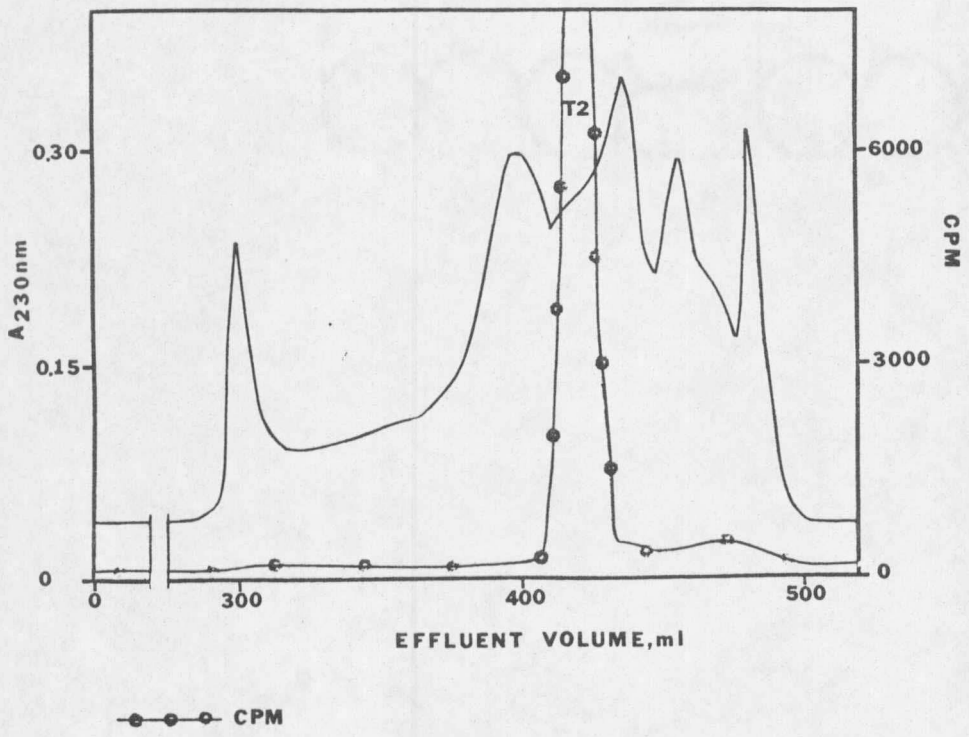


Figure 25. Gel filtration of pH 3.5 soluble SL tryptic peptides on Sephadex G25 in 0.01 M acetic acid

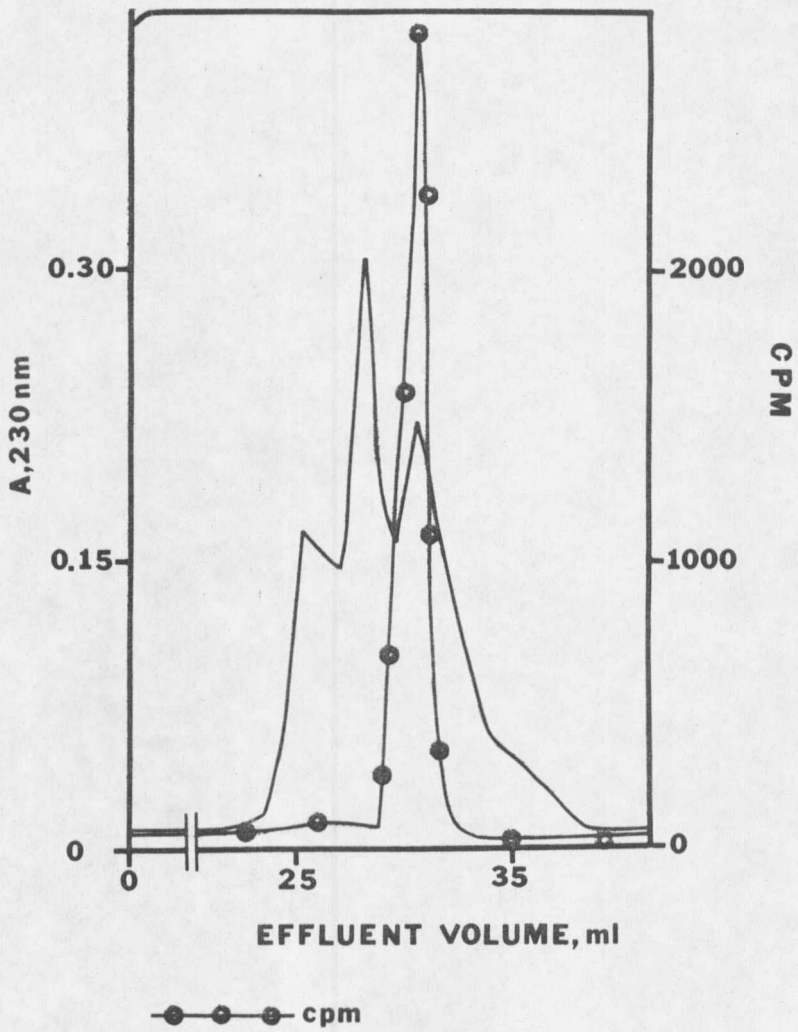


Figure 26. Repurification of peptide T2 from Figure 25 on Sephadex LH20

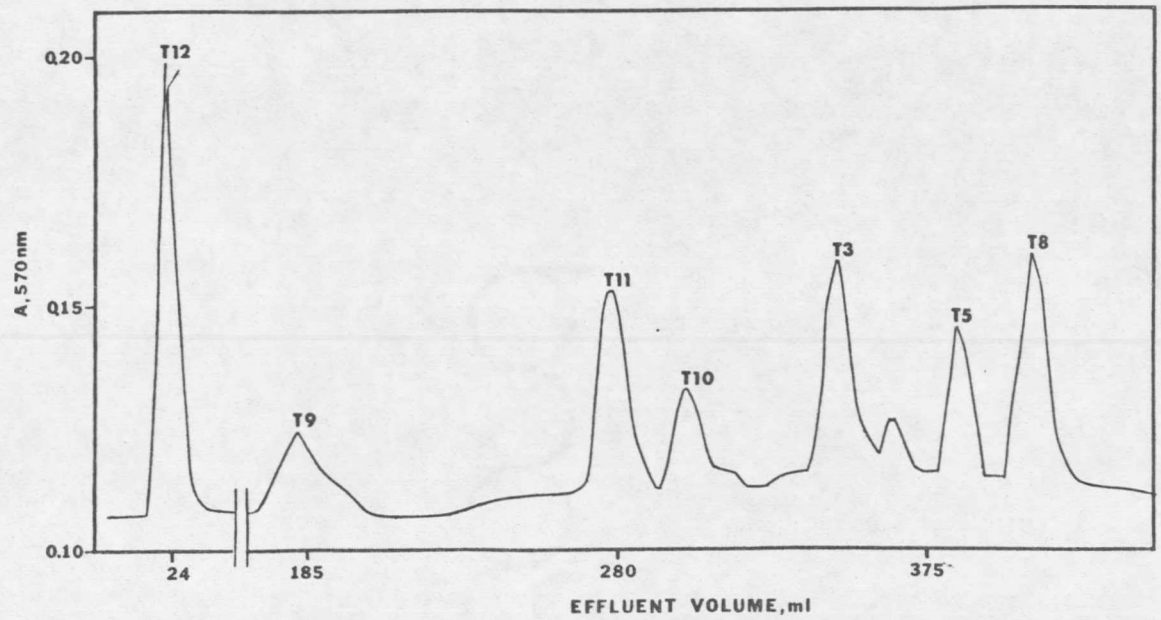


Figure 27. Cation exchange chromatography of pH 3.5 soluble SL tryptic peptides

Purification of the pH 3.5 Insoluble SL Tryptic Peptides

The pH 3.5 insoluble peptides were solubilized by raising the pH to 8.0. They were then gel filtered on Sephadex G50 as shown in Figure 28. The void volume peak contained a small core of high MW material. Fraction I contained peptide T6. Carbohydrate analysis located carbohydrate in fraction I, indicating that T6 was a glycopeptide. The purity of T6 varied with the completeness of SL trypsin digestion. In the gel filtration presented in Figure 28, T6 was recovered in 50% yield in homogeneous form. It was then lyophilized, dissolved in 0.1% SDS and submitted to automated Edman degradation. Fractions II, III and IV were poorly resolved from each other. Fractions II and IV contained peptides T7 and T4, respectively. The large 230 nm absorption associated with these fractions was due to the presence of Trp in T7 and T4. These peptides were not suitable as such for sequence analysis because of the presence of contaminants. They were repurified by reversed-phase HPLC. Figure 29 illustrates the repurification of T7. The peptide was recovered homogeneous in 8% yield, from the 280 nm absorption peak indicated in the elution profile. Several major absorption peaks were observed in the chromatogram. They corresponded to T7 and slightly modified counterparts of T7, i.e., T7 with oxidized tyrosine residues and T7 succinylated at some of

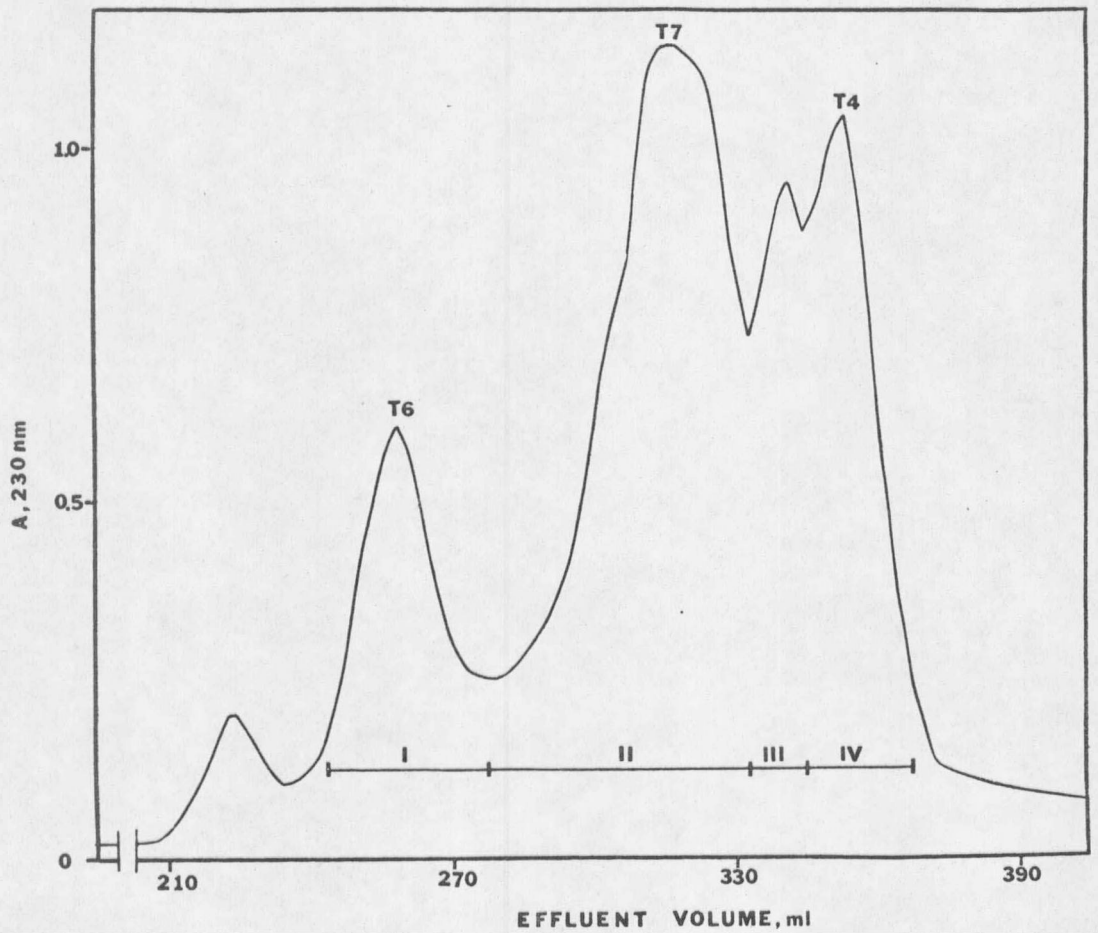


Figure 28. Gel filtration of pH 3.5 insoluble SL tryptic peptides on Sephadex G50

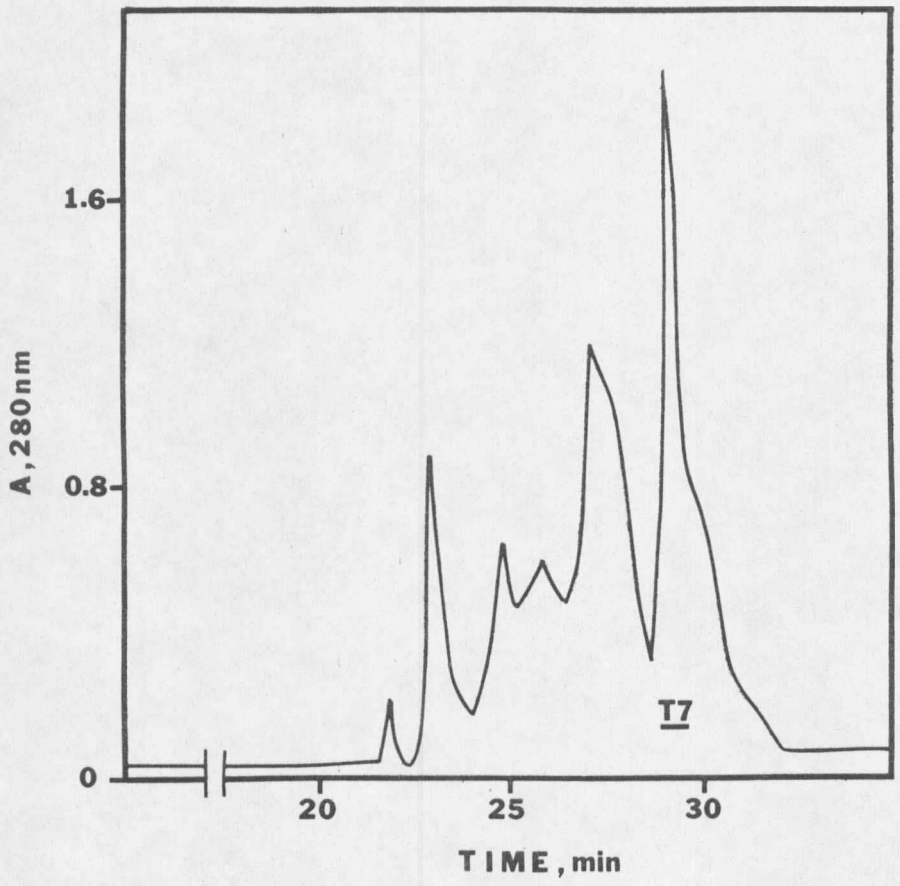


Figure 29. Repurification of peptide T7 by reversed-phase HPLC

its hydroxyamino acids. The presence of modified versions of T7 was responsible for the low yield of recovery of T7. Peptide T4 contained in the gel filtration fraction IV (Figure 28) could not be concentrated by rotary evaporation as it precipitated in the presence of high salt concentrations. It was concentrated by ultrafiltration and submitted to HPLC as shown in Figure 30. The major 280 nm absorption peak of the elution profile contained T4. The peptide was recovered homogeneous in 40% yield and was utilized for sequence analysis.

Peptide T6 was the largest SL tryptic peptide (46 amino acids). Its purification, presented above, was extremely simple, i.e. a single gel filtration step, and its detection was facilitated by the presence of the carbohydrate chain attached to the peptide. The ease of recovery of T6, together with the fact that it was almost entirely sequenced by automated Edman degradation (see below) rapidly provided the amino acid sequence of a major portion of SL.

The purity of T6 recovered from the gel filtration contrasted with the complexity of the multipeak region containing T7 and T4 (Figure 28). Peptides T7 and T4 were similar in size (40 and 30 amino acids, respectively) and therefore were not well resolved. The presence of two tryptophans in T7 as opposed to one in T4 may have resulted in adsorption of T7 to the sephadex matrix with

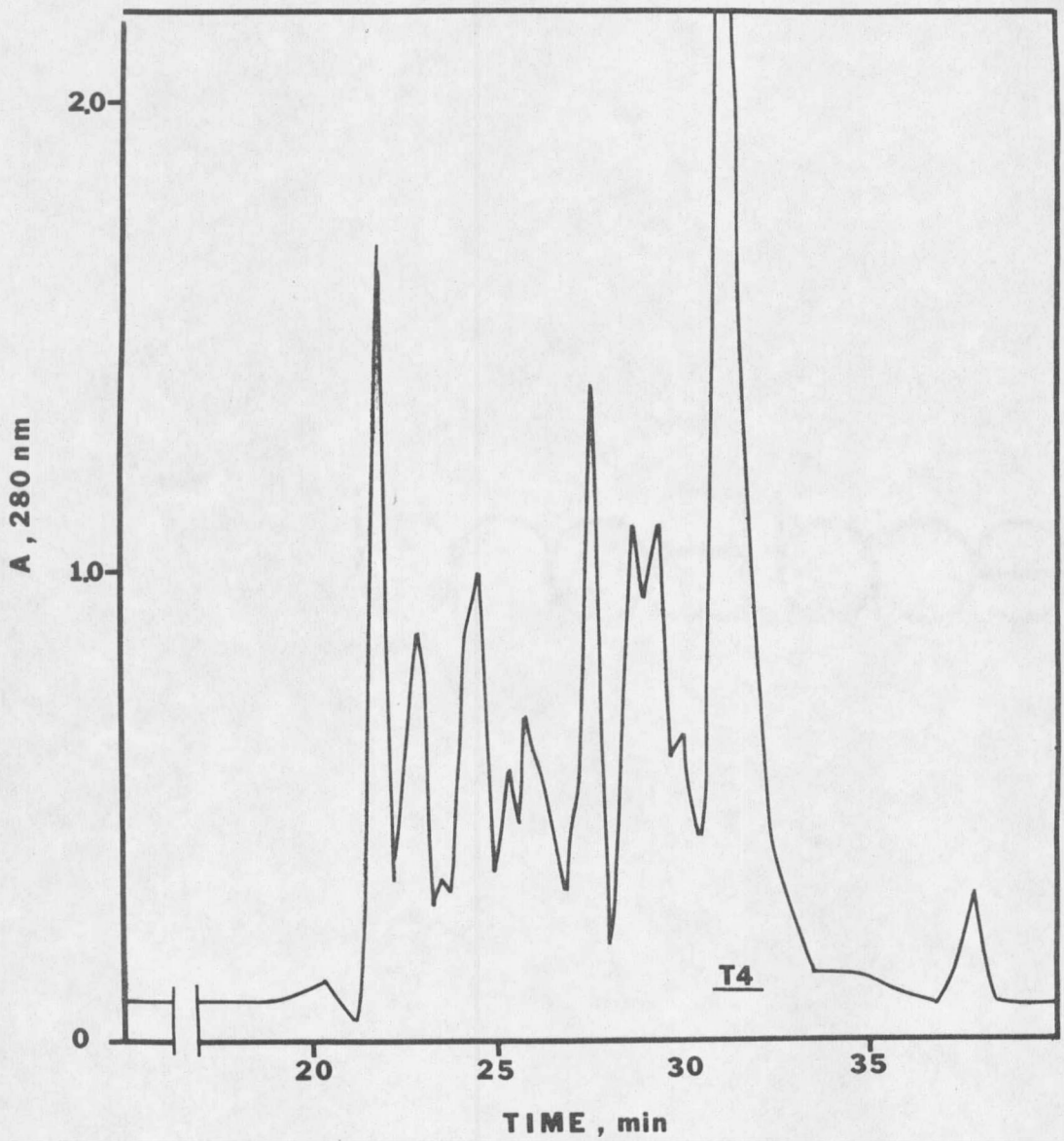
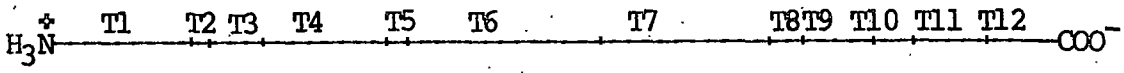


Figure 30. Repurification of peptide T4 by reversed-phase HPLC

consequent retardation of T7 and poor resolution of the peptides. The additional peptides contaminating T4 and T7 were presumably derived from unspecific cleavages of SL by trypsin. They also probably included succinylated peptide T1 (32 amino acids), derived from the amino terminus of SL, whose sequence was established by Edman degradation of intact SL (see below).

Analyses of SL Tryptic Peptides

All the tryptic peptides were sequenced by automated Edman degradation. The diagram below illustrates the position of the tryptic peptides in the SL polypeptide chain and can be referred to for the following descriptions.



Peptide T1. The amino acid sequence of T1 was established through Edman degradation of intact SL. Two hundred nanomoles (5.2 mg) lyophilized SL was dissolved in 10% formic acid and loaded in the spinning cup of the sequenator. Useful data were achieved for 32 cycles of Edman degradation performed on the protein. The derived sequence was: H_2N-A E N T V S F D F S K F L S G Q E N L I L Q G D T V TD D S N R_{32}. (Amino acid symbols are described in the Appendix.)

The assignments of the underlined residues were tentative because of low recovery of the PTH amino acids at these positions. No PTH amino acid was identified at position 28. This position was later identified as D by sequence analysis of a thermolytic peptide derived from this region of the protein. The same peptide also confirmed the assignments of residues 27, 30 and 31.

Peptide T1 was not recovered during the purification procedures of SL tryptic peptides. It probably behaved as a contaminant of the T4 and T7 region due to its similar size. The alpha-amino group of T1 was succinylated and would not have been useful for Edman degradation in any case. Lack of positive identification of all positions was due to inherent limitations in sequence analysis on very large peptides. The reagents used during the degradation cumulatively result in nonspecific cleavage of the protein and this results in appearance of additional amino acid residues which obscure the authentic sequence. The problem is normally overcome, as it was here, by isolation of an additional peptide fragment containing the relevant sequence. The alpha-amino group of SL was not blocked in these studies, contrary to the initial observation of Hapner & Robbins (1979).

Peptide T2 (Figures 25 & 26). The amino acid composition of T2 was: 1.0 C, 1.0 B, 1.1 T, 1.8 S, 1.5 V, 2.7 L, 1.0 R. Its amino acid

sequence was: C L V L T R_{38}. The threonine residue assignment was initially tentative. It was later confirmed by sequence analysis of a thermolytic peptide which overlapped T2 and T3.

The composition of T1 showed extraneous amino acids compared to the composition deduced from its sequence analysis. It indicated that the peptide contained some impurities. However, the identification of the PTH amino acids of the authentic sequence was unambiguous indicating that the amino acid sequence determination was correct.

Peptide T3 (Figure 27). The amino acid composition of T3 was: 2.7 B, 1.0 S, 1.9 Z, 1.0 P, 1.9 G, 1.5 V, 0.5 I, 0.3 L, 2.0 R. Its amino acid sequence was: E N N G R P V Q D S $\frac{V}{I}$ G R_{51}.

Peptide T3 contained one Arg-Pro bond which was not cleaved by trypsin; therefore, its composition contained two arginine residues. Valine and isoleucine were both present in equal amounts at the eleventh degradation cycle. The presence of two amino acids at the same position showed the existence of two allelomorphic forms of SL. It should be noted that allelomorphic replacements of this type could not contribute to the ionic heterogeneity observed earlier in SL due to the nonionic nature of the amino acids involved.

Peptide T4 (Figures 28 & 30). Amino acid analysis of T4 yielded the following composition: 2.0 B, 2.6 T, 1.7 S, 4.0 Z, 0.9 P, 1.1 A, 1.1 V, 2.9 I, 1.9 L, 1.5 Y, 3.5 F, 1.9 K, 0.7 H, 1.0 R. The amino acid sequence of the purified peptide was: V L Y Q T P I H L W D K Q I D K E A S F E T S F T F F I Y R_{81}.

Peptide T5 (Figure 27). Amino acid composition of T5 was: 1.7 B, 0.8 Z, 1.1 I, 1.0 R. Its amino acid sequence was: E N I N R_{86}.

Peptide T6 (Figure 28). Amino acid analysis of T6 yielded 6.0 B, 4.5 T, 3.5 S, 4.0 Z, 1.8 P, 6.4 G, 3.0 A, 2.6 V, 1.9 I, 2.2 L, 0.8 Y, 4.6 F, 1.9 K, 1.0 R, and 1.2 glucosamine. The sequence of the peptide was G G D G I T F F L A P T D T Q P K S G G G Y L G I F K D A E S N E T V V A V E F D T F S N R_{132}. No PTH amino acid could be



detected after the Asn residue at the forty-fifth cycle. Consequently, the carboxyl terminal arginine was identified by carboxypeptidase Y analysis of peptide T6.

Assignment of Carbohydrate Position. During the sequence analysis of peptide T6 reported above, no PTH amino acid was detected at position 118. The composition of the peptide indicated that position 118 contained Asx. The PTH derivative of asparagine

linked to a carbohydrate chain is not recovered from the sequenator spinning cup because of its polarity. Position 118 therefore contained the Asn to which SL carbohydrate chain is attached. Namen & Hapner (1979) showed the carbohydrate chain to be attached to Asn in the sequence ---SNET---. Asparagine 118 was found in the same sequence, indicating that it was the site of SL glycosylation.

Other Characteristics of SL Tryptic Glycopeptide. Peptide T6 was the largest SL tryptic peptide (46 residues). It contained no tryptophan and nearly half of SL glycine residues. Dubois analysis of the peptide confirmed that it contained the single carbohydrate chain in SL. End group analysis of T6 by the dansyl chloride method and the exhaustive alkylation method indicated glycine to be at the amino terminus of the peptide, in agreement with the sequence analysis. Polyacrylamide gel electrophoresis at pH 9.5 of T6 showed two bands running close to each other and of similar intensities (not shown). This charge heterogeneity was in apparent contradiction to T6 unique amino acid sequence and was probably of the same type as SL charge heterogeneity, i.e. due to amide differences.

Other methods were attempted for purification of peptide T6. These included affinity chromatography, based on the knowledge of SL carbohydrate chain structure (Namen & Hapner, 1979). Peptide T6 did

not bind ConA-Sepharose prior to or after beta xylosidase treatment and it did not interact with WGA-Sepharose. The peptide was also not bound by PBA-X30, a boronate affinity ligand which interacts with cis hydroxyl groups. The successful structure determination of T6 derived directly from gel filtration fortunately made the repurification attempts non-essential.

Peptide T7 (Figures 28 & 29). Amino acid composition of T7 was 7.0 B, 5.0 T, 4.2 S, 1.0 Z, 1.8 P, 2.0 G, 2.8 A, 4.0 V, 3.8 I, 1.4 L, 2.1 Y, 2.0 F, 3.0 K, 1.5 H, 1.0 R. Amino acid sequence of T7 was W D P A N S H I G I N V N S V K S K I T T P W G L K N D Y F T V T I T Y D A T R_{172}. The carboxyterminal arginine was not identified during the sequence analysis. No PTH amino acid was detected at this last cycle because the arginine was washed out of the sequenator cup. The arginine residue was identified by digestion of peptide T7 with carboxypeptidase Y and amino acid analysis of the released residues.

Peptide T7 was the second largest tryptic peptide, 40 amino acids long. It contained two tryptophans which were responsible for its high 280 nm absorption. No end group was detected by dansyl chloride analysis, in agreement with the sequence analysis (dansyl-Trp is destroyed during the analysis).

Peptide T8 (Figure 27). Amino acid analysis of T8 yielded the following composition: 3.4 S, 1.0 V, 1.3 L, 0.9 Y, 0.9 F, 1.0 R. Sequence of the amino acids was S L S V S S F Y R_{181}.

Peptide T9 (Figure 27). Amino acid composition of the peptide was 3.0 B, 0.9 T, 1.0 S, 1.0 P, 0.6 G, 0.8 A, 1.5 V, 0.9 I, 1.7 L, 0.8 F, 1.0 K, 0.6 H, 1.0 R. Sequence of the amino acids was N K P D D I F T V K A S V H L R_{197}.

The amino acid composition deduced from the sequence was not in perfect agreement with the composition of the peptide established by amino acid analysis. The sequence showed no Gly, one Leu and two Lys, while the composition indicated 0.6 Gly, 1.7 Leu and 1.0 Lys. The discrepancy in lysine content is best explained by the fact that the lysines were succinylated; succinylated lysines are not quantitatively recovered as lysines from acid hydrolysis. The identification of all the PTH amino acids during the sequence analysis of T9 was unambiguous. The excess Gly and Leu found in the peptide composition therefore belonged to contaminants.

Peptide T10 (Figures 23 & 24). Amino acid composition of T10 was 1.2 B, 1.3 Z, 1.0 P. 0.3 G, 1.1 A, 0.8 V, 1.5 L, 1.0 R. Absorption spectrum of the peptide showed the presence of tryptophan. Edman

degradation of the peptide gave the following amino acid sequence: D A L P Q W V R_{205}.

Peptide T11 (Figure 23). Composition of T11 was 1.1 B, 1.0 T, 1.0 S, 2.0 Z, 1.9 G, 2.0 A, 0.9 V, 1.1 I, 2.4 L, 0.9 H, 1.0 R. Its amino acid sequence was I G L S A A T G D L V E Q H R_{220}.

Peptide T12 (Figure 27). Amino acid composition of the peptide was 1.0 B, 0.8 T, 3.2 S, 0.6 P, 0.5 G, 0.6 V, 2.3 L, 0.6 Y, 0.7 F, 0.5 K. Sequence of the amino acids was L Y S W S F K S V L P L D S S T_{236}. The assignment of threonine and serine residues near the carboxyl terminus during the automated sequence analysis was only tentative. These residues were confirmed by carboxypeptidase Y digestion of T12.

The tryptic peptides produced from SL resulted from the cleavage at the arginine residues of the protein. The absence of arginine at the carboxyl terminus of T12 indicated that this peptide was derived from the carboxyl terminus region of SL.

Cleavage of SL with Thermolysin

Production and purification of thermolytic peptides from SL are illustrated by the flow chart in Figure 31. One μmole (26 mg) SL was digested with thermolysin at 37°C, pH 8, for 1 h. An aliquot of the digest was submitted to SDS electrophoresis. No intact protein

was left and no large peptide could be detected in the SDS gels. Thermolysin digestion of the protein was therefore extensive. The thermolytic digest was gel filtered on Sephadex G25 as shown in Figure 32. Nine 230 nm peaks were detected and the fractions were pooled accordingly. Aliquots of each pool were hydrolyzed and screened for their arginine content with the amino acid analyzer. The thermolytic peptides of interest were those that overlapped the tryptic peptides and, consequently, they contained arginine. The arginine contents of the nine pools are shown in the insert of Figure 32. Pools III to VII contained 500 nM or more arginine and were considered to contain overlaps. They were further fractionated by cation exchange chromatography on the Beckman AL5 column of the amino acid analyzer using a four-chamber gradient of pyridine-acetate buffers. Aliquots of the fractions eluted from the ion exchanger were lyophilized and reacted with the arginine reagent phenantrenequinone. Arginine positive regions of the effluent were analyzed with ninhydrin to better localize the peptide peaks. Amino acid analysis was then used to identify arginine peptides that contained overlapping regions of previously sequenced tryptic peptides.

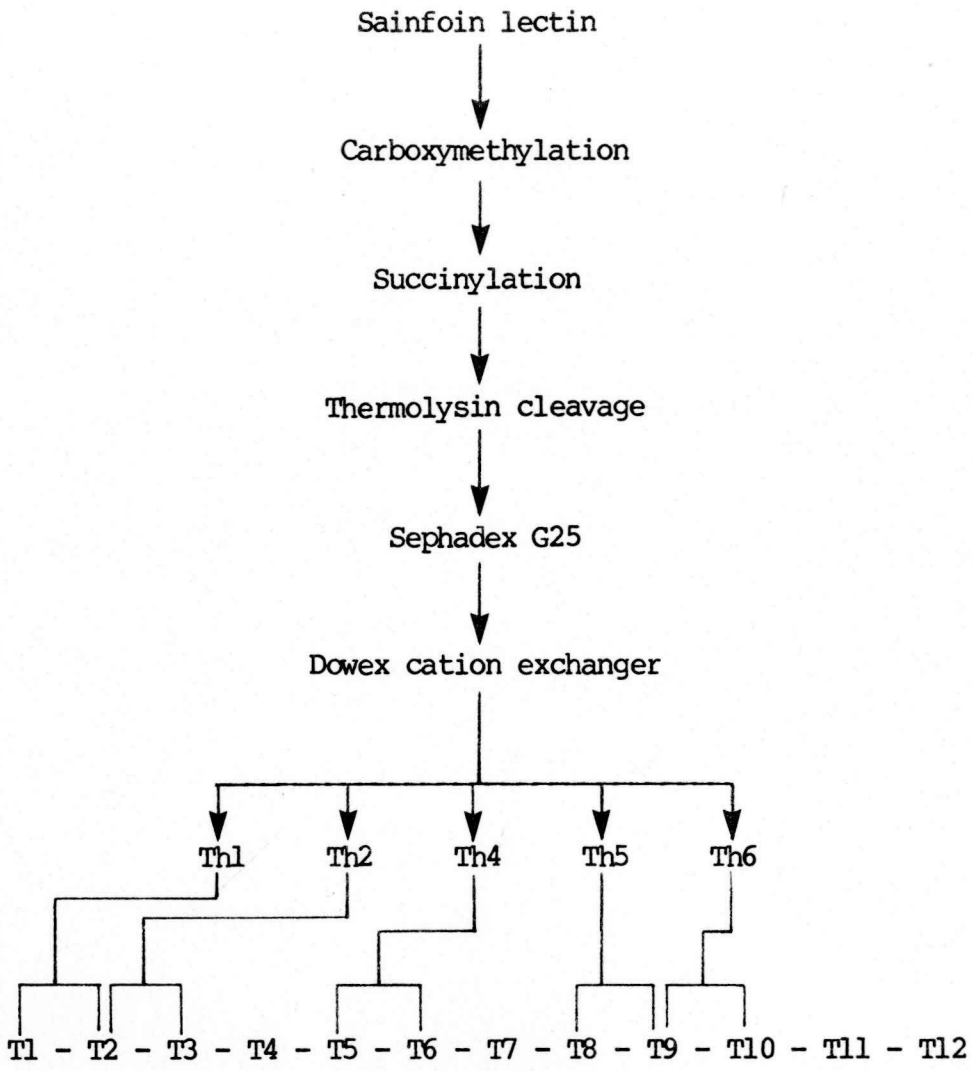
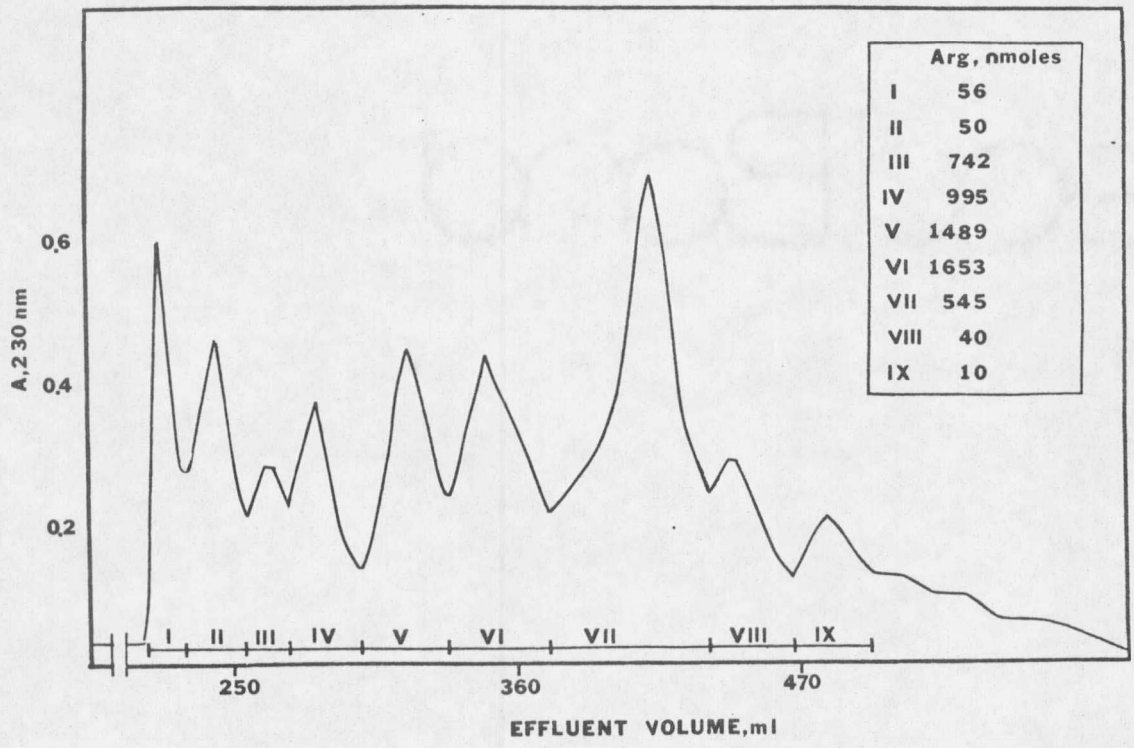


Figure 31. Production and purifiction of thermolytic overlaps



Insert: arginine content of the Sephadex G25 fraction

Figure 32. Gel filtration of SL thermolytic peptides on Sephadex G25

Prior to thermolysin digestion, SL was carboxymethylated and succinylated. Carboxymethylation prevented the possible oxidation of the sulfhydryl group and consequent lower yield of recovery of the cysteine containing thermolytic peptide. The purpose of the succinylation was to enhance solubility of the thermolytic peptides. Carboxymethylated-succinylated SL was readily digested with thermolysin. Thermolysin cleaves principally peptide bonds whose amino group adjoins to Ile, Leu, Phe, or Val (Heinrikson, 1977). Sainfoin lectin has 66 such sites and thermolysin certainly produced a multiplicity of small peptides. No attempt was made in the work presented here to separate and purify all the thermolytic peptides produced from SL. On the contrary, a method was developed to purify only those thermolytic peptides of interest, i.e. those overlapping SL tryptic peptides.

Analysis of Arginine-Containing SL Thermolytic Peptides

Peptide Thl. Peptide Thl was purified from pool III of the gel filtration on Sephadex G25 of SL thermolytic digest (Figure 32) by cation exchange chromatography. Amino acid composition of the peptide was 0.9 C, 3.5 B, 1.9 T, 1.1 S, 1.0 Z, 1.1 G, 0.9 V, 0.9 I, 1.0 L, 1.0 R. Its amino acid sequence was I L Q G D T V T D D S N R C_{33}.

Peptide Th1 overlapped the amino terminal peptide T1 and the cysteine residue of peptide T2 (Figure 31). It therefore positioned T2 on the carboxyl side of T1. Peptide Th1 confirmed the tentative assignments made during sequence analysis of intact SL at cycle number 27, 30 and 31. It also established Asp28 that had not been detected during Edman degradation of SL (see Analysis of SL Tryptic Peptides - Peptide T1).

Peptide Th2. Peptide Th2 was derived from pool II of the Sephadex G25 (Figure 32) and further purified by cation exchange chromatography. Its amino acid composition was 2.9 B, 1.0 T, 1.2 S, 1.8 Z, 0.8 P, 1.2 G, 0.9 V, 1.5 L, 2.0 R. Its amino acid sequence was established to be L V L T R E N N G R P_{44}.

Peptide Th2 overlapped the tryptic peptides T2 and T3 (Figure 31). It confirmed the assignment of threonine at cycle 5 of T2 that was only tentatively established during the sequence analysis of T2 (see Analyses of SL Tryptic Peptides - Peptide T2).

Peptide Th3. Peptide Th3 was purified from pool VII of the Sephadex G25 (Figure 32) by cation exchange chromatography. Its amino acid composition was 1.0 G, 0.5 V, 0.5 I, 1.0 R. The composition of peptide Th3 identified it as the thermolytic tripeptide G $\overset{I}{V}$ R_{51} corresponding to the carboxyl terminus region of T3. Presence of

one-half Val and one-half Ile in the peptide confirmed the existence of the two allelomorphic forms of T3 (see Analyses of Tryptic Peptides - Peptide T3). No sequence analysis was performed on Th3. The arginine residue of the peptide was at the carboxyl terminus. Consequently, Th3 was not an overlapping peptide and its sequence would not have produced any alignment information.

Peptide Th4. Peptide Th4 was purified form pool VI of the Sephadex G25 (Figure 32) by cation exchange chromatography. Its amino acid composition was 2.1 B, 3.2 G, 1.0 I, 1.0 R. Its amino acid sequence was I N R G G D G_{90}.

This peptide overlapped the tryptic peptides T5 and T6 (Figure 31).

Peptide Th5. Peptide Th5 was derived from pool V of Sephadex G25 (Figure 32) and further purified by cation exchange chromatography. Its amino acid composition was 2.9 B, 1.0 T, 1.0 P, 0.7 I, 0.4 Y, 1.6 F, 0.9 K, 1.0 R. Its amino acid sequence was F Y R N K P D D I F T_{189}.

This peptide overlapped tryptic peptides T8 and T9 (Figure 31).

Peptide Th6. Peptide Th6 was purified from pool VII of Sephadex G25 (Figure 32) by cation exchange chromatography. Its amino acid

composition was 1.0 B, 1.0 Z, 1.0 P, 1.0 A, 2.0 L, 1.0 R. Its amino acid sequence was L R D A L P Q W_{203}.

Peptide Th6 overlapped tryptic peptides T9 and T10 (Figure 31).

Peptide Th7. Peptide Th7 was purified from pool VII of the Sephadex G25 of Figure 32 by ion exchange chromatography. Its amino acid composition was 1.3 Z, 0.3 G, 0.7 V, 0.6 L, 0.6 H, 1.0 R.

The composition of the peptide showed it to be the hexapeptide L V E Q H R_{220} corresponding to the carboxyl terminal portion of peptide T11 (see Analyses of SL Tryptic Peptides - Peptide T11). It was not submitted to sequence analysis because it did not overlap any tryptic peptide.

Alignment of SL Tryptic Peptides

Sainfoin lectin tryptic peptides were aligned by two procedures: by overlapping thermolytic peptides and by homology with the leguminous lectin favin. Five of SL thermolytic peptides contained amino acid sequences overlapping certain of SL tryptic peptides. These overlapping sequences allowed the tryptic peptides to be aligned in seven blocks of continuous sequence, as illustrated in Figure 33. The first block of continuous sequence consisted of pepides T1, T2 and T3. The adjacent positions of these peptides were established by the thermolytic overlaps Th1 and Th2. Block

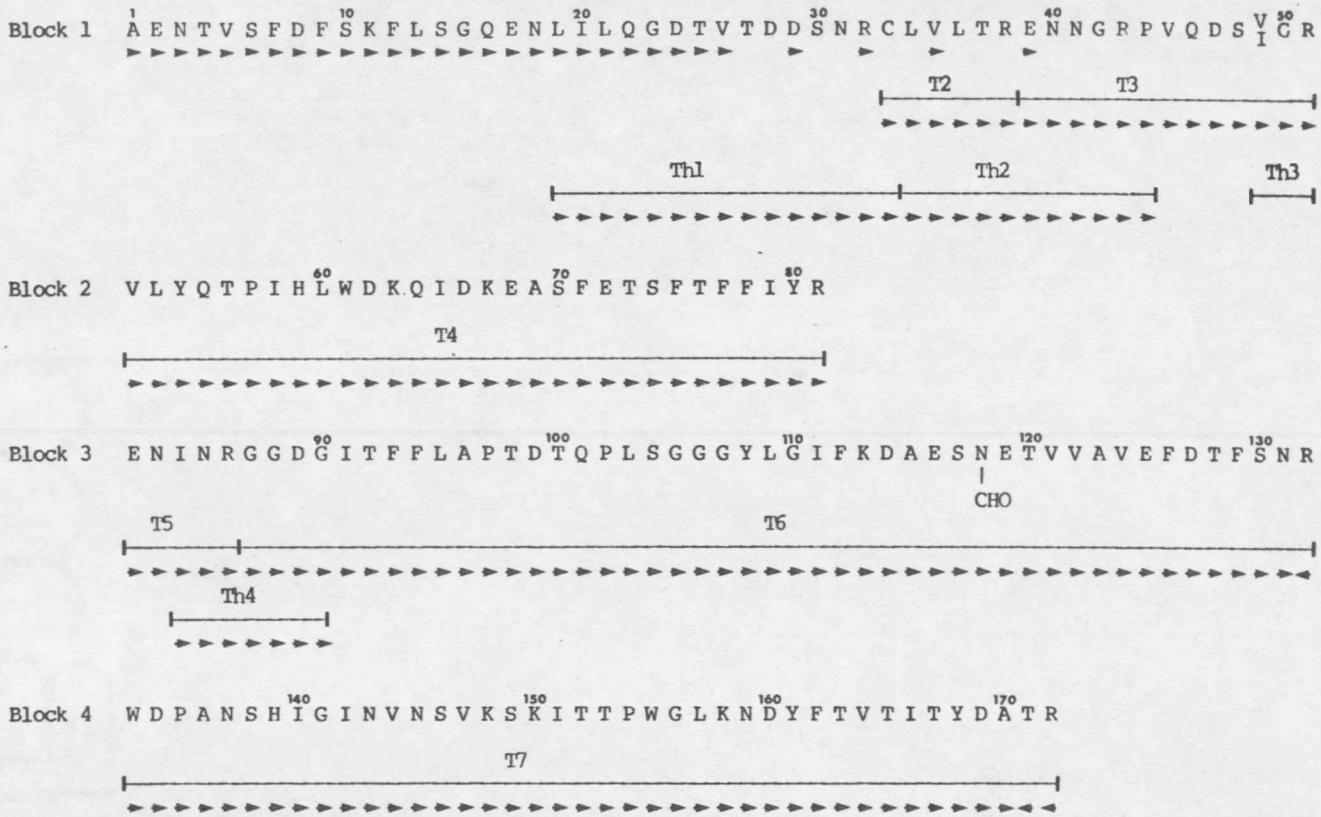
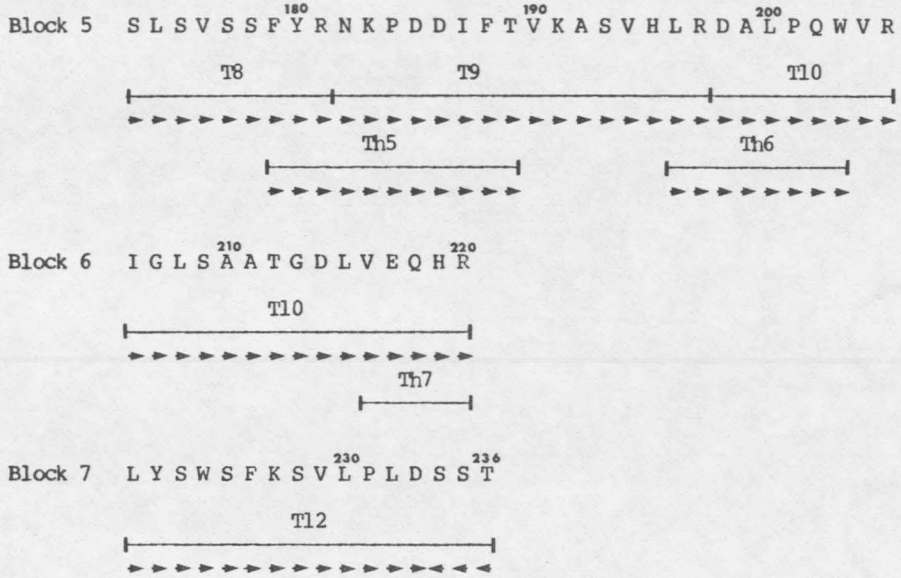


Figure 33. Reconstruction of seven blocks of continuous sequence from results obtained on SL component peptides

Figure 33 (continued)



[a]The arrows directly beneath the sequence indicate residue indentifications made from SL analysis. Residues directly identified from automated degradations and by carboxypeptidase Y digestion are denoted by right (–>) and left (<–) arrows, respectively. The asparagine at position 118 is glycosylated (CHO). T, tryptic peptide. Th, thermolytic peptide.

number two was composed of peptide T4. Block number three consisted of peptides T5 and T6. Thermolytic peptide Th4 overlapped these peptides. Block number 4 was composed of peptide T7. Block number 5 consisted of peptides T8, T9 and T10. The adjacent positions of these peptides were established by the thermolytic overlaps Th5 and Th6. Blocks number 6 and 7 were composed of peptides T11 and T12, respectively. The amino terminal location of block 1 was established by sequence analysis of the intact protein. The carboxyl terminal position of block 7 was proven by the absence of arginine of peptide T12. The five internal blocks were aligned through homologies with favin as shown in Figure 34. Each block of SL sequence was positioned under the portion of the favin molecule which showed a maximum number of identities with the sequence under consideration. This procedure resulted in the complete structuring of the SL amino acid sequence. Additional comments concerning the alignment assignment appear in the Discussion.



Figure 34. Alignment of SL peptides using favin as a template

CHAPTER FIVE

DISCUSSION

Sequence Strategy--Plan of Attack

The complete primary structure of sainfoin lectin (SL) will be determined by Edman degradation (Edman & Begg, 1967) of the protein and derived peptides. The peptides will be generated and purified at Montana State University. The actual sequence analyses will be performed at Washington University, St. Louis, Missouri, where a sequenator is available, through a collaborative arrangement with Dr. R. A. Bradshaw. The protein will be fragmented with an enzyme of narrow specificity, trypsin, in order to produce a limited number of peptides. The protein will be chemically modified at lysine in order to restrict trypsin cleavage to arginine and to increase the solubility of the fragments. The protein cysteine residue will be stabilized by alkylation with a ^{14}C tag for later identification of the cysteine containing peptide. The fragments will be purified and sequenced. They will be aligned by producing a second set of purified peptides and sequencing those containing overlapping regions of primary structure.

Production and Purification of Tryptic Peptides

Sainfoin lectin was succinylated exhaustively, carboxymethylated, and digested with trypsin. The succinylation was successful in restricting the action of trypsin to Arg-X bonds. Carboxymethylation with [^{14}C-1] iodoacetic acid prevented possible oxidative reactions which would have resulted in low recovery of the cysteine containing peptide. The label specifically introduced by SL carboxymethylation allowed easy detection of the cysteine peptide after cleavage with trypsin. The succinylated tryptic peptides were soluble in aqueous buffer and reasonably manageable. They were purified by standard purification techniques, and most of them were of appropriate length to be sequenced through to their carboxyl terminus.

The first purification step of the tryptic peptides was fractionation at pH 3.5. This simple step separated the small soluble tryptic peptides from the larger insoluble peptides and thereby reduced subsequent purification complexity. The eight small tryptic peptides were purified by gel filtration or cation exchange chromatography. The three large tryptic peptides were purified by gel filtration; two of them required subsequent repurification by HPLC.

Sainfoin lectin contains 12 arginines, one of which is found in an Arg-Pro bond (Arg43—Pro44, Figure 34). Arginyl-prolyl peptide bonds are resistant to trypsin cleavage (Smyth, 1967); therefore 12 fragments should result from cleavage by trypsin of succinylated SL. Eleven fragments were isolated that accounted for 202 out of the 234 amino acids of the intact protein. The remaining 32 residues were contained in the amino terminal peptide and were sequenced through Edman degradation of the intact protein. All amino acid residues of SL are therefore accounted for in the sequence presented here.

Sequence Analyses

Sainfoin lectin tryptic peptides did not need to be further cleaved in order to have their amino acid sequence established. Most of them were of suitable size for complete sequence determination with the sequenator. Sequence analysis of the largest peptides required the determination of long stretches of sequence, 30 amino acids or more. Sequence analysis of many residues is subjected to the following limitations. During Edman degradation coupling yields of the polypeptide with phenylisothiocyanate and cleavage of the amino terminal residue with heptofluorobutyric acid are typically 98% under optimum conditions (Allen, 1981). Consequently, increasing amounts of background amino acids are observed as the degradation proceeds through many cycles. Recovery

of PTH amino acids also varies depending on the amino acid involved. Correct sequence determinations therefore require precise quantitation of the PTH amino acids. This was made possible for the largest tryptic peptides by the detection and quantitation of the PTH amino acids by reversed-phase HPLC and associated high sensitivity of this detection system.

Another important aspect of the sequence analysis was the inclusion of polybrene in the spinning cup. Polybrene is a polymer containing quaternary amino groups that retain a positive charge during the degradation cycles and hold the polypeptides in the cup by ionic interactions (Thomas et al., 1981). This minimized losses resulting from extraction of the polypeptides (washout) and allowed sequencing of larger stretches of amino acids.

In the case of three tryptic peptides Edman degradation of the peptides was complemented by carboxypeptidase Y analysis of the peptides. The carboxyl terminal arginine of T6 and T7 was not detected by the automated sequence analysis. It was probably washed out of the spinning cup during the last degradation cycle. Presence of arginine at the carboxyl terminus of T6 and T7 was established through amino acid composition of the peptides, trypsin cleavage specificity, and carboxypeptidase Y analysis. Peptide T12 residues at and near the carboxyl terminus, threonine and serine, were also

confirmed with carboxypeptidase Y. Threonine and serine undergo extensive beta-elimination during the conversion of anilinothiazolinones to phenylthiohydantoins resulting in low recovery of these residues. Furthermore, the dehydroserine derivative is itself unstable and degraded to unidentified products (Allen, 1981). Peptide T12 was the 16 residue carboxyl terminal SL tryptic peptide. During its sequence analysis the PTH amino acids were identified by combination of thin layer chromatography and gas chromatography. Neither technique was sensitive enough to give a positive identification of threonine or serine at the 16th and 15th cycle and therefore carboxypeptide Y analysis was utilized for their identification.

Alignment of Tryptic Peptides

Certain SL tryptic peptides were aligned with thermolytic peptides which contained amino acid sequences overlapping two tryptic peptides. Five such thermolytic overlaps were found which joined T1 and T2, T2 and T3, T5 and T6, T8 and T9, T9 and T10. Peptides T4, T11, T12 could not be aligned by thermolytic overlaps because they contained thermolysin cleavable residues (Val, Ile, Leu) at their amino termini. The lack of thermolytic peptides containing the Arg81—Glu82, Arg132—Trp133 and Arg172—Ser173 bonds

and overlapping T4—T5, T6—T7 and T7—T8, respectively, was unexpected. It is possible that the enzyme cleaved these bonds. Minor cleavages at Glu, Trp and Ser by thermolysin have been reported (Heinrikson, 1977). It is also possible that some or all of these overlaps were produced, but they were not detected during the isolation procedure. The thermolytic overlaps were detected on the basis of their arginine content. In the main purification step of the peptides (ion exchange chromatography in pyridine buffers) the overlaps were detected with the qualitative arginine reagent phenantrenequinone. It is conceivable that some of the non-recovered overlaps were produced in relatively low yield and escaped the detection procedure.

The thermolytic peptides allowed the reconstruction of SL primary structure into seven blocks of continuous sequence (Figure 33). The amino terminal location of block number 1 which consisted of peptides T1, T2, T3, was dictated by the amino terminal position of peptide T1. The carboxyl terminal location of block number 7 which consisted of peptide T12 only, was ascertained by the absence of arginine in T12. Homologies were then used to complete the alignment. Blocks 2, 3, 4, 5 and 6 were aligned using the protein favin as a template (Figure 34). Favin is a leguminous lectin which exhibited 42% identities with SL. Such a high degree of homology

between the two proteins indicated resemblance in the proteins' primary structures and justified the alignment procedure. Four of the five blocks to be positioned consisted of large stretches of sequence of 29 amino acids or more. At least sections of these blocks were located in highly homologous regions so that their position could not be mistaken. The only small block, consisting of tryptic peptide T11, was also positioned under a region of favin that exhibited 42% identities with T11. The position of the blocks based on homology with the template favin was therefore unambiguous.

Observation of the position of each SL tryptic peptide relative to favin demonstrates that only three tryptic peptides showed little or no identities with favin, T5, T8, and T9 (Figure 34). Three thermolytic overlaps proved the location of these peptides. The position of each SL tryptic peptide in the protein primary structure was therefore established on the basis of homology or with a thermolytic overlap, or both.

Sainfoin lectin was found homologous to all the leguminous lectins whose sequence has been established, i.e. ConA, SBA, favin and LL (see below). The alignment of SL tryptic peptides established with favin would therefore have been the same if ConA, SBA, or LL had been chosen as a template. This fact confirmed the validity of SL peptide alignment presented in this research.

Sainfoin Lectin Amino Acid Sequence

The complete sequence of the 236 amino acids of SL was established (Figure 34). The subunit molecular weight calculated from the amino acid sequence was 26,509, in good agreement with previously reported molecular weight estimates (Hapner & Robbins, 1979).

The amino acid sequence composition of SL shown in Table 5 is compared to the hydrolyzate composition previously established. The composition from sequence analysis of SL was nearly identical to the previously published composition. Minor differences in the amounts of Thr, Ser, Gly and Ala could easily be explained by variations in recovery yield of these amino acids after acid hydrolysis of the protein.

The single cysteine residue of SL was located at position 33 of the protein. Although located near the amino terminus of SL, the cysteine side chain must be positioned in the interior of the molecule in order to explain its lack of reactivity in native SL (Hapner & Robbins, 1979).

The carbohydrate chain of SL was covalently bound to Asn118. This was the single glycosylation site, as established previously by Namen & Hapner (1979). Asparagine 118 was found in the sequence ---SNET--- that corresponded to the sequence published by Namen &

Table 5. Amino acid composition of sainfoin lectin[a,b]

Amino acid	Residues[a]/subunit	Residues[b]/26,509 g protein
Lys	11	11.3
His	4	3.5
Arg	12	11.3
Asx	33	32.7
Thr	21	19.6
Ser	24	24.9
Glx	18	18.4
Pro	9	9.3
Gly	15	16.0
Ala	11	11.9
Cys	1	0.9
Val	17.5	17.2
Met	0	0
Ile	13.5	12.7
Leu	18	18.3
Tyr	7	6.6
Phe	16	15.9
Trp	5	5.0
TOTAL	236	235.5

[a]from sequence analysis
[b]recalculated from Hapner & Robbins (1979)

Hapner (1979) of the SL glycopeptide produced by pronase digestion. This tetrapeptide sequence appears only once in the total sequence. Furthermore, the obligate sequence Asn-X-Thr(Ser) for N-glycosylation (Marshall, 1972) is not found in SL except for the ---SNET--- shown above.

One allelomorphic replacement was observed at position 49. The two residues observed at this position, Val and Ile, could not be responsible for the heterogeneity of SL observed on alkaline electrophoresis and isoelectricfocusing. An otherwise unique amino acid sequence was established for SL indicating the absence of isolectin species. The amino acid sequence of SL probably varies in amide content, i.e. Asp/Asn and Glu/Gln, that accounts for the ionic heterogeneity. The amide differences may have occured naturally or as a result of protein isolation procedures. They would not have been detected during the sequence analysis. Allelomorphic replacements in addition to the one at position 49 conceivably could have escaped detection. If two amino acids were present at one position that were recovered in different yields, one of them could have been overlooked as a contaminant.

Comparison of SL, Concanavalin A, Favin, Lentil Lectin and Soybean Agglutinin

Figure 35 shows the amino acid sequences of concanavalin A (ConA), SL, soybean agglutinin (SBA), favin and lentil lectin (LL) aligned in such a way as to maximize homologies among the proteins. The top line shows the ConA polypeptide chain starting at residue one at the top left of the figure and finishing at residue 237 at the bottom right. Sainfoin lectin and SBA are both composed of a

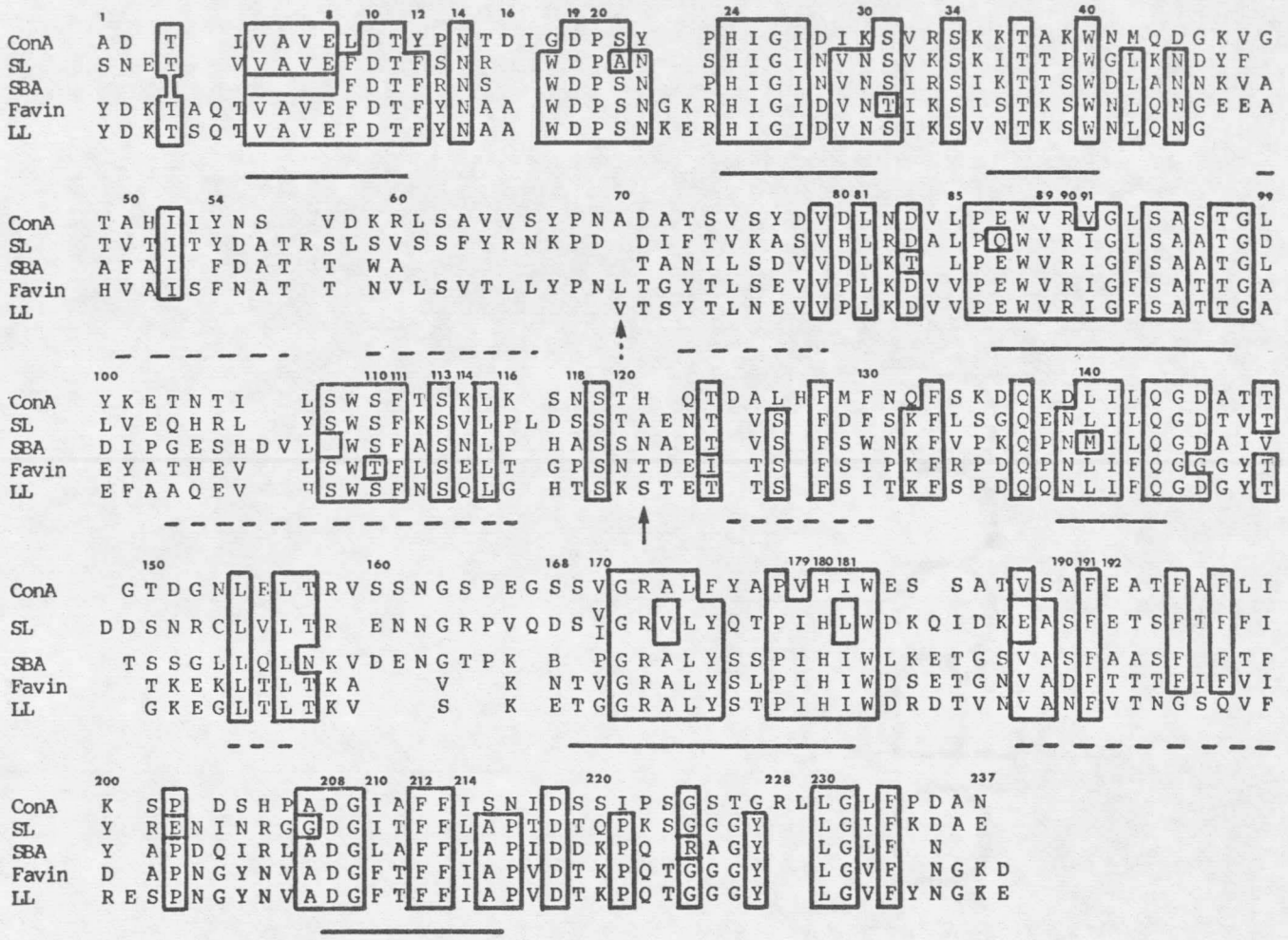


Figure 35. Comparison of the amino acid sequence of conA, SL, SBA, favin and LL

single polypeptide chain and their amino terminal amino acids correspond to ConA residue 121 (indicated by a solid arrow on Figure 35). Their carboxyl terminal amino acid corresponds to residue 120 of ConA. Favin and LL are composed of two subunits, a small alpha and a large beta polypeptide chain. The amino terminus of the beta subunit of favin coincides with the amino terminus of SL and SBA and is indicated by the solid arrow. The amino terminus of LL beta subunit corresponds to residue 2 of SL, favin beta chain and SBA. The amino terminus of the alpha subunit of favin and LL is located at residue 70 of ConA and is designated by a dashed arrow. The amino terminus of ConA therefore does not coincide with any amino termini of the other lectins indicating that the ConA alignment differs from that of the other proteins.

For a better understanding of the relative positions of the lectin sequences, the alignments of ConA, favin and SL have been schematized in Figure 36. The alpha subunit of favin is aligned between residues 70 and 120 of ConA. The beta subunit starts at residue 121 of ConA, runs through the ConA carboxyl terminus, picks up at ConA amino terminus to finish at residue 69 of ConA. Sainfoin lectin is aligned between the amino terminus of favin beta subunit and the carboxyl terminus of the alpha subunit. Lentil lectin and SBA which are not represented on the diagram (for simplicity) are

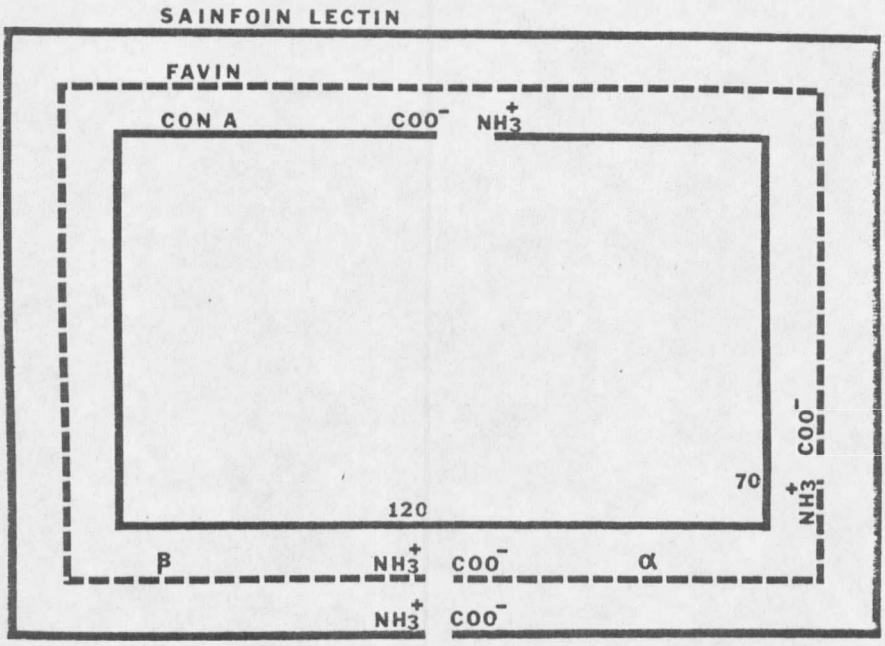


Figure 36. Schematic alignment of SL and favin with ConA

aligned like favin and SL, respectively. The circular permutation observed between ConA and the other lectins was first described by Cunningham et al. (1979) for ConA and favin. The genetic event resulting in the permutation is unknown. It would be instructive to establish the gene structure of ConA and the other lectins by gene isolation, cloning and DNA sequencing. Comparison of the genomic structures possibly would give an insight into the mechanism by which the circular permutation occured.

The number of amino acid identities exhibited by each couple of lectins of Figure 35 is shown in Table 6.

Table 6. Identities among ConA, SL, SBA, favin and LL

	ConA	SL	SBA	favin	LL
ConA	237	39%	37%	40%	34%
SL	93.5	236	41%	42%	40%
SBA	88	97	208	47%	48%
favin	94	100.5	110	236	69%
LL	81	94	101	163	211

The numbers on the diagonal are the total number of residues of each lectin. Percentages of identities for each lectin pair are on the right of the diagonal and the total number of identities are on the left. The degree of similarity between two homologous proteins is related to the evolutionary distance between the proteins (Doolittle, 1979). Although the evolutionary distances between ConA, SL, SBA, favin and LL cannot be computed from the data in Table 6, the relatedness of each pair of proteins can be qualitatively compared. The table shows that favin and lentil lectin are the most related proteins (69% identities), ConA and lentil lectin the least related (34%). The other lectin pairs show approximately 40% identities indicating that their evolutionary distances are similar except SBA and favin and SBA and LL which are apparently somewhat more closely related (48% identity).

Favin and LL are the only lectins containing two chains with subunit structure $(alpha\text{-}beta)_2$, among the lectins under consideration. Their similarity in quaternary structure, together with their high degree of resemblance suggest a close relationship of the plants fava bean and lentil. Taxonomic relationships in the Leguminosae family (Bell et al., 1978) places fava bean and lentil in the same tribe VICIEAE. The similar amino acid sequences of favin and LL therefore support their taxonomical relationship.

Sainfoin belongs to the tribe CORONILLEAE and soybean and the jack bean (from which ConA is purified) both belong to the tribe PHASEOLEAE (Bell et al., 1978). It is notable from Table 6 that ConA and SBA which belong to the same tribe are less related than SBA and favin which belong to different tribes, apparently contradicting the assigned taxonomic relationship of the plants.

Evolutionary relationships between the tribes of the subfamily PAPILIONOIDEAE of the family LEGUMINOSAE have been proposed by Bell et al. (1978). The three tribes to which ConA, SL, SBA, LL and favin belong were speculated to be of recent appearance. If this is true, it would indicate that the precursor of the lectins had an amino acid sequence aligned like SL, SBA, favin and LL, with one or two chains, and the genetic event that gave rise to ConA by circular permutation occured after the divergence of most tribes, perhaps even within the tribe PHASEOLEAE.

A phylogenetic tree of the lectins would be necessary to compare precisely the information given by the protein amino acid sequences to the existing taxonomy. Such a tree is computed by calculating the minimum number of genetic events required to obtain one amino acid sequence from another one (Fitch & Margoliash, 1967) or by reconstructing the ancestral sequence at each branch point of the tree (Dayhoff, 1972). Both methods require extensive use of a

computer for comparing extended amino acid sequences. It is beyond the scope of the work presented here.

Structure Conservation

Of the five lectins presented in Figure 35, only ConA has had its tertiary structure determined by X-ray crystallography (Becker et al., 1975; Reeke et al., 1975). The structure of ConA is characterized by two large antiparallel beta structures and almost no alpha structure. A prediction of the secondary structure of lentil lectin was made by Foriers et al. (1981). The calculations predicted a structure analogous to the one of ConA. Although the secondary structures of SL, SBA and favin have not been determined or predicted, the high degree of homology among these lectins and ConA and LL is predictive of similar tertiary structure.

Influence of Circular Permutation and Subunit Structure

Examination of ConA tertiary structure (Becker et al., 1975) reveals that the carboxyl terminus is located at the end of a series of irregular bends in the same region of the molecular surface as the amino terminus. It is conceivable that an additional or a different bend would permit residues 1 and 237 to form a peptide bond without significantly altering the general shape of the protein. This may be the case in SL, favin and LL which have the

corresponding residues linked. The amino terminus of SL and SBA and of the beta subunit of LL and favin is homologous to residue 121 of ConA. This histidine 121 is located in a loose turn and cleavage of the peptide bond 120-121 should not affect the overall structure of the protein. Concanavalin A is usually isolated as a mixture of intact single chain molecules and "nicked" molecules cleaved between residues 118 and 119 (Wang et al., 1971). The cleaved lectin is equally active indicating that the cleavage does not alter the protein folding. These observations suggest that the circular permutation between ConA and the other lectins may not result in major differences of the proteins' tertiary structure.

The amino terminus of the alpha subunit of favin and LL is homologous to Ala70 of ConA (Figure 36). The peptide bond 69-70 is located in a beta turn at the surface of ConA and could conceivably be cleaved without altering the molecule's general shape. This fact would suggest that the two chain lectins have similar spatial configuration to the single chain lectins.

In agreement with these observations, LL was predicted to be analogous to ConA, in spite of the circular permutation between the proteins and the difference in the quaternary structures (Foriers et al., 1981).

Secondary Structures

The residues of ConA that are conserved in the other lectins are examined below in order to hypothesize the type of ConA structures that may be common to the proteins. Identical residues observed in at least four of the five lectins are shaded in Figure 35 in order to highlight regions with a high degree of homology. Residues corresponding to each beta pleated sheet of ConA were underlined, with a solid line for the front beta structure and with a dashed line for the rear beta structure. It is notable that the residues of six of the seven strands of ConA front beta sheet are located in highly conserved regions. The rear beta structure is not as well conserved. Two of the six rear beta sheets are completely deleted in LL, partially in SBA, and correspond to low homology regions in the other lectins (residues 48 to 56, 59 to 66 of ConA, Figure 35). The other rear beta sheets show conservation of some residues but not to the extent of the front beta sheets. It is possible that the front beta structure of ConA will be found totally conserved in the other lectins, as predicted for lentil lectin (Foriers et al., 1981) while the rear beta structure will be only partially conserved. Other highly conserved residues are located in regions of ConA containing no regular secondary structures. This may suggest a role in maintenance of functional tertiary structure.

Metal Binding Sites

Two metal binding site, one for Ca^{++} and one for Mn^{++} are found in ConA (Kalb & Levitzki, 1968), LL (Ticha et al., 1970), favin (Cunningham et al., 1979) SBA (Galbraith & Goldstein, 1970), and SL (studies herein). In ConA, the Ca^{++} ion is bound to six ligands: it is coordinated to both carboxylic oxygens of Asp10, to one carboxylic oxygen of Asp19, to the one of Asn14, to the carbonyl oxygen of Tyr12 and to two water molecules (Becker et al., 1975). Figure 35 shows a total conservation of Asp10, Asn14, and Asp19 in all the lectins. Tyrosine 12 appears as phenylalanine in the other lectins. This change should not affect the binding site as the tyrosine side chain was not involved in the coordination of Ca^{++}. One of the water ligands is hydrogen bonded to Asp208 in ConA. This residue is also conserved in the other lectins. The complete conservation of the Ca^{++} ligands of ConA supports the identity of the Ca^{++} binding site in ConA and the other lectins.

The Mn^{++} ion has six ligands: the side chains of Glu8, Asp10, Asp19, His24 and two water molecules (Becker et al., 1975). The amino acid ligands are conserved in the other lectins except Glu8 which is deleted in SBA. The Mn^{++} binding site of ConA therefore seems to be totally conserved in the other lectins, with the exception of SBA which has at least one different amino acid ligand.

Presence of near identical Ca^{++} and Mn^{++} binding sites in the five lectins may reflect the importance of the metals for the functional protein configuration and reinforces the prediction of similar tertiary structure of the lectins.

Carbohydrate Binding Site

The carbohydrate binding site of ConA has been identified in crystals of ConA alpha-methylmannoside complex (Hardman & Ainsworth, 1976). The side chains in closest proximity to the bound carbohydrate belong to Tyr12, Asn14, Asp16, Leu99, Tyr100, Ser168, Asp208 and Arg228. Involvement of aspartyl and tyrosyl groups in the carbohydrate binding site of ConA is suggested by chemical modification experiments (Hassing et al., 1971; Doyle & Roholt, 1968). Only two of these residues are conserved in all the lectins, Asn14 and Asp208 (Figure 35). Tyrosine 12 has been changed to phenylalanine in all the lectins. Aspartate 16 is deleted in SL and SBA and Arg228 is deleted in SL, SBA and LL. The only other conserved residue besides Asn14 and Asp208 is Leu99 which is also found in SBA. It therefore seems legitimate to deduce that the carbohydrate binding site of ConA is poorly, if at all conserved in the other lectins. This result is surprising for SL, favin and LL, all of which have similar binding specificity to ConA, i.e., D-mannose and D-glucose (Goldstein & Hayes, 1978). The binding site

in these lectins was apparently no better conserved than that in SBA which shows a different binding specificity (N-acetyl galactosamine). Detailed comparison of inhibition of agglutination by ConA, favin and LL has shown differences in the affinity of the lectins for various saccharides (Allen et al., 1976). Furthermore, several 3-O substituted glucose derivatives were bound poorly by ConA whereas they were excellent inhibitors of favin and lentil lectins. These binding differences possibly result from the low conservation of the amino acids located in ConA carbohydrate binding site among the other lectins.

Hydrophobic Cavity

Concanavalin A tertiary structure shows the presence of a deep hydrophobic cavity which binds nonpolar molecules such as o-iodophenol and o-iodobenzoic acid (Becker et al., 1975) and which is surrounded by the side chains of Tyr54, Leu81, Leu85, Val89, Val91, Phe111, Ser113, Val179, Ile181, Phe191, Phe212, and Ile214. Six of these twelve residues are conserved in all the lectins (Leu81, Val89, Phe111, Ser113, Phe191 and Phe212, Figure 35). Isoleucine 181 was conserved in all the proteins but SL where it was replaced by the homologous amino acid leucine. Leucine 85 was conserved in SL and SBA, replaced by the homologous residue Val in

favin and LL. Isoleucine 214 was conserved in favin and LL, replaced by Leu in SL and SBA. Valines 91 and 179 were replaced by Ile in all the other proteins. Tyrosine 54 was conserved in SL, replaced by Phe in SBA and favin, deleted in LL. The residues surrounding the ConA hydrophobic cavity were therefore conserved in the other lectins or replaced by homologous amino acids. The high degree of conservation of this structure suggests a possible importance in the molecular functioning of plant lectins. Concanavalin A hydrophobic cavity was demonstrated to be a weak binding site for the plant hormone auxin beta indolacetic acid (Edelman & Wang, 1978). These authors reported that favin also binds the plant hormone and suggested that the lectins may serve as a storage or transport protein for the hormone.

Monomer-Monomer and Dimer-Dimer Contacts

The interactions involved in the ConA dimer formation are associated with residues 87, 88, 90, 117, 119, 123, 124, 126, 130, 131, 132, 134, 136, 137, 139, 175, 176, 177, 178 and 180 (Becker et al., 1975). Thirteen out of these 20 residues are conserved in the other lectins or replaced by homologous amino acids. This conservation indicates that the contacts between protomers in the dimer of SL, favin, LL and SBA may be analogous to the contacts in ConA.

Differences in monomer-monomer orientation are expected in SBA wince it has only one carbohydrate binding site per dimer (Lotan et al., 1974). Existence of a single binding site in two identical polypeptide chains indicates that the two subunits are in a different conformation or oriented in such a fashion as to render one binding site inaccessible. Such a non-equivalence of the protomers is unique to SBA among the lectins under consideration.

The interactions involved in ConA tetramer formation are associated with residues 49, 51, 53, 55, 57, 58, 60, 62, 64, 66, 68, 72, 76, 108, 110, 112, 114, 116, 118, 121, 122, 125, 127, 129, 138, 187, 188, 190, 192 and 194 (Becker et al., 1975). Salt links between Lys114 and Lys116 of one dimer and Glu192 of the other dimer are also involved in tetramer maintenance (Becker et al., 1975). Of the 30 contact residues, only four are conserved in the other lectins and 11 were totally deleted in LL. Residues responsible for the salt bridge were not conserved either. The salt bridge absence together with the low degree of conservation of the dimer contact residues probably reflect the difference of quaternary structure of the lectins: sainfoin lectin is a dimer of identical subunits (Hapner & Robbins, 1979) and favin and LL are dimers of alpha-beta subunits (Goldstein & Hayes, 1978). Soybean agglutination which is

a tetramer like ConA does not appear to have analogous dimer-dimer contacts.

Carbohydrate Attachment Site

Sainfoin lectin, SBA and favin are glycoproteins, while ConA and LL contain no carbohydrate. The carbohydrate chain of SL is linked to Asn 118 (this work) that which is homologous to residue 2 of ConA (Figure 35). The carbohydrate chain of favin is linked to Asn168 (Hopp et al., 1982) corresponding to ConA Asn55 (Figure 35). Soybean agglutinin also contains a carbohydrate chain linked to an asparagine residue (Lis & Sharon, 1977). The position of the carbohydrate in SBA is not yet available. However, Asn41 (homologous to Asn162 of ConA, Figure 35) is the only asparagine residue of SBA located in a sequence of the type Asn-X-Thr(Ser) that is necessary for glycosylation to occur (Marshall, 1972). It is therefore reasonable to predict that Asn41 is the glycosylation site of SBA. For an unknown reason, the corresponding sequence in ConA (Asn-Gly-Ser) is not glycosylated in spite of the fact that the tripeptide sequence is exposed in a beta turn at the surface of the molecule. Sainfoin lectin, SBA and favin show different glycosylation sites, Asn residues 118, 41 and 168, respectively. This observation, together with the fact that ConA and LL are not glycoproteins, implies that the carbohydrate content of the lectins,

when present, does not play a major role in the protein overall configuration and the protein binding characteristics.

Structure Conservation. Summation

In brief, the amino acid sequences of ConA, SL, SBA, favin and LL predict an analogous three-dimensional structure of these lectins. Comparison of the residues highly conserved in the five lectins in conjunction with the knowledge of ConA tertiary structure predicted the conservation of the seven front antiparallel beta pleated sheets of ConA and conservation of at least three of the six rear beta pleated sheets. Analogy of tertiary structure of the lectins was implied by the high conservation of ConA residues forming the metal binding sites and those forming the hydrophobic cavity. Conservation or replacement by homologous amino acids of more than half of ConA residues implied in monomer-monomer contact suggests some analogy in the proteins' quaternary structure. Most of the dimer-dimer contact resioues and the salt link involved in ConA tetramer maintenance were not conserved, as expected since SL, favin and LL do not form tetramers. The most surprising result from the comparison of the lectins' amino acid sequences was the low conservation of the sugar binding site, even in SL, favin and LL that have the same carbohydrate binding specificity as ConA.

Crystallographic data will be needed to confirm the analogy of tertiary structure of the lectins suggested by the homologous primary structures.

Other Leguminous Lectins

Several leguminous lectins, composed of alpha and beta subunits similar to favin and LL, have had their small alpha subunit sequenced: pea lectin (Richardson et al., 1978), the lectin from Vicia sativa (Gebauer et al., 1981) and the two chain lectin of Vicia cracca (Baumann et al., 1982). Comparison of the amino acid sequences of the alpha subunit of these three lectins and favin and LL showed a high degree of homology (Baumann et al., 1982). The lectins all belong to the same leguminous tribe VICIEAE. Their similarity of primary structure was expected and did not bring new information to the structural comparison of leguminous lectins.

The sequence of the terminal residues of peanut lectin, the beta subunit of pea lectin, the R and L subunits of PHA (lectin derived from Phaseolus vulgaris), soybean lectin, and the beta subunit of lentil lectin were compared (Foriers et al., 1977). Amino terminal residues of Dolichos biflorus lectin (Etzler et al., 1977) and of Lathyrus odoratus lectin (Kolberg et al., 1980) were also sequenced and compared to the above lectins. Lentil lectin, Lathyrus odoratus lectin, and pea lectin showed near identity, while

all the lectins were homologous to each other. If the homology of the amino terminal residues extends to the whole polypeptide chains, it would indicate that lectins of several carbohydrate binding specificities are homologous. Lentil, Lathyrus odoratus, and pea lectins belong to the D-mannose, D-glucose binding family, together with favin, ConA and SL. Peanut lectin binds D-galactose, SBA and Dolichos biflorus lectin bind N-acetyl galactosamine; PHA binds a complex poorly defined carbohydrate structure (Goldstein & Hayes, 1978). It is therefore possible that the high level of homology of primary structure observed in ConA, SL, favin, SBA and LL be extended to all leguminous lectins, regardless of their carbohydrate binding specificity.

The high conservation of structure among the lectins supports the hypothesis of a common precursor for the proteins (Foriers et al., 1977). Most important, the structure conservation suggests that the lectins have essential functions in the legumes and that they fulfill these functions with a similar mechanism. It would be interesting to compare the available lectin sequences to that of the clover lectin. The latter has been extensively studied for its physiological function and has been found clearly involved in the binding of rhizobium to clover roots (Dazzo, 1980). The clover lectin amino acid sequence has not been determined. Structural

analogy with this lectin may support the hypothesis that lectins mediate the binding of specific rhizobia to legumes.

It is possible that lectins derived from legumes form a family of proteins with homologous primary and tertiary structure, similar overall conformation, and specific differences localized in the carbohydrate binding region. The general structure would have been conserved in order to preserve the mechanism by which the lectins accomplish their yet unknown function. Different binding sites would allow for the specificity of the molecules on which the function is performed. In this sense, this family of proteins appears to be analogous, for example, to the serine proteinases.

CHAPTER SIX

SUMMARY AND CONCLUSIONS

Sainfoin lectin exists as a dimer in solution at neutral pH and monomerizes progressively as the pH is decreased. Monomerization is accompanied by competing denaturation so that complete dissociation into native monomers cannot be achieved. The protein is protected to some extent from low pH denaturation by the presence of glucose, suggesting some stabilizing effect of the interacting carbohydrate. Solubility of SL in aqueous buffers is also greatly enhanced when the carbohydrate binding site is occupied.

Sainfoin lectin succinylated at an average of 10 out of 11 lysine residues is predominantly in the dimeric conformation, while totally succinylated SL is predominantly in the monomeric conformation. This observation raises the possibility that a lysine residue could be involved in a salt link between SL monomers and responsible for dimer maintenance. The situation would be analogous to the salt bridge of conA, formed by two lysines of one dimer and one glutamic acid of the other dimer, responsible for tetramer maintenance (Reeke et al., 1975).

Sainfoin lectin is trypsin resistant in its native conformation, and after low or high pH exposure. The protein is not

an inhibitor of trypsin, indicating the unavailability of tryptic sites in SL. Totally succinylated SL is also trypsin resistant in spite of the fact that it has lost its carbohydrate binding ability. Succinylation therefore induces a conformational change, at least at the level of the carbohydrate binding site, but it does not expose any tryptic cleavage sites. Sensitization of SL to trypsin can be induced by alkylation of the protein's single cysteine residue. Trypsin digestion then occurs very rapidly with all tryptic sites being cleaved.

Sainfoin lectin can be fractionated into four ionic species by anion exchange chromatography. These species exhibit hemagglutination activity, similar amino acid composition, and similar immunodiffusion behavior. They are homogeneous in pH 9.5 electrophoresis but remain heterogeneous on isoelectricfocusing. These observations, together with the determination of a unique amino acid sequence, indicate the absence of isolectins in sainfoin. The charge heterogeneity is best explained by differences in amide content that could have arisen during isolation of the protein, or it may be a natural feature of sainfoin lectin as isolated from seeds.

Sainfoin lectin is a metalloprotein containing one Ca^{++} per monomer and non-stoichiometric amounts of Mg^{++} and Mn^{++}. By analogy

with other lectins of the D-mannose (D-glucose) binding family (Goldstein & Hayes, 1978), SL is believed to have one Ca^{++} binding site and one site for Mn^{++}. The transition metal may be partially lost during purification procedure as occurs in conA (Christie et al., 1978). Its site may then be filled with other available divalent cationic metals.

Affinity electrophoresis of SL at pH 4.0 separates monomer and dimer species of the protein. Dissociation constants for the protein-free glucose complexes calculated with this technique are 1.59×10^{-3} M and 1.97×10^{-3} M for dimer and monomer, respectively. Dissociation constants for the protein-entrapped glucose complexes are 0.81×10^{-3} M (dimer) and 3.29×10^{-3} (monomer). The larger affinity of the dimer for the immobilized ligand may be due to its larger number of binding sites (Horejsi et al., 1977).

Equilibrium dialysis of SL indicates the protein to have a single binding site per subunit. Dissociation constant for SL-glucose calculated by this technique is 3.52×10^{-3} M and the n value is 1.03.

Mild acid hydrolysis of SL does not specifically cleave the single Asp-Pro peptide bond of the protein. This treatment results in the production of several fragments that are unmanageable because of poor solubility and aggregation properties.

Cyanylation of SL followed by incubation at pH 9 results in cleavage of the protein at its single cysteine. The cleavage produces two fragments of approximate MW 3,500 and 23,400. The large fragment contains the cysteine derivative, indicating SL cysteine residue to be located approximately 30 amino acids from the polypeptide chain amino terminus. Cleavage at cysteine of SL was not useful for the production of peptides for sequence analysis. The small amino terminal peptide had its sequence established through sequence analysis of the whole protein. The large peptide was blocked at its amino terminus and could not be submitted to Edman degradation.

The basic fragmentation method of SL which led to the production of peptides for SL amino acid sequence was trypsin digestion of the succinylated, carboxymethylated protein. Eleven peptides were thereby produced and purified by standard techniques (isoelectric precipitation, gel filtration, ion exchange chromatography and HPLC). These eleven peptides, together with the 32 amino terminal residues of SL, accounted for the whole protein. The amino acid sequence of SL was established by Edman degradation of the tryptic peptides, of the intact protein, and of a set of overlapping thermolytic peptides. The tryptic peptides were ordered

with the thermolytic overlaps and by homology with other leguminous lectins.

A unique amino acid sequence was established for SL that contains one allelomorphic replacement Val(Ile)49 (Figure 33). The lectin contains 236 amino acids and has a MW of 26,509. The single cysteine is at position 33. The carbohydrate chain of SL is attached to Asn118.

Comparison of SL amino acid sequence to conA, favin, LL and SBA reveals a high degree of homology among the lectins, approximately 40% identities between each lectin pair except favin and LL, which are even more closely related (69% identities). This observation suggests a common precursor for the five lectins.

Sainfoin lectin is aligned like SBA, i.e. its amino terminus corresponds to SBA amino terminus and its carboxyl terminus corresponds to SBA carboxyl terminus. The two lectins resemble favin and LL when the latter proteins are aligned as if they were a single polypeptide chain beta-alpha. Concanavalin A differs from all the above lectins in that it exhibits a circular permutation, i.e. its residue 121 corresponds to the amino terminus of the other lectins.

Concanavalin A residues highly conserved in the other lectins include those forming the conA front beta sheet, the metal binding

sites, the hydrophobic cavity and the monomer-monomer contacts. The sequence homologies also support a lesser conservation of the rear beta sheet and the carbohydrate binding site. Analogous secondary and tertiary structures are therefore expected for conA, SBA, SL, LL and favin.

Comparison of all published partial amino acid sequences of leguminous lectins to the sequences of the five lectins discussed above suggests that all the leguminous lectins will be found homologous to each other. The structure conservation of the leguminous lectins supports an essential function for these proteins that has been evolutionarily retained.

REFERENCES

Allen, G. (1981) in Sequencing of Proteins and Peptides (Work, T. S. & Burdon, R. H., Eds.), Elsevier/North Holland, Amsterdam.

Allen, A. K., Desai, N. N. & Neuberger, A. (1976) Biochem J. 155, 127-135.

Ambler, R. P. (1972) Methods Enzymol. 25, 262-272.

Ashford, D. & Neuberger, A. (1980) TIBS 5, 245-248.

Ashwell, G. & Harford, J. (1982) Annu. Rev. Biochem. 51, 531-554.

Barondes, S. H. (1981) Annu. Rev. Biochem. 50, 207-231.

Bartles, J. R., Frazier, W. A. & Rosen, S. D. (1982) Int. Rev. Cytol. 75, 61-99.

Bauer, W. D. (1981) Annu. Rev. Plant Physiol. 32, 407-449.

Baumann, C. M., Strosberg, A. D. & Rudiger, H. (1982) Eur. J. Biochem. 122, 105-110.

Becker, J. W., Reeke, G. N. Jr., Wang, J. L., Cunningham, B. A. & Edelman, G. M. (1975) J. Biol. Chem. 250, 1513-1524.

Bell, E. A., Lackey, J. A. & Polhill, R. M. (1978) Biochem. Syst. Ecol. 6, 201-212.

Bradshaw, R. A., Bates, O. J. & Benson, J. R. (1980) J. Chromatog. 187, 27-33.

Brauer, A. W., Margolies, M. N. & Haber, E. (1975) Biochemistry 14, 3029-3035.

Brown, J. C. & Hunt, R. C. (1978) Int. Rev. Cytol. 52, 277-349.

Christie, D. J., Alter, G. M. & Magnuson, J. A. (1978) Biochemistry 17, 4425-4430.

Cocking, E. C. & Yemm, E. W. (1954) Biochem. J. 58, xii-xiii.

Cohen, E. (1974) Ann. N.Y. Acad. Sci. 234, 1-412.

Coulet, M. (1979) Ann. Immunol. 130C, 1-16.

Cunningham, B. A., Wang, J. L., Waxdal, M. J. & Edelman, G. M. (1975) J. Biol. Chem. 250, 1503-1512.

Cunningham, B. A. Hemperly, J. J., Hopp, T. P. & Edelman, G. M. (1979) Proc. Natl. Acad. Sci. U.S.A. 76, 3218-3222.

Dayhoff, M. O. (1972) Atlas of Protein Sequence and Structure, Vol. 5, Natl. Biomed. Res. Found., Silver Springs, Maryland.

Dazzo, F. B. (1980) 38th Symposium of the Society for Developmental Biology, pp. 277-304, Academic Press, New York.

Degani, Y. & Patchornik, A. (1971) J. Org. Chem. 36, 2727-2728.

Del Campillo, E., Shannon, L. M. & Hankins, C. N. (1981) J. Biol. Chem. 256, 7177-7180.

Doolittle, R. F. (1979) in The Proteins (Neurath, H. & Hill, R. L., Eds.), Vol. IV, pp. 1-118, Academic Press, New York.

Doyle, R. J. & Roholt, O. A. (1968) Life Sci. 7, 841-846.

Dubois, M., Gillis, K. A., Hamilton, J. K., Rebers, P. A. & Smith, F. (1956) Anal. Chem. 28, 350-356.

Easley, C. (1965) Biochem. Biophys. Acta 107, 386-388.

Edelman, G. M. & Wang, J. L. (1978) J. Biol. Chem. 253, 3016-3022.

Edman, P. & Begg, G. (1967) Eur. J. Biochem. 1, 80-91.

Ellman, G. L. (1959) Arch. Biochem. Biophys. 82, 70-77.

Etzler, M. E., Talbot, C. F. & Ziaya, P. R. (1977) FEBS Lett. 82, 39-41.

Etzler, M., Gibson, D. & Scates, S. (1979) Plant Physiol. 63, 134 (Abstr.).

Fisch, W. W., Mann, K. G. & Tanford, C. (1969) J. Biol. Chem. 244, 4989-4994.

Fisch, W. W., Hamlin, L. M. & Miller, R. L. (1978) Arch. Biochem. Biophys. 190, 693-698.

Fitch, W. M. & Margoliash, E. (1967) Science 155, 279-284.

Foriers, A., Wuilmart, C., Sharon, N. & Strosberg, A. D. (1977) Biochem. Biophys. Res. Commun. 75, 980-986.

Foriers, A., Lebrun, E., Van Rapenbusch, R., de Neve, R. & Strosberg, A. D. (1981) J. Biol. Chem. 256, 5550-5560.

Fornsted, N. & Porath, J. (1975) FEBS Lett. 57, 187-191.

Fraser, K. J., Poulsen, K. & Haber, E. (1972) Biochemistry 11, 4974-4977.

Galbraith, N. & Goldstein, I. J. (1970) FEBS Lett. 9, 197-201.

Gebauer, G., Schiltz, E. & Rudiger, H. (1981) Eur. J. Biochem. 113, 319-325.

Gibson, D. & Etzler, M. (1979) Plant Physiol. 63, 134 (Abstr.).

Goldstein, I. J. & Hayes, C. E. (1978) Adv. Carbohydr. Chem. Biochem. 35, 127-340.

Gordon, J. I., Smith, D. P., Andy, R., Alpers, D. H., Schonfeld, G. & Strauss, A. W. (1982) J. Biol. Chem. 257, 971-978.

Gray, W. R. (1972) Methods Enzymol. 25, 121-138.

Gunther, G. R., Wang, J. L., Yahara, I., Cunningham, B. A. & Edelman, G. M. (1973) Proc. Natl Acad. Sci. U.S.A. 70, 1012-1016.

Hankins, C. N. & Shannon, L. M. (1978) J. Biol. Chem. 253, 7791-7797.

Hankins, C. N., Kindinger, J. I. & Shannon, L. M. (1980) Plant Physiol. 66, 375-378.

Hapner, S. J. & Hapner, K. D. (1978) J. Histochem. Cytochem. 26, 478-482.

Hapner, K. D. & Robbins, J. E. (1979) Biochem. Biophys. Acta 580, 186-197.

Hardman, K. D. & Ainsworth, C. F. (1976) Biochemistry 15, 1120-1128.

Hassing, G. S., Goldstein, I. J. & Marini, M. (1971) Biochem. Biophys. Acta 243, 90-97.

Hauzer K., Ticha, M., Horejsi, V. & Kocourek, J. (1979) Biochem. Biophys. Acta 583, 103-109.

Hayashi, R. (1977) Methods Enzymol. 47, 84-93.

Hayes, C. E. & Goldstein, I. J. (1974) J. Biol. Chem. 249, 1904-1914.

Heinrikson, R. L. (1977) Methods Enzymol. 47, 175-189.

Hemperly, J. J., Becker, J. W. & Cunningham, B. A. (1981) in Proteins in Biology and Medicine (Bradshaw, R., Hill, R. & Tang, J., Eds.), Academic Press, New York.

Herzog, V. (1981) TIBS 6, 319-322.

Hirs, C. H. W. (1967) Methods Enzymol. 11, 197-199.

Hirs, C. H. W. (1967) Methods Enzymol. 11, 325-339.

Hopp, T. P., Hemperly, J. J. & Cunningham, B. A. (1982) J. Biol. Chem. 257, 4473-4483.

Horejsi, V. (1979) Les Colloques de l'Inserm 86, 391-398.

Horejsi, V. (1981) Anal. Biochem. 112, 1-8.

Horejsi, V. & Kocourek, J. (1974) Methods Enzymol. 34, 361-367.

Horejsi, V., Ticha, M. & Kocourek, J. (1977) Biochem. Biophys. Acta 499, 290-300.

Itoh, M., Kondo, K., Kamada, H., Izutsu, K., Shimabayashi, Y. & Takahashi, T. (1980) Agric. Biol. Chem. 44, 125-133.

Jacobs, J. W. & Niall, H. D. (1975) J. Biol. Chem. 250, 3629-3636.

Jacobson, G. R., Schaffer, M. H., Stark, G. R. & Vanaman, T. C. (1973) J. Biol. Chem. 248, 6583-6591.

Janzen, D. H., Juster, H. B. & Liener, I. E. (1976) Science 192, 795-796.

Kalb, A. J. & Levitzki, A. (1968) Biochem. J. 109, 669-672.

Kauss, H. & Glaser, C. (1974) FEBS Lett. 45, 304-307.

Klotz, I. M. (1967) Methods Enzymol. 11, 576-580.

Kolberg, J., Michaelsen, T. E. & Sletten, K. (1980) FEBS Lett. 117, 281-283.

Landon, M. (1977) Methods Enzymol. 47, 145-149.

Liener, I. E. (1976) Annu. Rev. Plant Physiol. 27, 291-319.

Lis, H. & Sharon, N. (1973) Annu. Rev. Biochem. 42, 541-574.

Lis, H. & Sharon, N. (1977) in The Antigens (Sela, Ed.), Vol. IV, pp. 429-529, Academic Press, New York.

Lotan, R., Siegelman, H. W., Lis, H. & Sharon, N. (1974) J. Biol. Chem. 249, 1219-1224.

Marshall, R. D. (1972) Annu. Rev. Biochem. 41, 673-702.

Matsumoto, I., Jinbo, A., Kitagaki, H., Golovtchenko-Matsumoto, A. M. & Seno, N. (1980) J. Biochem. 88, 1093-1096.

McKenzie, G. H., Sawyer, W. H. & Nichol, L. W. (1972) Biochem. Biophys. Acta 263, 283-293.

Means, G. E. & Feeney, R. E. (1971) Chemical Modification of Proteins, Holden-Day, Inc., San Francisco.

Miller, R. L. (1982) Fed. Proc. 42, 1033 (Abstr.).

Miller, J. B., Hsu, R., Heinrikson, R. & Yachnin (1975) Proc. Natl. Acad. Sci. U.S.A. 72, 1388-1391.

Millerd, A. (1975) Annu. Rev. Plant Physiol. 26, 53-72.

Mirelman, D., Galun, E., Sharon, N. & Lotan, R. (1975) Nature 256, 414-416.

Misaki, A. & Goldstein, I. J. (1977) J. Biol. Chem. 252, 6995-6999.

Mishkind, M., Raikhel, N. V., Palevitz, B. A. & Keegstra, K. (1982) J. Cell Biol. 92, 753-764.

Morris, C. J. O. R. & Morris, P. (1976) in Separation Methods in Biochemistry, 2nd Ed., pp 853-873, Halsted Pres, New York.

Nagata, Y. & Burger, M. M. (1974) J. Biol. Chem. 249, 3116-3122.

Namen, T. & Hapner, K. D. (1979) Biochem. Biophys. Acta 580, 198-209.

Neurohr, K. J., Young, N. M. & Mantsch, H. H. (1980) J. Biol. Chem. 255, 9205-9209.

Nicolson, G. (1974) Int. Rev. Cytol. 48, 90-190.

Ouchterlony, O. (1948) Ark. Kemi Miner. Geol. 26B, 1-9.

Paau, A. S., Leps, W. T. & Brill, W. J. (1981) Science 213, 1513-1514.

Pisano, J. J. & Bronzert, T. J. (1969) J. Biol. Chem. 244, 5597-5607.

Poretz, R. D. & Goldstein, I. J. (1970) Biochemistry 9, 2890-2896.

Reeke, G. N. Jr., Becker, J. W. & Edelman, G. M. (1975) J. Biol. Chem. 250, 1525-1547.

Reisner, A. H., Nemes, P. & Bucholtz, C. (1975) Anal. Biochem. 64 509-516.

Richardson, C., Behnke, W. D., Freisheim, J. H. & Blumenthal, K. M. (1978) Biochem. Biophys. Acta 537, 310-319.

Schatchard, G. (1949) Ann. N.Y. Acad. Sci. 51, 660-672.

Sequeira, L. (1978) Annu. Rev. Phytopathol. 16, 453-481.

Sharon, N. (1975) Complex Carbohydrates, Addison-Wesley Publishing Co., Massachusetts.

Shoham, M., Kalb, A. J. & Pecht, I. (1973) Biochemistry 12, 1914-1917.

Simpson, D. L., Thorne, D. R. & Loh, H. H. (1978) Life Sciences 22, 727-748.

Smyth, D. G. (1967) Methods Enzymol. 11, 214-231.

Stark, G. R. (1977) Methods Enzymol. 47, 129-132.

Talbot, C. F. & Etzler, M. E. (1978) Biochemistry 17, 1474-1479.

Thomas, K. A., Silverman, R. E., Jeng, I., Baglan, N. C. & Bradshaw, R. A. (1981) J. Biol. Chem. 256, 9147-9155.

Ticha, M., Entlicher, G., Kostir, J. V. & Kocourek, J. (1970) Biochem. Biophys. Acta 221, 282-289.

Uhlenbruck, G. (1981) Naturwissenschaften 68, 606-612.

Yamada, S. & Itano, H. A. (1966) Biochem. Biophys. Acta 130, 538-540.

Wang, J. L., Cunningham, B. A. & Edelman, G. M. (1971) Proc. Natl. Acad. Sci. U.S.A. 68, 1130-1134.

Wang, J. L., Cunningham, B. A., Waxdal, M. J. & Edelman, G. M. (1975) J. Biol. Chem. 250, 1490-1502.

Weber, K. & Osborn, M. (1975) in The Proteins (Neurath, H. & Hill, R., Eds.), Vol. I, Academic Press, New York.

Weinzierl, J. & Kalb, A. J. (1971) Fed. Eur. Biochem. Soc. Lett. 18, 268-270.

APPENDIX

Amino Acid Symbols

A	alanine	M	methionine
B	aspartic acid/asparagine	N	asparagine
C	cysteine	P	proline
D	aspartic acid	Q	glutamine
E	glutamic acid	R	arginine
F	phenylalanine	S	serine
G	glycine	T	threonine
H	histidine	V	valine
I	isoleucine	W	tryptophan
K	lysine	Y	tyrosine
L	leucine	Z	glutamic acid/glutamine

MONTANA STATE UNIVERSITY LIBRARIES
stks D378.K847@Theses
Sainfoin lectin : molecular characteriza
RL
3 1762 00169322 3

D378
K847
cop.2

Kouchalakos, R. N.
Sainfoin Lectin:
Molecular
characterization and
amino acid sequence

DATE	ISSUED TO

D378
K847
cop 2