Molecular cloning and sequencing of a cDNA from the grasshopper Melanoplus differentialis
by Matthew Craig Rognlie

A thesis submitted in partial fulfillment of the requirements for the degree of Master of Science in Biochemistry
Montana State University

Abstract:
Lectins are a significant component of the molecular defense response of invertebrates. The carbohydrate-binding capacity of these proteins may mediate recognition of foreign substances thereby targeting them for humoral or cellular defensive reactions. The hemolymphatic agglutinin (GHA) in the grasshopper Melanoplus differentialis has undergone initial biochemical characterization. The lectin appears to function in the in vivo and in vitro recognition of certain foreign particles including fungal blastospores.

The objective of this study was to determine the amino acid sequence of GHA. The difficulty of isolating large amounts of protein and the availability of an antibody probe led to the decision to use molecular cloning procedures for pursuit of this goal.

Fat body, the site of GHA synthesis, was dissected from the insect and used for isolation of poly (A+) rich RNA. Double stranded cDNA was synthesized from this preparation, and cloned into the expression vector λgtll to produce cDNA libraries. These libraries were screened with the anti-GHA antibody to detect phage plaques containing an antigenic β-galactosidase fusion protein. A positively reacting clone was isolated and shown to contain a 300 bp insert. This cDNA was labeled with [α-32P]dCTP via nick-translation and used to screen remaining libraries in search of a longer cDNA containing the coding region of GHA.

An 875 bp cDNA was isolated and its nucleotide sequence determined with dideoxy nucleotide-mediated chain termination procedures. The sequence was searched for open reading frames that could correspond to GHA, but such a coding region was not obvious. The cDNA sequence and possible translations were compared to other invertebrate lectins and to sequence databases. Minimal homology with other protein and nucleic acid sequences has so far been observed. Amino acid sequence of two tryptic peptides isolated from GHA were also compared to deduced amino acid sequences from the isolated 875 bp cDNA. This comparison revealed a partial identity. The exact identity of the 875 bp cDNA is not clear based on the analyses of the nucleotide sequence. Further examination is required to confirm correlation to a GHA protein.

MOLECULAR CLONING AND SEQUENCING OF A cDNA FROM THE GRASSHOPPER *Melanoplus differentialis*

by

Matthew Craig Rognlie

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

of

Master of Science

in

Biochemistry

MONTANA STATE UNIVERSITY
Bozeman, Montana

April 1991

APPROVAL

of a thesis submitted by

Matthew Craig Rognlie

This thesis has been read by each member of the thesis committee and has been found to be satisfactory regarding content, English usage, format, citations, bibliographic style, and consistency, and is ready for submission to the College of Graduate Studies.

April 25 1991
Date

Kenneth D Hagen
Chairman, Graduate Committee

Approved for the Major Department

April 25, 1991
Date

Edwin H. Abbott
Head, Department of Chemistry

Approved for the College of Graduate Studies

May 2, 1991
Date

Henry J Parsons
Graduate Dean

STATEMENT OF PERMISSION TO USE

In presenting this thesis in partial fulfillment of the requirements for a master's degree at Montana State University, I agree that the Library shall make it available to borrowers under rules of the Library. Brief quotations from this thesis are allowable without special permission, provided that accurate acknowledgment of the source is made.

Permission for extensive quotation from or reproduction of this thesis may be granted by my major professor, or in his absence, by the Dean of Libraries when, in the opinion of either, the proposed use of the material is for scholarly purposes. Any copying or use of the material in this thesis for financial gain shall not be allowed without my written permission.

Signature Matthew C. Rognli

Date 4-22-91

ACKNOWLEDGMENTS

I wish to thank my major professor, Dr. Kenneth D. Hapner, for his guidance and support throughout my studies. I thank my Graduate Committee for their input: Dr. Samuel J. Rogers, Dr. William E. Dyer and Dr. John R. Amend. I have also depended upon others, too many to mention, for expert advice. My family, and my wife Juli, have supported me and made my studies possible.

Above all, I acknowledge the Lord Jesus Christ with whom all things are possible, including the scientific study of His creation.

TABLE OF CONTENTS

TABLE OF CONTENTS-Continued

TABLE OF CONTENTS-Continued

LIST OF TABLES

LIST OF FIGURES

LIST OF FIGURES - Continued

ABSTRACT

Lectins are a significant component of the molecular defense response of invertebrates. The carbohydrate-binding capacity of these proteins may mediate recognition of foreign substances thereby targeting them for humoral or cellular defensive reactions. The hemolymphatic agglutinin (GHA) in the grasshopper Melanoplus differentialis has undergone initial biochemical characterization. The lectin appears to function in the in vivo and in vitro recognition of certain foreign particles including fungal blastospores.

The objective of this study was to determine the amino acid sequence of GHA. The difficulty of isolating large amounts of protein and the availability of an antibody probe led to the decision to use molecular cloning procedures for pursuit of this goal.

Fat body, the site of GHA synthesis, was dissected from the insect and used for isolation of poly(A+) rich RNA. Double stranded cDNA was synthesized from this preparation, and cloned into the expression vector λgt11 to produce cDNA libraries. These libraries were screened with the anti-GHA antibody to detect phage plaques containing an antigenic ß-galactosidase fusion protein. A positively reacting clone was isolated and shown to contain a 300 bp insert. This cDNA was labeled with [α-^{32}P]dCTP via nick-translation and used to screen remaining libraries in search of a longer cDNA containing the coding region of GHA.

An 875 bp cDNA was isolated and its nucleotide sequence determined with dideoxy nucleotide-mediated chain termination procedures. The sequence was searched for open reading frames that could correspond to GHA, but such a coding region was not obvious. The cDNA sequence and possible translations were compared to other invertebrate lectins and to sequence databases. Minimal homology with other protein and nucleic acid sequences has so far been observed. Amino acid sequence of two tryptic peptides isolated from GHA were also compared to deduced amino acid sequences from the isolated 875 bp cDNA. This comparison revealed a partial identity. The exact identity of the 875 bp cDNA is not clear based on the analyses of the nucleotide sequence. Further examination is required to confirm correlation to a GHA protein.

INTRODUCTION

Invertebrate Lectins

The insect defense system is well adapted for protection against foreign pathogens. This system is composed of physical barriers (i.e., cuticle and peritrophic membrane) and cellular and humoral defensive responses. Cellular response consists of encapsulation of foreign bodies (Karp, 1990) and/or phagocytosis (Bayne, 1990). Humoral response generally involves antibacterial factors including lysozyme and cecropin and attacin-like proteins (Dunn, 1990). Insects, unlike vertebrates, are thought not to be capable of 'learning' a response to invasion. This notion, however, has recently been challenged (Karp, 1990). Insects possess the capability to recognize foreignness or 'non self' as differing from 'self'. The molecular basis for recognition and discrimination is an unsolved fundamental problem in the study of the insect immune/defense system (Lackie, 1981).

Agglutinins (lectins) have been implicated as an important factor in non-self recognition in invertebrates (Dunn, 1986; Ratcliffe et al., 1985). These proteins or glycoproteins bind to cells and glycoconjugates according to their carbohydrate binding capability. Agglutinins are ubiquitous among organisms, and are present in plants, animals

and microorganisms (Lis and Sharon, 1986). A recent review (Sharon and Lis, 1989) describes the role of lectins in many organisms, indicating that the proteins are important in various normal and pathological processes.

The ability of lectins to bind cell surface glycoconjugates may contribute to the recognition of non-self materials in invertebrates. In fact, the lectin from Sarcophaga peregrina may play a dual role, important for both morphogenic development and defense reactions (Natori, 1990). Although models have been presented concerning roles of agglutinins in arthropods, the physiological function of these proteins for defense purposes is not clear (Lackie, 1988).

The biochemistry of grasshopper hemagglutinin (GHA) from Melanoplus differentialis and Melanoplus sanguinipes has been studied. The hemolymphatic agglutinin agglutinates human and some other animal erythrocytes (Jurenka et al., 1982; Hapner, 1983). It is a high-molecular weight glycoprotein whose activity is inhibited by D-glucosides, D-galactosides or EDTA (Stebbins and Hapner, 1985). The EDTA inhibition is due to the lectin's carbohydrate binding dependence on Ca^{++}. The characteristic carbohydrate inhibition of agglutination suggests that carbohydrate binding specificity is directed toward glucosidic and galactosidic glycoconjugates.

The protein is a 600-700,000 mw noncovalent aggregate of 70,000 mw units. The monomer unit was shown to contain two covalently attached subunits of 40,000 and 28,000 mw (Stebbins

and Hapner, 1985), but recent evidence suggests that it is a homodimer of two identical subunits of 35,000 mw and an uncharacterized smaller protein coisolate (Hapner, unpublished results).

Agglutinin from M. differentialis is synthesized mainly in the fat body with smaller amounts in ovary and testes tissues as determined by metabolic labeling studies (Stiles et al., 1988). GHA is associated with the outer membrane of a fraction of granular hemocytes (Bradley et al., 1989).

In studies of the action of agglutinin, it was found that grasshopper agglutinin did not act as an opsonin or recognition molecule between certain foreign particles and hemocytes (Bradley et al., 1989). Recently, however, the agglutinin has been shown to opsonize fungal blastospores toward hemocytic phagocytosis and in vivo clearance (Hapner, unpublished results).

The amino acid sequence of lectins from four invertebrates has been determined. A comparison of their biochemical characteristics and that of GHA is included in Table 1.

The lectin from the flesh-fly Sarcophaga peregrina (Sarcophaga lectin) is present in hemolymph of induced or injured larvae and is a 190,000 mw heterohexamer with α and β subunits of 32,000 and 35,000 mw, respectively (Komano et al., 1980). This is the only insect lectin whose sequence has been determined by recombinant DNA methods (Takahashi et al.,

Table 1. Invertebrate Lectins of Known Structure.

Lectin	Denatured size(mw)	Monomer size(mw)	Ca^{++} dependence?	Specificity
BRA-2	140,000	22,000	Yes	Galactose
BRA-3	64,000	16,000	Yes	Galactose
ECH	26,000	13,000	Yes	Galactose
TUN	NA	15,000	Yes	Galactose
FLY	190,000	32,000-α 30,000-ß	Yes	Galactose
GHA	70,000	35,000	Yes	Galactose

Abbreviations: BRA = lectins from Megabalanus rosa; ECH = lectin from Anthocidaris crassispina; TUN = lectin from Polyandrocarpa misakiensis; FLY = lectin from Sarcophaga peregrina.

1985). The coelomic fluid of the barnacle Megabalanus rosa contains three different molecular species of lectins, and the amino acid sequences of two have been determined. BRA-2 (Muramoto and Kamiya, 1990) and BRA-3 (Muramoto and Kamiya, 1986) are 140,000 and 64,000 mw homooligomers of 22,000 and 16,000 mw subunits, respectively. Echinoidin, a lectin found in the coelomic fluid of the sea urchin Anthocidaris crassispina is a 300,000 mw aggregate of 26,000 mw homodimers (Giga et al., 1985). Complete amino acid sequence of this lectin was determined by conventional means (Giga et al., 1987). Another lectin has been isolated and sequenced from homogenates of the tunicate Polyandrocarpa misakiensis. This lectin exists as a monomer of 15,000 mw (Suzuki et al., 1990).

These five sequenced lectins exhibit homology at the structural as well as the biochemical level. Figure 1 displays a gap-introduced amino acid sequence alignment of

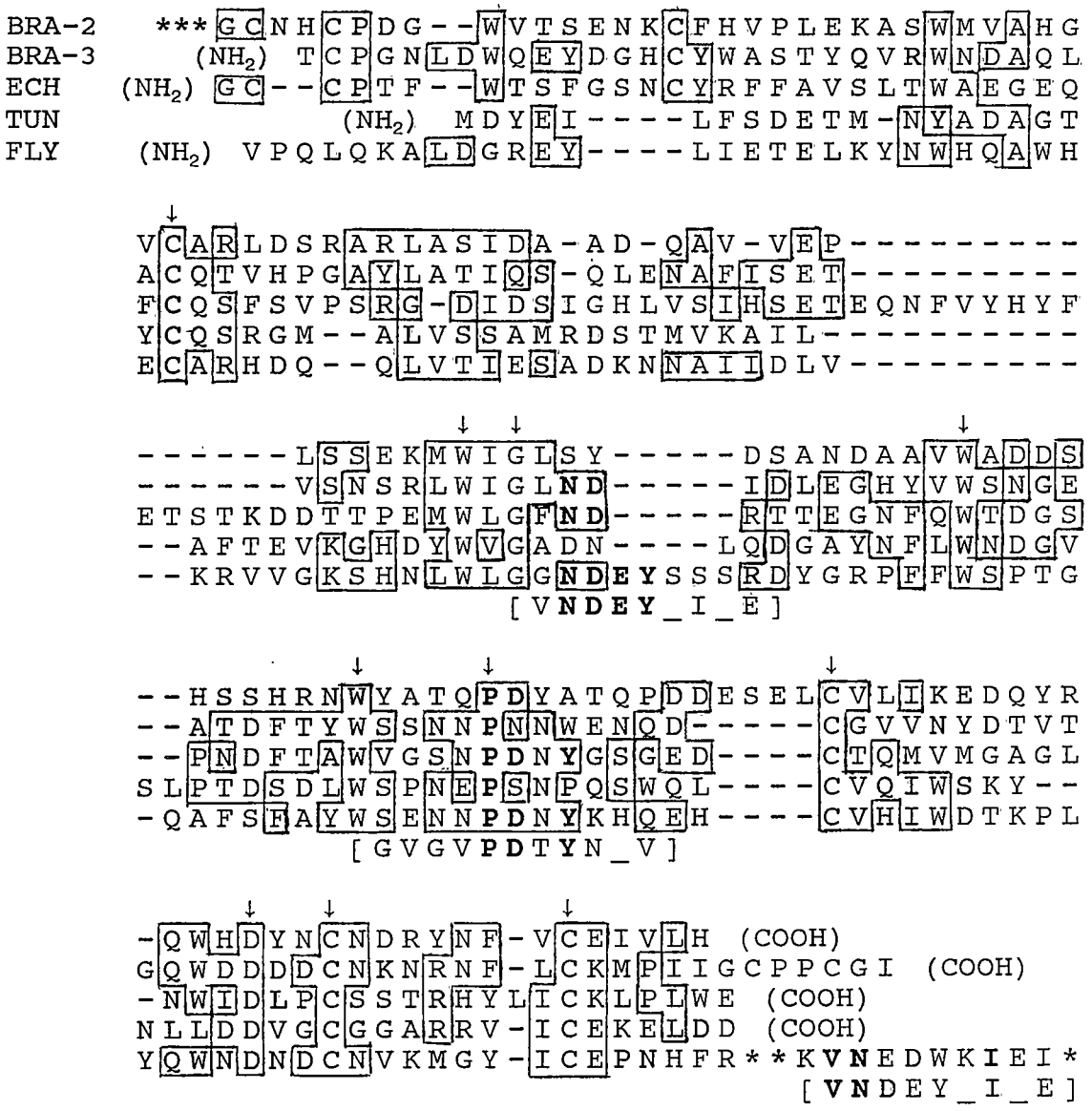

Figure 1. Alignment of Invertebrate Lectin Sequences. See Table 1 for abbreviations. Complete sequences are shown, except for BRA-2 which contains an additional 45 amino acids toward the amino terminus (***), and FLY containing an additional 60 amino acids toward the carboxyl terminus (*). Two asterisks in FLY represent 44 residues. Boxes denote sequence identity in two or more lectins. Bold figures indicate identity with two tryptic peptides of GHA, that are enclosed in brackets (underscore refers to unknown residues). Arrows mark ten invariant residues among the molecules.

the five lectins modified from K. Muramoto and H. Kamiya (1990), using the one-letter amino acid notation. Boxes enclose regions of sequence identity.

The amino acid sequences of two tryptic peptides isolated from GHA are shown in brackets (Figure 1) (Hapner, unpublished results). These two segments exhibit possible similarity at three separate positions in the Sarcophaga lectin sequence (bold residues, Figure 1). It is presumed that structural homologies may be at least as great between the fly lectin and GHA, as that among the other lectins.

Rationale for the Research

The rationale for the work to be described is based on two general assumptions. (1) Because of its importance in various defensive responses, study of the insect lectin protein will contribute to a better understanding of insect defensive physiology and invertebrate immune systems. (2) New knowledge regarding the insect defense system may be applied to nonchemical insect management programs. Applications could include propagation of useful insects as well as bio-control of harmful species. It is therefore reasonable to study the agglutinin protein in the grasshopper.

It was decided that primary structural characterization of the lectin would contribute to understanding of its biochemical functions. This structural information could be compared to other related proteins to infer additional

characteristics of the grasshopper lectin. GHA is present in low amounts in the hemolymph of the insects, and isolation of a sufficient quantity for direct amino acid sequence determination would be difficult and time consuming. Therefore, use of molecular cloning techniques to determine a cDNA nucleotide sequence was chosen for deduction of the protein sequence. Besides allowing the structural comparisons mentioned above, isolation of the cDNA for GHA will also promote continuing studies of the molecular biology of the GHA system, including gene structure, copy number, and expression characteristics.

Research Objectives

The primary objective of this project is to determine the nucleotide sequence of a cDNA corresponding to the hemagglutinin from the grasshopper Melanoplus differentialis. Upon completion of this objective, the study will be extended to include deduction of the amino acid sequence of GHA, and comparison of cDNA and protein sequences to other invertebrate lectins and to sequence databases.

Approach to the Problem

A fat body cDNA library in bacteriophage λgt11 was provided by Dr. K.D. Hapner. Plaques produced from this library will be screened to isolate a clone that produces an antigenic fusion protein. A rabbit anti-GHA polyclonal

antibody is available, and can be used for this screen and plaque purification. Isolation of the cDNA from this antigenic clone will follow. This cDNA can then be used to probe remaining library solutions for a full-length cDNA corresponding to GHA. Nucleotide sequence determination of the full-length cDNA will be determined.

The identity of the deduced primary structure with GHA will be authenticated by two means. First, the determined structure should contain sequence corresponding to the two tryptic peptides isolated from GHA, and secondly, amino acid homology with other known invertebrate lectins should be apparent.

MATERIALS AND METHODS

cDNA Library Construction

This section of experimentation, cDNA library construction, was carried out previously by other laboratory personnel (Hapner, unpublished results) and formed the beginning point for research reported in this thesis. Formation of the λgt11 cDNA libraries is briefly described here for completeness.

Insects

Melanoplus differentialis grasshoppers were raised by the USDA Rangeland Insect Laboratory (Montana State University) from permanent colonies. They were maintained on a diet of bran and ryegrass. Fifth instar nymphs or adult females were used for dissection of fat body.

Dissection of Fat Body

Insects were cold-anesthetized and surface sterilized with liquid detergent, bleach and 70% ethanol, followed by sterile water. The dorsal surface was cut open, pinned back and the gut removed. Fat body lining the body wall was removed with forceps under a dissection microscope, dipped in pH 7 phosphate-buffered saline with 1 mg/ml glutathione, and

immediately placed in a small tared petri dish cooled on solid CO_2.

Isolation of Total RNA

Isolation of total cellular RNA was done by the guanidinium thiocyanate (GdSCN) extraction method (Chirgwin et al., 1979). Fat body was homogenized in GdSCN extraction buffer. Debris were removed by centrifugation, then total RNA was pelleted by ultracentrifugation through a cushion of 4 M cesium chloride.

Poly(A+) RNA Selection and Purification

Poly(A+) RNA was isolated from purified total RNA by affinity chromatography on an oligo(dT) cellulose column (Edmonds et al., 1971; Aviv and Leder, 1972). Concentration and purity were determined by spectrophotometry.

cDNA Library Construction

Synthesis of cDNA. cDNA synthesis and cloning into bacteriophage λgt11 (Young and Davis, 1983) was accomplished using a λgt11 cloning kit (Bethesda Research Laboratories, Gaithersburg MD). First and second strand cDNA synthesis was carried out based on a method by Gubler and Hoffman (1983) according to manufacturer's instructions. The cDNA formed was blunt-ended using T4 DNA polymerase.

Purification of cDNA. The cDNA preparation was applied to a standardized Sepharose 4B column for purification and size separation. The cDNA molecules were separated into two size aliquots of less than, and greater than 600 bp.

Ligation into λgt11 DNA. An EcoRI Linker Ligation kit (Promega, Madison WI) was used to ligate cDNAs into λgt11 DNA. The fractionated cDNA was treated with EcoRI methylase, then 10mer EcoRI linkers were ligated to the cDNA ends using T4 DNA ligase. Digestion with EcoRI followed, to produce cDNA with EcoRI 'sticky' ends. The cDNAs were ligated into EcoRI-cleaved λgt11 DNA using T4 DNA ligase.

Packaging into Viable Phage Particles. The Promega Packagene® system was used to assemble recombinant λgt11 DNA into viable phage. Five different samples were prepared from both aliquots of purified cDNA, and nine separate cDNA libraries were produced. Packaging efficiency was determined by infecting E. coli Y1090 and plating on LB plates with 100 μg/ml ampicillin.

Analysis of Bacteriophage Library

Recipes of all media, buffers and reagents mentioned in the following text may be found in Appendix I. All chemicals used were reagent grade and water used was 15 MΩ pure (Millipore, Bedford MA).

Western Blotting

Activity and specificity of antibodies toward agglutinin was analyzed by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) (Laemmli, 1970) and Western blotting procedures (Towbin et al., 1979). SDS-PAGE was done using a BioRad (Richmond CA) Protean apparatus. Discontinuous slab gels (16 cm x 18 cm) were made, with 4% acrylamide stacking and 7.5% resolving portions. Gels were run at 30mA until the bromophenol blue marker dye exited the gel (ca. three hours). Gels were removed to a BioRad transfer apparatus, and proteins were transferred to nitrocellulose by manufacturer's instructions at 100 V for two hours. The nitrocellulose was then washed in TBST for five hours, and placed in a Hoefer PR200 Decaprobe multichannel apparatus to allow probing of each lane with separate antibodies. Incubation of transferred proteins with primary antibodies for 12 hours followed. The nitrocellulose was removed, rinsed and transferred to a 1:2000 dilution of alkaline phosphatase (AP) conjugated secondary antibody for 1.5 hours, followed by development with 4-bromo-3-chloro-2-indolyl-phosphate/nitro-blue tetrazolium (BCIP/NBT; 167 μg/ml and 330 μg/ml in AP buffer, respectively). Development times varied (usually 2-10 min.), and development was stopped by rinsing in H_2O.

Antibody Purification

Rabbit anti-agglutinin #152 polyclonal antibody, previously purified by Affligel-Blue chromatography, was further purified by affinity chromatography. This purification was done to remove antibodies cross-reacting with E. coli protein (de Wet et al., 1984). A crude protein extract from an E. coli Y1090 culture was prepared according to Sambrook et al. (1989). Bacterial cells harvested from a two liter culture were resuspended in 100 ml coupling buffer. Hen egg white lysozyme (200 mg) was added and the suspension was incubated 20 min. at room temperature. Two milligrams of pancreatic DNAase I and 200 µl of Triton X-100 were added, followed by incubation for one hour at 4°C. The solution was then brought to pH 9.0.

The concentration of protein in the lysate was determined by the Pierce BCA method (Pierce, Rockford IL). Ten milligrams of E. coli lysate protein was bound to 0.3 g CNBr activated Sepharose 4B as per manufacturer's directions (Pharmacia, Piscataway NJ), and the washed complex used for purification of antibody solutions.

Screening of Bacteriophage Plaques

Guidelines for immunological screening of bacteriophage plaques were found in Mierendorf et al. (1987) and were based on a procedure developed by Benton and Davis (1977). A single colony of E. coli strain Y1090 or LE392 was picked from a stock culture on an LB plate containing 100 µg/ml ampicillin,

and grown overnight, with shaking, at 37°C in LB broth with 0.2% maltose and 10 mM $MgSO_4$. This culture (100 µl) was infected with about 1 x 10^4 plaque forming units (pfu) from a bacteriophage library for 20 min. at 37°C, added to four ml of molten (50°C) LB with 0.6% agarose (top agarose) and poured on a 90 mm LB agar plate. Plates were incubated for 3.5 hours at 42°C, overlayed with 10 mM isopropyl-ß-D-thiogalactoside (IPTG)-soaked 0.45 µm nitrocellulose filters (MSI, Westboro MA) and incubated for 3.5 hours at 37°C. The filters were removed and washed twice for 20 min. in TBST.

Non-specific binding sites on the filters were blocked by washing in TBST/1% gelatin for two hours at room temperature. Filters were then incubated for 1.5 hours in TBST containing 10 µg/ml rabbit anti-agglutinin polyclonal antibody to detect antigenic ß-galactosidase fusion protein. Nonspecifically bound antibody was removed by subsequent washing in TBST (3X 15 min.).

Positive plaques were detected by incubation for 40 min. with AP conjugated goat anti-rabbit polyclonal antibody (BioRad) at 1:1000 in TBST. Unbound conjugate was removed by washing in TBST. Bound conjugate antibody was detected with BCIP/NBT reagent, in a manner similar to Western blotting.

Isolation and Purification of Bacteriophage Clones

Positive plaques, as determined by immunological screening, were isolated by removing a plug from the agar

plate which contained the plaque of interest, and transferring it to one milliliter SM buffer. Another round of immunological screening on low plaque density plates was done to determine purity of sample. If some plaques were not positive, another single positive was picked and re-plaque purified. This procedure was continued until screening indicated that all plaques derived from a single isolated plaque were positive (plaque purification).

Analysis of Bacteriophage DNA

Purification of Phage DNA

E. coli strain Y1090 was used to prepare confluent phage plaques in top agarose (1×10^4 to 1×10^5 plaques). Phage DNA was purified by a phage miniprep/plate lysate method (Sambrook et al., 1989). Top agarose was scraped from the plates into SM buffer (5 ml/plate) and incubated for 30 min. at room temperature with gentle agitation. Debris was removed by centrifugation at 4°C at 8000 x g for 10 min. The supernatant was incubated for 30 min. at 37°C in 1 μg/ml RNAase H and 1 μg/ml DNAase I to destroy host nucleic acid. Bacteriophage was precipitated in 10% polyethylene glycol (PEG 8000) and 1 M NaCl for 12 hours at 0°C and collected at 9000 x g for 20 min at 4°C. The pellet was dissolved in SM (0.5 ml/plate) and debris removed at 8000 x g for 10 min at 4°C. Phage were disrupted in 0.5% SDS and 5 μM EDTA at 68°C for 15 min. The DNA was purified by phenol extraction, precipitated in

isopropanol, and dissolved in water with 0.2 mg/ml RNAase H. Yields of DNA from the phage miniprep were determined by spectrophotometry at 260 nm (A_{260} of 1.0 = 50 μg DNA/ml).

Identification of Inserts

The DNA solution was analyzed by horizontal agarose gel electrophoresis in a Fisher (Pittsburgh PA) horizontal electrophoresis apparatus (7 cm x 10 cm). Agarose gels, 0.6% to 2.0%, in 0.5X TBE were used throughout, depending on DNA size range to be analyzed. Ethidium bromide (0.5 μg/ml) was added to molten agarose prior to electrophoresis. Electrophoresis was done at 4 V per cm electrode path. A pure sample displayed a single band upon visualization with 254 nm ultraviolet light, with no other DNA present. The presence and size of cDNA inserts were analyzed by cleavage with restriction endonucleases followed by agarose gel electrophoresis. Migration distances were compared with that of standard DNA markers (Bethesda Research Laboratories).

Purification of Inserts

cDNA inserts removed from vector DNA by restriction endonuclease digestion were separated by agarose gel electrophoresis in a preparative horizontal gel unit (13 cm x 19 cm). Isolation of the fragment from the gel was based on a procedure by Ogden and Adams (1987). A well was cut completely through the gel immediately in front of the band, and filled with 0.5X TBE. The end of the well nearest the

anode was lined with dialysis membrane (15,000 mw) to prevent migration of DNA beyond the well. Electrophoresis was resumed until the band migrated completely into the well. Polarity of the chamber was reversed and electrophoresis continued for 60 seconds to desorb DNA from the dialysis membrane. The buffer containing the DNA was removed, extracted with phenol and precipitated with two volumes of ethanol in 375 mM sodium acetate. Purity and yield of the isolated DNA band was determined by agarose gel electrophoresis.

Subcloning of cDNAs into Plasmid Vector

Ligation into Plasmid Vector

The plasmid vector pGEM®-7Zf(+) (Promega) was digested with appropriate restriction enzymes to provide ends compatible with the insert to be subcloned, and subsequently purified by phenol extraction and ethanol precipitation. Insert and vector DNA was combined in a 3:1 molar ratio and heated to 65°C for five min. to melt annealed ends. One unit of T4 DNA Ligase (Promega), 10X ligase buffer (0.1 vol.) and 0.1 mg/ml bovine serum albumin (fraction V, Sigma, St. Louis MO) were added to the DNAs, and the ligation reaction was allowed to proceed for four hours to overnight at 16°C.

Transformation of Bacteria

E. coli strain JM101 was rendered competent by the calcium chloride method (Cohen et al., 1972) and transformed

with recombinant plasmid from the ligation reaction product. A single JM101 colony was picked from an M9 minimal plate, and grown in 50 ml LB until the $O.D._{600}$ = 0.15 to 0.20 (ca. four hours). The cells were cooled on ice, and harvested at 4000 x g for 10 min. at 4°C. Cells were resuspended in 10 ml of 0°C 0.1 M $CaCl_2$, and incubated between 30 min. and three hours. The cells were again harvested and resuspended in 2 ml 0.1 M $CaCl_2$ and incubated 12 to 24 hours on ice.

Transformation of bacterial hosts with plasmid DNA utilized procedures described by Sambrook, et al. (1989). DNA (50 ng) was used to transform 200 µl of the competent cells. The DNA was added to the cells, and the suspension was incubated for 30 min. on ice. The cells were heat shocked at 42°C for 90 seconds, cooled on ice for three min., and allowed to recover in 800 µl SOC medium at 37°C for 45 min. The solution was plated on LB plates containing: 100 µg/ml ampicillin, 0.5 mM IPTG and 40 µg/ml X-Gal (5-bromo-4-chloro-3-indolyl-ß-D-galactoside). Bacteria harboring a recombinant plasmid were able to grow, due to plasmid-derived ß-lactamase activity. These colonies appeared white because the lacZ α-peptide plasmid sequence had been interrupted by the inserted DNA. Bacteria containing a ligated, non-recombinant plasmid (with the lacZ α-peptide sequence intact) yielded blue colonies and were ignored.

Preparation and Analysis of Plasmid DNA

A plasmid miniprep protocol utilizing PEG precipitation of plasmid DNA was performed (Sambrook et al., 1989) to produce sequenceable plasmid DNA. Lysis of host cells was carried out by a modification of procedures developed by Birnboim and Doly (1979) and Ish-Horowicz and Burke (1986). Positive (white) transformed JM101 colonies were picked and grown in LB with 100 μg/ml ampicillin overnight at 37°C with shaking. A 1.5 ml microcentrifuge tube was filled with the culture, and the cells were collected by centrifugation at 13,000 x g for 30 sec. Cells were resuspended in 100 μl 0°C Solution I, and 200 μl of r.t. Solution II was added to lyse the cells and the solution was incubated 5 min. on ice. Solution III (150 μl - 0°C) was added to precipitate protein and genomic DNA and incubated for 5 min. on ice. The precipitate was removed, and the DNA remaining in the supernatant was ethanol precipitated. The DNA was redissolved in 100 μl TE (pH 7.6), and 100 μl 5 M LiCl (0°C) added to precipitate high-molecular weight RNA, which was removed by centrifugation. Isopropanol (200 μl) was added to precipitate DNA, and the DNA was then redissolved in 200 μl TE with 20 μg/ml RNAase H and incubated for 30 min. at room temperature. Plasmid DNA was precipitated by addition of 200 μl of 1.6 M NaCl/13% PEG at 0°C. The DNA was pelleted and redissolved in TE, phenol extracted and ethanol precipitated in the presence

of 375 mM ammonium acetate. The final plasmid DNA solution was subject to agarose gel, restriction and sequence analyses.

Sequencing of Recombinant Plasmid DNA

Sequencing Reactions

Determination of recombinant plasmid nucleotide sequence was carried out by the Sanger method of dideoxy nucleotide-mediated chain termination (Sanger et al., 1977). The enzymes used were avian myeloblastosis virus reverse transcriptase (RT) and Klenow fragment of E. coli DNA polymerase I (K), both from the Promega K/RT® sequencing system. Sequenase® 2.0 modified T7 DNA polymerase (United States Biochemical, Cleveland OH) was also used for some sequencing reactions.

One to four μg of plasmid DNA was denatured in 0.2 N NaOH/0.2 mM EDTA. The solution was neutralized in 375 mM ammonium acetate and the denatured plasmid DNA precipitated in ethanol. Sequencing reactions were done as per manufacturer's instructions. The label, 20 mM [α-^{35}S]dATP (10 mCi/ml), was incorporated into the newly synthesized strand. Reactions were stored at -50°C until electrophoretic analysis.

Electrophoresis and Autoradiography

Polyacrylamide electrophoresis sequencing gels were constructed in the 21 cm x 50 cm BioRad SequiGen® Nucleic Acid Sequencing Cell. Standard gels used were 8% acrylamide (19:1 acrylamide:bis-acrylamide) and 8M urea in 1X TBE. Gels were

pre-run at 2300 V and 50 W until gel temperature reached 50°C. Sequencing reaction samples (2.5 μl) were loaded, and two and a half hours later a second loading of each reaction on the gel was done to maximize sequence data. The gel was run an additional 2 to 2.5 hours. Electrophoresis was maintained at a constant 2300 V (50 W) and gel kept at 50°C for the duration of the run. After electrophoresis, gels were fixed in 10% methanol and 10% acetic acid for 20 min. then dried onto Whatmann 3MM chromatography paper at 80°C for one hour under vacuum. Dried gels were exposed to X-ray film (Kodak X-OMAT; 35 cm x 43 cm) for about 36 hours at room temperature.

Sequence Analysis

Autoradiographic sequence data were analyzed with two computer software packages. Genepro (Riverside Scientific, Bainbridge Is. WA) was used for alignment with selected nucleic acid sequences, and for translational, restriction site and open reading frame analyses. The GCG (University of Wisconsin, Madison WI) software package was used for Pearson and Lipmann (Pearson and Lipmann, 1988) protein and nucleotide sequence alignment against GenBank, EMBL and NBRF sequence databases, and other analyses. The GCG software and databases were accessed by TelNet communications link with the Pittsburgh Supercomputing Facility.

Library Screening with Radioactive Probe

Preparation of Probe

The insert containing the 300 bp cDNA was removed from the recombinant plasmid by restriction endonuclease digestion and gel purified. This preparation was used for nick translation using a kit from Promega. cDNA (0.5 μg) was combined with DNAase I (0.5 ng), DNA polymerase (2.5 units), reaction buffer, 100 μM each of dGTP/dTTP/dATP, and 7 μl [α-^{32}P]dCTP (10 mCi/ml at 3000 Ci/mmol) in a volume of 25 μl. Reactions were allowed to proceed at 15°C for one hour, and were stopped by addition of EDTA. Percent incorporation was determined by comparison of liquid scintillation counting results from the reaction product and trichloroacetic acid-precipitated reaction product. The product was purified of excess labeled dNTP's by spun-column chromatography through a Sephadex P-20 matrix. The purified probe was used for screening experiments.

Plaque Growth and Transfer

Plaques of library bacteriophage were prepared in top agarose on LB plates. Plaques were blotted onto dry nitrocellulose circles (92 mm, MSI), denatured in 0.5 N NaOH/1.5 M NaCl, neutralized in 0.5 M Tris (pH 7.4)/1.5 M NaCl, dried at 80°C in a vacuum oven for one hour, and incubated for 1.5 hours at 42°C in hybridization buffer. Denatured, nick translated cDNA probe (2 x 10^5 cpm/ml) was

added to the buffer and filters, and incubated at 50°C for 5-12 hours. Filters were washed in 2X SSPE/0.1% SDS (4X 10 min. at 22°), followed by two stringent washings in 1X SSPE/0.1% SDS at 68°C for one hour each to remove all probe not bound to a DNA complement. Lifts were exposed to X-ray film (Kodak X-OMAT) backed by an intensifying screen overnight at -50°C. Positive plaques were isolated and purified in similar manner as used for immunological screening.

PCR Amplification of cDNAs from Phage Plaques

Polymerase Chain Reactions

AmpliTaq® DNA polymerase (Perkin-Elmer Cetus, Norwalk CT) (Saiki et al., 1988) was used according to manufacturer's suggestions. One unit of AmpliTaq, 0.5 μM of both primers, 1.25 mM dNTP's, 10X buffer and 10^4-10^5 plaque forming units of a phage suspension were combined in one 25 μl reaction. Reactions were denatured at 95°C and subjected to 30 rounds of the following cycle in a thermal cycler (Perkin-Elmer Cetus): denaturing for 1 min. at 95°C, annealing of primer for 2 min. at 55°C, and extension of primer for 4 min. at 72°C.

Analysis of Amplified cDNAs

The sizes of the PCR products were analyzed by agarose gel electrophoresis Southern blotting techniques (Southern, 1970). After electrophoresis, the DNA bands were transferred to nitrocellulose by capillary transfer in 15X SSPE. The DNA

bound to the nitrocellulose was screened with the radioactive probe in an identical procedure to plaque screening, to confirm amplification of the desired cDNA. PCR products were digested with EcoRI to remove the incorporated primers, gel purified, and subcloned into a plasmid vector for preliminary sequence analysis by procedures mentioned previously.

Oligonucleotide Primer Construction

Oligonucleotide primers were synthesized on an ABI model 391 DNA synthesizer (Applied Biosystems Inc., Foster City CA). Manufacturer's supplied protocol was used for this process. Trityl 5' protecting groups were removed and cleavage from the column and removal of other protecting groups was done in 30% NH_4OH. The primers were lyophilized, dissolved in H_2O and used directly in sequencing reactions. Yield was determined spectrophotometrically at 260 nm (A_{260} of 1.0 = 33 μg/ml).

RESULTS

cDNA Library Construction

Total RNA Isolation

Preparations of RNA yielded an intense, high-molecular weight band and no smearing at lower molecular weight, upon denaturing agarose gel electrophoresis and staining with toluidine blue. This indicated that the RNA had not been seriously degraded by contaminating enzymes and was potentially useful for subsequent experimentation (Hapner, unpublished results).

Selection of Poly(A+) RNA

The isolated total RNA (43 mg) was applied to an affinity column of oligo(dT) cellulose to select for poly(A+) rich RNA. Elutions from the column resulted in 225 μg of poly(A+) rich RNA. Translation of the eluant with an *in vitro* translation kit (Promega) and ^{35}S-methionine indicated the presence of translatable material (poly(A+) RNA) (Hapner, unpublished results).

cDNA Library Construction

The poly(A+) rich RNA (5.9 μg) was used for cDNA synthesis. Fractions of double stranded cDNA eluted from the

calibrated Sepharose 4B column were pooled into two aliquots, accounting for about 80% of the material eluting from the column. Packaging of the λgt11 DNA, with cDNA inserts, into viable bacteriophage particles, produced five libraries for the large cDNA fraction, and four libraries for the small cDNA fraction. Titering of the libraries indicated 7.9×10^5 recombinant pfu were contained in the five large libraries, and 1×10^6 recombinant pfu were present in the four small libraries.

Analysis of Bacteriophage Library

Isolation of Positive Clone

Immunological screening of the <600 bp cDNA library at a density of 1×10^4 plaques per 90 mm LB agar plate produced three potential positive plaques. Plaque purification of these indicated that only one plaque was a true positive, producing an antigenic ß-galactosidase fusion protein (Hapner, unpublished results).

Experiments performed to confirm the positive nature of the isolated clone were initially unclear due to a background problem. The #152 primary antibody bound to both positive and known negative plaques making the indication of positives difficult. These results are shown in Figure 2. To resolve this difficulty, the antibody preparation was purified as discussed below.

Agglutinin Antibody Activity Toward E. coli Lysate

A Western blot was carried out on E. coli lysate to observe if the binding of the antibody to known negative λgt11 plaques was due to E. coli protein present from cell lysis. Figure 3 displays the results of this experiment. Cross-reactivity of the GHA antibody with an E. coli protein was confirmed by appearance of an approximately 70,000 mw band in lanes 5 and 6. This experiment also indicated that #152 polyclonal antibody properly recognized both non-reduced and reduced forms of GHA from electrophoresed hemolymph and purified GHA.

Validation of Positive Clone

Immunoscreens were repeated on plaques prepared from the positive phage suspension. The antibody used in the experiment was purified by affinity chromatography on a Y1090 lysate-Sepharose 4B column. Figure 4 compares the immunoscreens of positive and negative plaques. These results show that the purified antibody did not react with negative plaques, and confirmed antigenicity of the isolated and purified phage clone with the GHA antibody.

Analysis of Bacteriophage cDNA Inserts

Preparation of λgt11 DNA

Purified positive phage suspensions were used for preparation of phage DNA. Yields from these DNA preparations

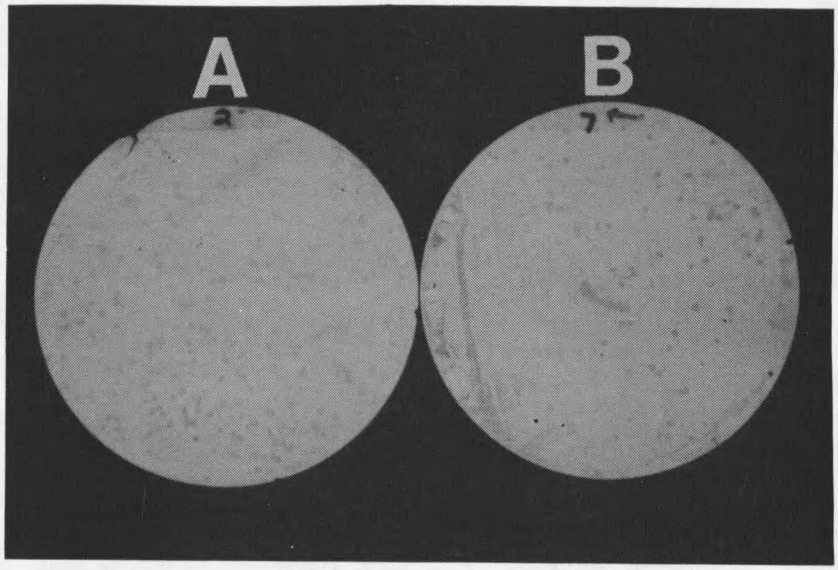

Figure 2. Cross-reactivity of Anti-GHA Antibody With Non-recombinant Phage Plaques (B) and Positive Plaques (A).

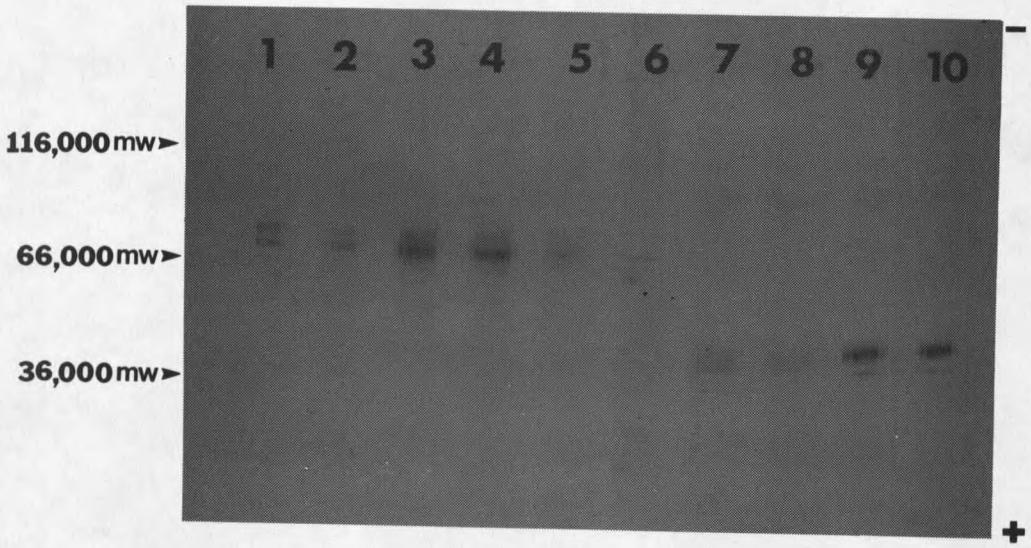

Figure 3. Western Blot of GHA With Agglutinin Antibody. The even lanes are probed with rabbit immune anti-serum. The odd lanes are probed with purified anti-GHA IgG. Reduced samples were boiled in the presence of 1% ß-mercaptoethanol.

-Lanes 1,2 : grasshopper hemolymph
-Lanes 3,4 : purified GHA
-Lanes 5,6 : E. coli lysate
-Lanes 7,8 : reduced hemolymph
-Lanes 9,10: reduced, purified GHA

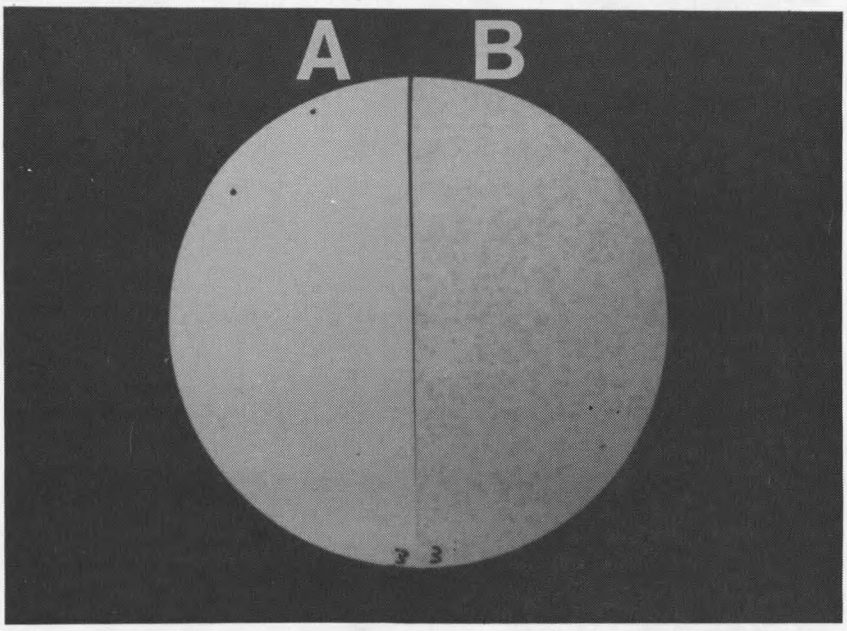

Figure 4. Differentiation of Nonrecombinant Phage Plaques (A) from Positive Plaques (B) With Affinity Purified Antibody. Antibody species cross-reacting with E. coli lysate protein were removed by affinity chromatography (compare with Figure 2), and allowed differentiation of positive and negative plaques.

were generally 5-20 μg phage DNA per plate. The results of an agarose gel analysis of the DNA are shown in Figure 5. The presence of a single band in lane 7 indicates that the preparation was pure and free of bacterial nucleic acids. This single band of DNA migrated slightly slower than a 23 Kb standard. The purified recombinant λgt11 DNA was digested with restriction endonucleases to confirm the presence and size of a cDNA insert.

Identification of cDNA Inserts

DNA isolated from the first purified positive plaque was resistant to cleavage by EcoRI restriction endonuclease. This was unexpected because the cDNA had been cloned into λgt11 DNA at the EcoRI site, and excision with EcoRI should be possible. Alternative enzymes were used to remove the cDNA insert from the λgt11 DNA. Cleavage of λgt11 DNA with both SacI and KpnI restriction endonucleases produces a 1990 bp fragment containing the EcoRI cloning site. Agarose gel analysis of this reaction is found in lane 3 of Figure 5. Lane 4 of the same figure displays the results of KpnI/SacI digestion of the recombinant λgt11 DNA. The 1990 bp fragment shifted to 2300 bp upon SacI/KpnI digestion. This increase in size confirmed the presence of a cDNA insert of approximately 300 bp.

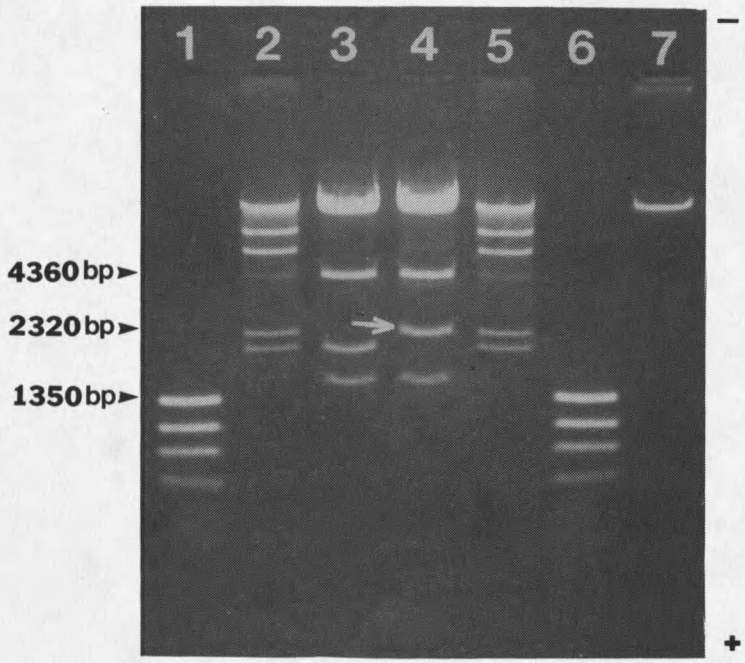

Figure 5. Agarose Gel (0.8%) Analysis of Restricted Native and Recombinant λgt11 DNA: Evidence of a cDNA insert.
-Lanes 1,6: HaeIII-digested ΘX174 DNA standards
-Lanes 2,5: HindIII-digested λ DNA standards
-Lane 3 : SacI/KpnI restricted native λgt11 DNA
-Lane 4 : SacI/KpnI restricted recombinant λgt11 DNA
-Lane 7 : recombinant λgt11 DNA preparation
(arrow indicates a larger KpnI/SacI band (2300 bp) in the recombinant phage preparation)

Gel Purification of Insert

Following the identification of size and presence of the cDNA in the λgt11 DNA, the 2300 bp fragment was isolated. Recombinant phage DNA (40 μg) was subjected to restriction with both KpnI and SacI. The resulting 2300 bp fragment (see arrow, lane 4, Figure 5) was gel purified and yielded 100 ng of purified DNA. This fragment was subcloned into pGEM®-7Zf(+), a plasmid vector designed for double stranded sequence analysis.

Subcloning of cDNAs into pGEM®-7Zf(+)

Ligation into Plasmid Vector

The plasmid vector pGEM®-7Zf(+) was chosen for subcloning of isolated cDNAs because it contains both SacI and KpnI restriction sites in the polylinker. This characteristic of the vector is necessary since the purified cDNA (with attached λgt11 DNA) has one SacI and one KpnI compatible end. T4 DNA Ligase was used to ligate the 2.3 Kb fragment into SacI/KpnI digested pGEM®-7Zf(+). The reaction product was not analyzed by gel electrophoresis, but used directly for transformation of *E. coli*.

Transformation of *E. coli* JM101

E. coli strain JM101 was used for plasmid transformation experiments. LB agar plates containing ampicillin (for selection of plasmid-bearing cells) and IPTG/X-Gal (for

selection of recombinant plasmid-bearing cells) were used for transformation experiments. Plating of transformation reactions with undigested vector indicated efficiencies of greater than 10^6 transformants per microgram DNA. Ligation efficiency was 10%, determined by ligation reactions using pGEM®-7Zf(+) opened at the KpnI restriction site. Several white colonies appeared on plates of JM101 transformed with the ligation product of both insert and vector. These colonies were picked and used for isolation of plasmid DNA.

Preparation and Analysis of Plasmid DNA

Yields from plasmid minipreps were about 2 μg plasmid DNA per milliliter of transformed bacterial culture used in the procedure. Figure 6 displays an agarose gel analysis of the prepared plasmid DNA. Lane 5 shows pure DNA, in various migration forms, with no contaminating host genomic DNA. Lane 2 shows that plasmid DNA rendered linear by digestion with a single restriction endonuclease, produced a single band of 5300 bp as expected (3000 bp plasmid DNA plus 2300 bp of λgt11/grasshopper cDNA). Treatment with both SacI and KpnI produced two bands of 2300 bp and 3000 bp (lane 3, Figure 5). This indicated that the cDNA/λgt11 insert was excised from the plasmid. Also, treatment of the sample with EcoRI removed the 300 bp insert from the recombinant plasmid (not shown). A map of the predicted structure of the recombinant plasmid is shown in Figure 7. This was constructed based on the results of

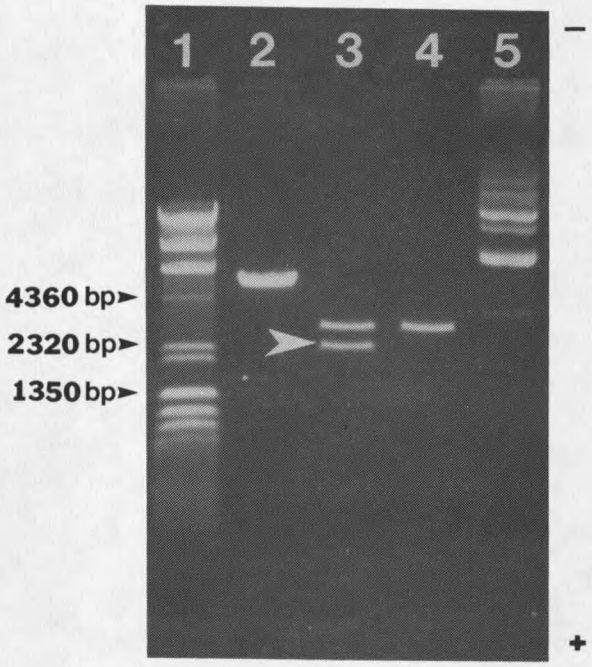

Figure 6. Agarose Gel (0.7%) Analysis of pGEM:MR2.3
-Lane 1: HindIII-digested λ, and HaeIII-digested ΘX174
-Lane 2: KpnI-digested pGEM:MR2.3 (linear, 5300 bp)
-Lane 3: KpnI/SacI-digested pGEM:MR2.3
-Lane 4: KpnI-digested pGEM®-7Zf(+) (linear, 3000 bp)
-Lane 5: undigested pGEM:MR2.3 DNA preparation
The arrow in lane 3 designates the cloned λgt11/cDNA 2300 bp KpnI/SacI fragment cut from the plasmid by digestion with the restriction endonucleases.

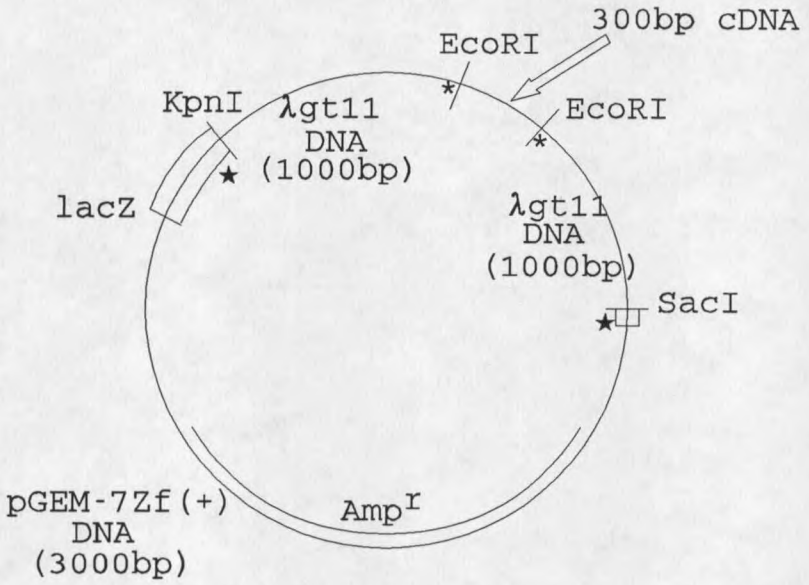

Figure 7. Plasmid Map of pGEM:MR2.3. Asterisks indicate λforward and λreverse priming sites. Stars indicate plasmid priming sites. Boxed region symbolizes lacZ coding region and cloning polylinker. The map was constructed from the results of restriction endonuclease digestion and agarose gel analysis found in Figure 6.

restriction endonuclease digestion. This recombinant plasmid was labeled pGEM:MR2.3.

Sequencing of pGEM:MR2.3

Sequencing Reactions

In the plasmid pGEM:MR2.3, λgt11 DNA flanks the cDNA insert. Primers supplied with the plasmid which bind to the promoter sites flanking the polylinker (Figure 7) were not useful as they were each separated from the inserted cDNA by about 1000 bp. Forward and reverse primers of λgt11 (λF, λR) were therefore used for sequencing reactions. These primers bind within 30 bp on either side of the λgt11 EcoRI cloning site (see asterisks, Figure 7). Four reactions were done using pGEM:MR2.3 as template: each of two sequencing enzymes (K or RT) with each of the two λgt11 primers (λF or λR).

Sequence Analysis

From the autoradiogram of the dried sequencing gel, greater than 300 bp of nucleotide sequence was read from reactions for each primer, producing complete sequence data on both strands of the cDNA insert. Data from the reverse transcriptase reaction were generally easier to read, while Klenow sequence data were used for confirmation. The determined nucleotide sequence for the 300 bp cDNA is shown in Figure 8.

One plasmid promoter primer (T7) was extended in a sequencing reaction using pGEM:MR2.3 as a template. The reaction produced about 200 bp of λgt11 sequence which was

```
1    ATGTCTTTTA GGTGTCTTCC TGTTTCGGTA ACGAAGTCAC CTTCCACAGC
51   CTCATCTGTG AATCCTGTCA GGTAGAACTC TTCTTTCGGC TCGAGGAGTC
101  GCAGCAGGGT CTCGACACGG TCCCAGGTGT CGGGCACAGC GAGCGTCGCG
151  TTCTCGGATC GGCACACCCT GGCTGCCTCC GCGTGCACGT GCTTCTCGGC
201  GTGCACCTTG TAGAAGCGGC TCGCGTCCTT CAGCCACACA TAGGAGGGTG
251  GTATGCCCAC CGAGAGCCGG CGCTCGCAGA AGAATGGCGC TGTATTCCCG
```

Figure 8. Nucleotide Sequence of 300 bp Grasshopper cDNA

identical to nucleotide sequence of the KpnI/SacI λgt11 fragment (obtained from New England Biolabs, Beverly MA). The identity confirmed the presence of λgt11 DNA and the plasmid structure as predicted for pGEM:MR2.3 (Figure 7).

Library Screening With a Radioactive Probe

Isolation and Nick Translation of a 300 bp cDNA

The cDNA contained in pGEM:MR2.3 was utilized as a hybridizing probe to screen remaining λgt11 libraries for detection of additional positive clones. The intent was to discover clones containing inserts longer than 300 bp which could be subsequently sequenced. Previous results suggested that the cDNA insert could be excised with EcoRI without any attached λgt11 DNA. This was desired, as presence of λgt11 DNA would interfere with plaque screening. Agarose gel isolation of the 300 bp cDNA from 50 μg of EcoRI restricted pGEM:MR2.3 yielded 2.5 μg of cDNA. Nick translation reactions using 0.5 μg of the 300 bp cDNA produced a nucleotide probe with specific activity of about 10^8 cpm/μg DNA and 40% incorporation of the label.

Screening of Library Plaques

Two libraries prepared from the >600 bp cDNA preparation were used to prepare plaques for screening with the DNA probe. Plaques were prepared at a density of 1×10^4 plaques per 90 mm agar plate. Approximately 1×10^5 recombinant plaques were screened with the probe. After autoradiography of hybridized plaque lifts, twenty two potential positive plaques were located. These plaques were picked, and nine of them were plaque purified.

PCR Amplification of Inserts from Positive Phage

Gel and Southern Analysis of PCR Products

The sizes of cDNA inserts from the nine positive plaques were next analyzed, to choose the longest cDNA. Because phage DNA preparation of all nine represented a lengthy task, another process was used to determine cDNA sizes. The polymerase chain reaction (PCR) was used to specifically amplify the cDNAs directly from phage suspensions. λgt11 cDNA inserts from the nine positive phage were amplified by PCR primed with the λgt11 primers which anneal to sites flanking the EcoRI cloning site. PCR products were analyzed by agarose gel electrophoresis and the results are displayed in Figure 9. Sizes of the λgt11 cDNA inserts ranged from 250 bp to 3800 bp. Southern blot analysis of the gel with the 300 bp cDNA probe was done to confirm correspondence to the 300 bp cDNA. All

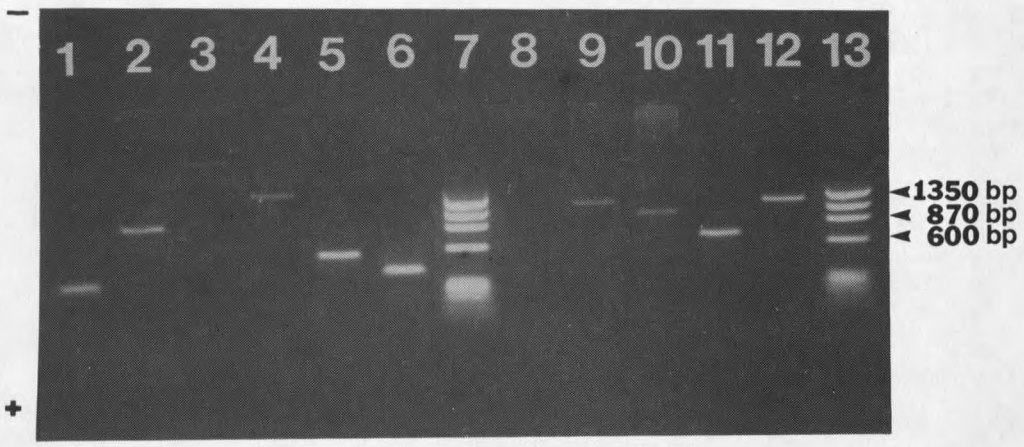

Figure 9. Gel Analysis of Amplified λgt11 cDNA Inserts.
-Lanes 1-5, 9-12: positive phage samples
-Lanes 7,13: HaeIII-digested ΘX174
-Lane 6: positive control: original recombinant λgt11 with 300 bp insert
-Lane 8: negative control: PCR reaction with no added DNA

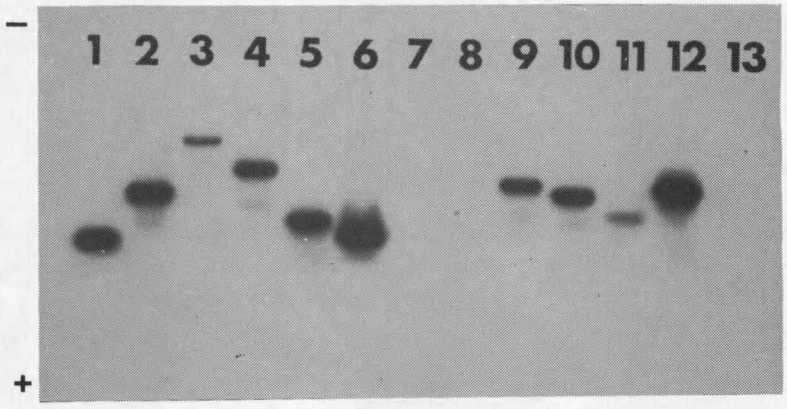

Figure 10. Southern Analysis of Figure 9 Gel.
(see Figure 9 legend for lane identification)

but one of the inserts (lane 11 represents the exception) hybridized strongly to the cDNA isolated from the λgt11 clone producing an antigenic ß-galactosidase fusion protein (Figure 10). The reason for the weak association of the amplified insert (in lane 11, Figure 10) with the probe is unknown. One of the products was actually smaller than the 300 bp probe (lane 1) at 250 bp. Seven purified clones then remained whose cDNAs were larger than 300 bp and strongly complementary to the probe.

Restriction Analysis of PCR Products

PCR products were subjected to restriction analysis with EcoRI. This was done to confirm that the inserts were a single cDNA, and not concatamers of linkered cDNAs formed in the original EcoRI cloning into λgt11. All but two of the longest cDNAs (>800 bp) exhibited some form of cleavage with EcoRI, and were not studied further. The two amplified inserts were approximately 850 bp and 950 bp.

Subcloning and Sequence Analysis of PCR Products

Preparation and Subcloning of PCR Products

Subcloning and preliminary sequence analysis of the amplified 850 bp and 950 bp cDNAs was done to determine if they contained nucleotide sequence identical to the original 300 bp cDNA. Larger volume (100 μl) PCR reactions were done to prepare a sufficient amount of amplified insert for

subcloning. These reactions yielded about 2 to 5 μg of amplified DNA. The products were purified and restricted with EcoRI to provide cohesive ends for subcloning. The EcoRI-digested samples were ligated into EcoRI-digested pGEM®-7Zf(+). Ligation efficiency was about 40%, and transformation efficiency was greater than 1 x 10^6 transformants/μg DNA. Transformation yielded several white colonies. These colonies were picked, grown in LB broth to stationary phase, and used in plasmid mini-preps to prepare plasmid DNA for nucleotide sequencing.

Preliminary Sequence of PCR Products

The primers used for sequencing reactions on plasmid DNA harboring a PCR product were complementary to the promoter region near the EcoRI site in pGEM®-7Zf(+). Nucleotide sequence was obtained from both primers on two subcloned PCR amplified cDNAs (plasmids pGEM:MR.85 and pGEM:MR.95). One end of each corresponded to the original 300 bp cDNA insert as expected. The largest insert, approximately 900 bp, was chosen for complete sequence analysis. This cDNA insert was consequently isolated from a phage miniprep and subcloned for nucleotide sequencing. The insert was isolated from phage DNA because cloned PCR products may contain unreliable sequence information. This possibility is discussed in the Discussion section.

Complete Sequencing of the cDNA Insert

Subcloning of cDNA

A DNA miniprep from the λgt11 phage harboring the approximately 900 bp cDNA produced about 25 μg DNA. An aliquot was digested with both KpnI and SacI, yielding a 3000 bp SacI/KpnI fragment. This size indicated the presence of about a 1000 bp cDNA insert. Additional phage DNA was digested similarly. The 3000 bp fragment was gel purified and ligated into SacI/KpnI-digested pGEM®-7Zf(+). The ligation reaction was used to transform $CaCl_2$-competent JM101 E. coli. A white colony was selected and amplified for plasmid preparation. The plasmid was named pGEM:MR3.0.

Sequence Analysis

Nucleotide sequence of pGEM:MR3.0 was determined using Sequenase®. The reactions were primed with the λF and λR primers and each produced 330 to 350 bases of sequence data. A portion of the autoradiogram of the sequencing gel is shown in Figure 11. One sequencing reaction produced sequence identical to the original 300 bp cDNA (Figure 8), proving that it was a longer cDNA derived from the same message. One oligonucleotide primer (20mer) was synthesized based on sequence data found furthest from that produced by the λF primer (underscore, Figure 12) and named "3'A". This primer was used in another sequencing reaction with pGEM:MR3.0. The sequence produced by this reaction overlapped with the

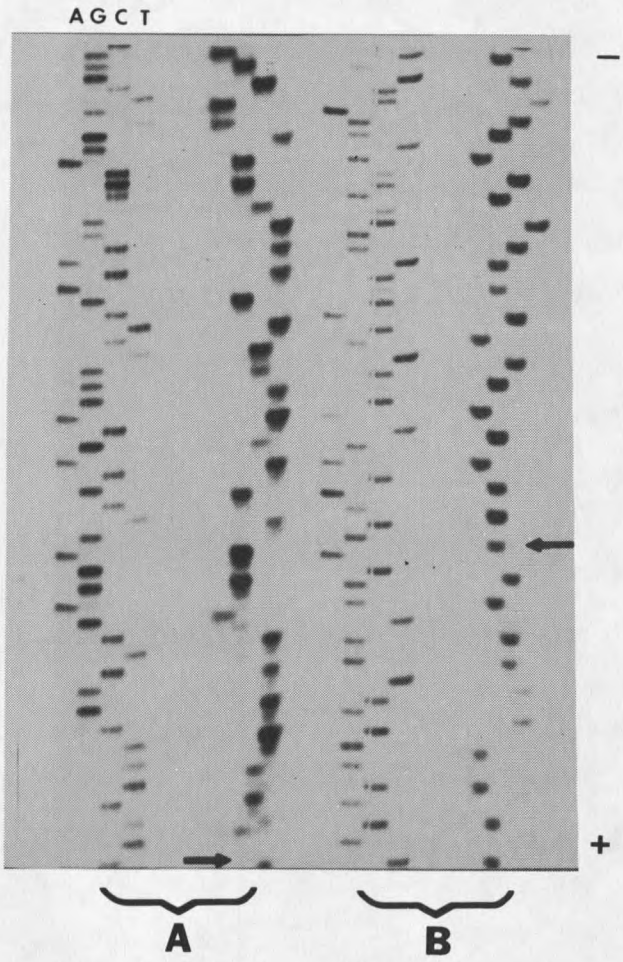

Figure 11. Portion of Sequencing Gel of pGEM:MR3.0. (A) Data complementary to the end of the 875 bp cDNA in Figure 12. The arrow designates base number 874. (B) Data corresponding to beginning of 875 bp cDNA in Figure 12. The arrow designates base number 1. The sets of four lanes are loaded in the order AGCT, left to right.

sequence produced by the two λgt11 primers. This overlap indicated that nucleotide sequence data had been obtained for the entire cDNA insert. Figure 12 displays the complete sequence of the 875 bp insert.

The nucleotide sequence found in Figure 12 was analyzed for presence of open reading frames (ORF). The results of this computer assisted analysis are shown in Figure 13. An ORF is expected to be present because the cDNA was produced from purified messenger RNA that must code for a protein product. Also, if the cDNA does correspond to GHA, a stop codon should be present at one end of the sequence, with an ORF extending to the other. A cDNA corresponding to GHA should contain an ORF of approximately 900 to 1000 bp long to account for an approximately 35,000 mw protein. An 875 bp cDNA may not be sufficiently long to account for the complete GHA coding region. Therefore, an ORF should be present throughout most of the length of the cDNA. Figure 13 indicates that a single ORF of the necessary size does not span the entire cDNA. The significance of this result will be considered later in Discussion.

The nucleotide sequence (Figure 12) was compared to cDNA sequence of Sarcophaga lectin, and to a nucleotide sequence database. The 875 bp cDNA contained 36% identity with the Sarcophaga lectin cDNA. This comparison is found in Figure 16. The identity appeared evenly throughout the sequence, without larger, continuous regions of identity. Alignment

```
1   GGGAGAGCAC GGCTGCAGCT CGCGCCAGGA GGAGGAGTTG ATAAGCAGGA
51  TGCAGCTGGT GACGGTGTGC GCGGCGCTGG TGGCAGCGAC AGTACCCTGC
101 ACCCTGGCCG CCGTGGACCT GTTCTGCAGC TGCCAGGTGC GCCACCACAG
151 GGACTCGACG ACGGCCGTGC ACTGCTCAGG GGAACAGAGT GGGAACAAAA
201 CGATTTCTTG CCAAAAAGCT CAAGTGCCGG ACATTCCACG TGACTACCAC
251 TACGTGCCAG GCTACGCCTC GTCAAGCTGT ACCGCATAAT GATGACATGG
301 GAGGAAGCCA AAAAGGCCTG CGAAGCCGAG GGAGCAAAAT TAGCAGTCCC
351 AAGAGACAAC CACGCCTACG ATGGCCTGAA GCAGATCTTC AAGTTAGGGT
401 TTGGGGTGTA CTGGGCCAAC ATCGGAATCA CAGATCATGA GAGCGAGGGA
451 TATTCAGCGG AGTGGATGGT CATCCAGTGT CGTTCCTGCC ATGGAATCCT
501 AATGAACCCA ACAACCCGGA GGCAACGAGA ACTGTGTTAA CGTCAACGAC
551 AAAGGACAGC TGAACGACTG GCATTGCGGG ATACAGCGCC ATTCTTCTGC
601 GAGCGCCGGC CCTCGGTGGG CATACCACCC TCCTATGTGT GGCTGAAGGA
651 CGCGAGCCGC TTCTACAAGG TGCACGCCGA GAAGCACGTG TACGCGGAGG
701 CAGCCAGGGT GTGCCGATCC GAGAACGCGA CGCTCGCTGT GCCCGACACC
751 TGGGACCGTG TCGAGACCCT GCTGCGACTC CTCGAGCCGA AAGAAGAGTT
801 CTACCTGACA GGATTCACAG ATGAGGCTGT GGAAGGTGAC TTCGTTACCG
851 AAACAGGAAG ACACCTAAAA GGCAT
```

Figure 12. Complete Nucleotide Sequence of 875 bp cDNA. The twenty nucleotides underscored represent the annealing site of the synthesized 20 base oligonucleotide primer. The structure represents determined nucleotide sequence found immediately inside the EcoRI sites. Residues 575 through 875 represent the original 300 bp sequence in Figure 8.

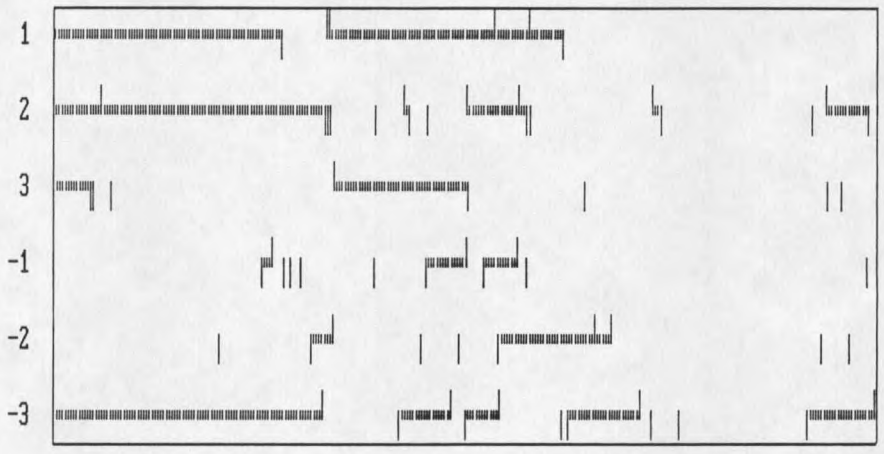

Figure 13. ORF Analysis of 875 bp cDNA. Vertical lines extending up denote methionines and ones extending downward denote stop codons. Bars indicate open reading frames. Numbers on the left side of the diagram designate the six possible frames in the cDNA. The positive frame numbers correspond to the sequence shown in Figure 12.

against the GenBank database produced little identity with other invertebrate nucleotide sequences, and no extensive similarities with non-invertebrate DNA sequences.

Three of the possible translations of the 875 bp cDNA are displayed in Figure 14, and represent the three possible frames of the sequence shown in Figure 12. These sequences, with the stops removed, were aligned with sequences in the NBRF database and the five invertebrate lectins (Figure 1). The results of these alignments suggested lack of any strong homologies. Little identity was observed with the other lectins, especially in regards to the invariant residues. These results are tempered with the fact that none of the translations in Figure 14 represent a complete ORF. Confirming a proper ORF in the sequence should contribute to alignment reliability. The significance of the lack of strong similarity with other sequences is discussed later.

Lastly, the six possible translations were compared to the two tryptic peptide sequences found in Figure 1. One peptide sequence, VNDEY_I_E, did not appear in the translations. Identity with the other, GVGVPDTYN_V, was not observed, but a tetrapeptide match was found and is underscored in Figure 14. Reliability of this comparison also depends on presence of a proper reading frame. Therefore, the peptide sequences may not be found in the cDNA sequence until an ORF has been established.

```
1    GRARLQLAPG GGVDKQDAAG DGVRGAGGSD STLHPGRRGP VLQLPGAPPQ
51   GLDDGRALLR GTEWEQNDFL PKSSSAGHST *LPLRARLRL VKLYRIMMTW
101  EEAKKACEAE GAKLAVPRDN HAYDGLKQIF KLGFGVYWAN IGITDHESEG
151  YSAEWMVIQC RSCHGILMNP TTRRQRELC* RQRQRTAERL ALRDTAPFFC
201  ERRPSVGIPP SYVWLKDASR FYKVHAEKHV YAEAARVCRS ENATLAVPDT
251  WDRVETLLRL LEPKEEFYLT GFTDEAVEGD FVTETGRHLK G
```

```
1    GEHGCSSRQE EELISRMQLV TVCAALVAAT VPCTLAAVDL FCSCQVRHHR
51   DSTTAVHCSG EQSGNKTISC QKAQVPDIPR DYHYVPGYAS SSCTA***HG
101  RKPKRPAKPR EQN*QSQETT TPTMA*SRSS S*GLGCTGPT SESQIMRARD
151  IQRSGWSSSV VPAMES**TQ QPGGNENCVN VNDKGQLNDW HCGIQRHSSA
201  SAGPRWAYHP PMCG*RTRAA STRCTPRSTC TRRQPGCADP RTRRSLCPTP
251  GTVSRPCCDS SSRKKSST*Q DSQMRLWKVT SLPKQEDT*K A
```

```
1    ESTAAARARR RS**AGCSW* RCARRWWQRQ YPAPWPPWTC SAAARCATTG
51   TRRRPCTAQG NRVGTKRFLA KKLKCRTFHV TTTTCQATPR QAVPHNDDMG
101  GSQKGLRSRG SKISSPKRQP RLRWPEADLQ VRVWGVLGQH RNHRS*ERGI
151  FSGVDGHPVS FLPWNPNEPN NPEATRTVLT STTKDS*TTG IAGYSAILLR
201  APALGGHTTL LCVAEGREPL LQGARREARV RGGSQGVPIR ERDARCARHL
251  GPCRDPAATP RAERRVLPDR IHR*GCGR*L RYRNRKTPKR H
```

Figure 14. Translations of Nucleotide Sequence of the 875 bp cDNA (from Figure 12). The three possible reading frames are represented. These correspond to 1, 2 and 3 in the ORF analysis in Figure 13, respectively. Asterisks denote stop codons in the nucleotide sequence. Underscore in the top translation indicates a tetrapeptide match with one GHA tryptic peptide, displayed in Figure 1.

DISCUSSION

Perspective

The results presented in the previous section indicate that the complete nucleotide sequence of an 875 bp grasshopper cDNA has been successfully determined. A summary of the events leading up to characterization of this cDNA follow.

A Melanoplus differentialis fat body cDNA library was constructed in the E. coli bacteriophage λgt11. Plaques of the library phage were screened with anti-GHA antibodies. These antibodies identified one plaque (clone) as containing an antigen, suggesting presence of a cloned cDNA corresponding to GHA. This phage clone was subsequently purified, and DNA analysis indicated presence of a 300 bp insert in the EcoRI cloning site. The cDNA from this clone was purified, and rendered radioactive by nick translation with [α-^{32}P]dCTP.

Repeated library plaque screening with the specific 300 bp cDNA probe permitted isolation of additional positive phage clones. Selection of the longest identified insert provided the 875 bp cDNA. This cDNA was subcloned into a plasmid vector, and sequence data for the entire insert was obtained. Because of its length, complete nucleotide sequence determination required synthesis of an oligonucleotide primer

complementary to sequence present in the internal portion of the cDNA. This primer was extended in a sequencing reaction to produce sequence data for the internal portion of the insert. Finally, the determined sequence of the 875 bp insert was analyzed.

Completion of this project required use of numerous procedures in the field of molecular biology. Many of these were standard protocols found in laboratory handbooks, but some experiments required modification of existing procedures for completion. Explanation of the key experiments and necessary modifications are presented in the following sections.

Central Strategy

Choice of Library Vector

To complete the objective of isolation of a GHA cDNA, a suitable probe had to be available to locate this cDNA. A DNA probe is usually desired because cDNAs in recombinant vectors can be detected regardless of insertion frame or direction.

Construction of DNA probes is often based on partial amino acid sequence of the protein of interest. It was decided that DNA probes based on structure of the two GHA tryptic peptides (Figure 1) would be highly degenerate. A probe of this nature may not be sufficiently specific, and non-specific binding could interfere with location of true

positives. Therefore an alternate method of detection of cDNAs was desirable.

The anti-GHA antibody offered a useful probe. Recombinant vectors able to express inserted cDNAs can be screened with antibodies. This led to choice of the expression vector λgt11.

Recombinant λgt11 vectors contain cDNA insertions at the EcoRI site at the end of the λlac5 gene. Under conditions of induction with IPTG, the gene/cDNA is transcribed. Translation of the message results in a ß-galactosidase fusion protein. If the cDNA is inserted in the appropriate direction and frame, the fusion protein will contain a polypeptide coded for by the cDNA. Antibodies able to recognize the denatured and reduced form of the desired protein may be able to associate with this fusion protein. These recombinant clones can therefore be detected using such a probe.

The rabbit anti-GHA antibody preparation identified one λgt11 plaque as producing an antigenic fusion protein. This clone was chosen for further characterization, based on the antigen's presumed correlation to a grasshopper agglutinin.

Subcloning Strategy

Two things needed to be done after identification of the insert: sequence analysis followed by isolation of the insert. Determination of the cDNA nucleotide sequence would provide confirmatory data for future sequenced cDNAs. Isolation of the cDNA in pure form was done next so that it

could be converted into a probe to search the library for a full-length cDNA.

Sequence analysis of cDNA inserts while cloned in λgt11 is possible under carefully controlled conditions (Kim and Jue, 1990), but is not consistently successful. Preparation of large amounts of λgt11 DNA for sequencing and insert isolation was also difficult. For these reasons, it was concluded that removal of the insert and subcloning into a plasmid vector would aid analysis of cDNAs from λgt11. The plasmid pGEM®-7Zf(+) was chosen for subcloning because of its polylinker restriction site compatibility with the ends of the cDNA.

Use of the cDNA Probe

The 300 bp cDNA was isolated from the recombinant plasmid. This cDNA then served as a probe for extended screening of the library.

In conducting the plaque screening experiments, the goal was to locate longer cDNAs containing a portion identical to the 300 bp probe. Plaque screening using the probe was carried out at high stringency (i.e. washing of hybridized filters at 68°C) because a 300 bp DNA strand will associate strongly with a complement. Increased stringency results in dissociation of weakly bound (background) probes, thus targeting the true positives. Screening of plaques from the

library in this manner clearly distinguished positive plaques. The 875 bp cDNA was identified by means of this screen.

Size analysis of numerous positive clones detected by screening with the DNA probe indicated the longest insert to be 875 bp. Longer inserts were identified, but cleavage by EcoRI indicated probable concatamerization of cDNAs before initial cloning into λgt11 DNA. Further analysis of these inserts was therefore unsuitable.

Choice of Sequencing Protocol

The Sanger dideoxy-chain termination method was chosen for sequencing of recombinant plasmids because of its common usage and the availability of protocols. Two methods are generally used for rendering dideoxy-terminated DNA strands radioactive: either end-labeling of primers with ^{32}P or incorporation of radioactive residues (usually [α-^{32}P]dNTP or [α-^{35}S]dNTP) into the newly synthesized strand. Incorporation of [α-^{35}S]dATP into DNA was chosen for two reasons. First, autoradiographic bands produced by sequencing gels using ^{35}S are sharper and easier to read. Secondly, the isotope is safer and easier to use than the ^{32}P isotopes because of its lower emission energy.

After selection of the sequencing protocol, nucleotide sequence was obtained for approximately two-thirds of the 875 bp insert. The center one-third of the cDNA remained to be sequenced. Two options for obtaining that portion of the

sequence were considered. (1) The nucleotide sequence would be read for possible restriction sites using computer software. Suitable sites would be chosen for fragmentation of the insert, then fragments would be subcloned. (2) An oligonucleotide primer would be constructed based on sequence near the unknown region. This primer would be used to sequence the internal region directly, in a 'walk through' fashion.

Method (2) was chosen for three reasons. First, the sequencing experiment could be done much faster. Secondly, facilities for construction of the primer are readily available in an adjacent laboratory. Lastly, primer construction would provide valuable experience with DNA synthesis techniques and the newest automated technology.

These key experiments, and a comprehension of the procedures involved, permitted the determination of complete nucleotide sequence of the 875 bp cDNA.

Unique Tools Used in Project

Opportunities were presented throughout the project that allowed use of special procedures. These procedures generally enhanced the project by producing cleaner data, or decreasing the time needed to complete the project.

Polymerase Chain Reaction

The polymerase chain reaction amplifies specific regions of double stranded DNA through use of oligonucleotide primers.

The procedure requires two primers, one specific for each strand, that anneal to sites flanking the region of interest.

The numerous positive clones resulting from plaque screening with the cDNA probe required analysis. This presented a lengthy task if standard phage DNA preps were used. PCR allowed rapid size analysis of the cDNA inserts using the λgt11 primers which flank the inserts. The PCR method was appealing because of three characteristics. (1) Phage suspensions were used directly, without purification of DNA. (2) Analysis of the inserts could be completed in one day. (3) The PCR products could be subcloned directly for preliminary sequence analysis.

Sequence data produced from cloned PCR products can only be preliminary because the clones represent a single amplified DNA molecule. The Taq DNA polymerase used in the reaction lacks a 3'→5' editing function and nucleotide misincorporation occurs at a rate of 2×10^{-4} nucleotides in each round of amplification (Saiki et al., 1988). Cloned PCR products therefore can contain unreliable sequence information because they represent an individual amplified molecule. Direct sequencing of an entire amplification reaction, although, can produce reliable data. When the complete product of the PCR is used for a DNA sequencing template, the random misincorporated nucleotides become statistically insignificant. Two experiments based on protocol in Sambrook et al. (1990) of direct sequencing were unsuccessful. The

sequence analysis therefore had to be done on cDNA's isolated from phage DNA preparations to be reliable. Recently however, direct sequencing of PCR products has been successful in a neighboring laboratory, using a protocol developed by Higuchi et al. (1989). Direct sequencing by this procedure may therefore be used to accelerate future analysis of cDNAs.

Use of Sequenase®

The latest version (2.0) of USB's modified T7 DNA polymerase produced higher quality data than that produced by other enzymes used in the project. Procedures for sequencing with this system could be completed in less than one hour, representing half the time of Promega's K/RT® system. Sequencing reactions using both systems indicated that Sequenase® produced significantly more reliable and consistent data. Figure 15 displays a comparison of Sequenase® and reverse transcriptase enzymes. These observations suggest that this enzyme should be used in future projects in this laboratory.

Advanced Computing

Use of computer software and other computing services contributed greatly to completion of the project. Both networked and PC software were used throughout. Personal computer systems were used to run software such as GenePro. The software packages provided rapid investigation of

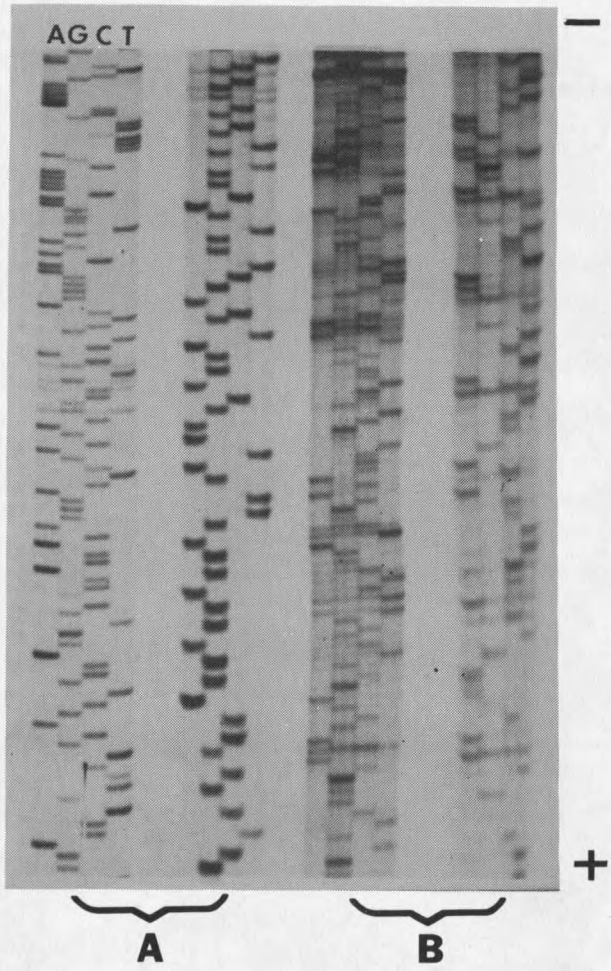

Figure 15. Comparison of Sequenase® (A) and K/RT® (B) Sequencing Enzymes. These results indicate that Sequenase® produces higher quality data than the Promega (K/RT) system. The sets of four lanes are loaded AGCT from left to right.

nucleotide and protein sequence data, including restriction site, open reading frame, and homology analyses.

A proposal to the Pittsburgh Supercomputing Facility was prepared, securing a starting account at the facility. The account provided sufficient use privileges to access all software with biomedical applications. Networking capabilities on the campus allowed connection to the facility and ability to transfer files. Computing at this facility made use of extensive sequence databases and the extensive GCG software package.

Procedural Difficulties and Solutions

Standard protocols found in 'handbooks' occasionally were inadequate for completion of experiments. Solutions to these problems were found in other literature or by creative modification of procedures.

Immunological Screening

Experiments to confirm antigenicity of the phage clone isolated by Dr. K.D. Hapner at Purdue University gave uncertain results. The anti-GHA primary antibody bound strongly to all plaques, suggesting cross-reactivity with *E. coli* lysate. Adjustment of experimental conditions, such as incubation times and reagent concentrations, did not improve the background binding problem. The cross-reactivity of the antibodies was confirmed by Western blotting (Figure 3). A method to remove cross-reactive antibody species was found in

Sambrook, et al. (1990). The problem was solved by construction of an affinity column containing immobilized E. coli lysate protein. The cross-reactive antibodies were removed by passing the anti-GHA solution over the column. Use of this purified antibody confirmed the presence of an antigenic λgt11 clone (Figure 4).

Phage DNA Preparation

Preparation of λgt11 DNA from lysates was found to be a difficult procedure. Liquid lysate preparation was generally unsuccessful, requiring use of the described plate lysate protocol. It was determined that the key step in the process was the growth of phage plaques. Plate lysates often gave weak plaques resulting in very low yields of poor quality DNA. Some factors found to be important are: fresh media, LB bottom agar in lieu of agarose, phage/host ratio upon infection and incubation time. Choice conditions have been outlined, but further optimization is necessary.

Manipulation of Recombinant Phage DNA

Analysis of recombinant phage DNA followed its purification. Restriction analysis of the original 300 bp cDNA/λgt11 clone was difficult. Treatment with EcoRI restriction endonuclease was expected to remove the cDNA since it was cloned in the vector at the EcoRI site. This was unsuccessful, and cleavage at alternate sites was necessary to observe presence of the insert (Figure 5). It was assumed

that the EcoRI sites were altered upon cloning and were unavailable.

A plasmid vector was chosen that would allow ligation of an insert with SacI and KpnI compatible ends. The insert was then separated from plasmid priming sites by 1000 bp. Because of this, primers specific for λgt11 DNA were used in place of the plasmid primers. These modifications resulted in the ability to successfully determine nucleotide sequence of the cDNA inserts.

An unexpected observation occurred during plasmid analysis. Treatment with EcoRI removed the 300 bp cDNA from the plasmid, whereas it could not remove it from the λgt11 DNA. This indicated that resistance of the recombinant λgt11 was probably due to impurities in the phage DNA sample. The capability to remove the cDNA from the plasmid proved to be useful for isolation and preparation of the probe.

Data Analyses of the 875 bp cDNA

Three central analyses were completed based on the determined nucleotide sequence of the 875 bp cDNA. The purpose of these analyses was to confirm identity of the cDNA. These analyses were: (1) open reading frame, (2) nucleotide and deduced peptide sequence homology with other sequences and (3) deduced amino acid sequence homology with two tryptic peptides isolated from GHA. Each of these analyses will be presented in turn, and summarized at the end of the section.

Open Reading Frame

The analysis in Figure 13 indicates that a large ORF is not present, at least one that could correspond to a 35,000 mw protein. An ORF in the cDNA is expected, as stated in the previous chapter. It is therefore possible that some mistakes were produced in the sequencing reactions or in reading the autoradiograms, and caused the ORF analysis to be unclear. To date, nucleotide sequence of the equivalent of only one strand of the cDNA has been determined. Sequencing of both strands of the cDNA will provide reliable nucleotide sequence and confirm previously obtained sequence. Sequence analysis of both strands may correct any analysis and produce a clear ORF. Sequence data should therefore be obtained for both strands of the entire 875 bp cDNA so that the ORF analysis is dependable.

Similarity With Other DNA and Protein Sequences

The 875 bp sequence was further examined by searching for similarities with other nucleotide and amino acid sequences. Several regions of semiconserved sequence exist among the invertebrate lectins (Figure 1). Because of this, it was assumed that a translation of the cDNA sequence would exhibit similarity with the other invertebrate lectin sequences. The three translated frames (Figure 14) of the cDNA were aligned with these lectin sequences. Significant identity of the translations with the invertebrate lectin amino acid sequences was not observed. Most notably, portions of the cDNA

translations were not found that corresponded to the semiconserved regions in the invertebrate lectins. It is important to note that identity is low between the five lectin structures in Figure 1 (less than 30%). Strong homology may therefore not be seen with the translated 875 bp cDNA. The most important correlation might be identity with the ten residues conserved among the five lectins.

Additionally, the three translations were aligned against the NBRF protein sequence database. No significant regions of amino acid homology were observed upon this analysis, and few of the identified homologies were with invertebrate proteins. It is assumed that lack of homology with sequences is related to lack of an obvious ORF. This suggests that a reliable ORF should be established before pursuing protein sequence alignments.

The Sarcophaga lectin structure has been solved via molecular cloning methods (Takahashi et al., 1985). A cDNA nucleotide sequence of the Sarcophaga lectin was therefore available. The 875 bp grasshopper cDNA sequence was compared to the Sarcophaga lectin cDNA sequence. It was thought that the two sequences may be similar since they both correspond to an insect lectin. Similarity between the two would suggest that the 875 bp cDNA may correspond to GHA. An identity of 36% was observed between the Sarcophaga lectin cDNA and the 875 bp grasshopper cDNA (Figure 16). A greater than one-third identity between two nucleotide sequences may be significant.

significant. Further manipulation of alignment parameters should be done to confirm the highest possible identity. The similarity also suggests that reliable sequence data should be obtained to produce the best alignment.

Lastly, the translations were searched for identity or homology with the two tryptic peptides isolated from GHA. Precise matches with the peptides were not found, but a tetrapeptide match was observed in frame 1 (Figure 14) with one of the tryptic peptides. The lack of identity with the peptides may be due to the lack of an obvious ORF. Because the peptides were isolated from a protein, their corresponding nucleotide sequences are expected to be found only in an ORF. The tetrapeptide match (VPDT) is promising, and will be studied further upon establishment of an ORF.

Summary of Analyses

The identification of the 875 bp cDNA, as coding for GHA, has not been determined based on initial data from open reading frame and homology analyses. The cDNA presumably corresponds to grasshopper agglutinin, but this relationship is not confirmed by analysis of the nucleotide sequence of the insert.

The uncertainty about identity of the 875 bp cDNA results from three observations. (1) Homology with semiconserved regions of invertebrate lectins was expected, but not observed. (2) An open reading frame in the sequence relating to a protein of 35,000 mw (about 900 bp) is apparently not

present. More rigorous sequence analysis (i.e. complete analysis of both strands) should be done before the ORF analysis is considered reliable. (3) Identity of a portion of a cDNA translation with the tryptic peptides was expected. Although identity was not observed, the tetrapeptide match may suggest a correlation. These results suggest that further characterization of the cDNA is necessary to confirm its identity. Therefore, the 875 bp cDNA must have nucleotide sequence determined for both strands. Also, a longer insert should be isolated so that a complete ORF may be identified. Defining of a complete ORF will allow deduction of protein sequence.

Future study will be made easier through use of the procedures presented in this study. Even if a new cDNA library needs to be made for further immunological screening and plaque isolation, the modified procedures will greatly quicken progress. These procedures include: the improved antibody purification, bacteriophage and phage DNA manipulation improvements, and the use of the polymerase chain reaction, both for rapid analysis and the possibility of direct sequence analysis of amplified material.

Future Work

The ambiguity of the cDNA sequence needs to be resolved. Several immediate possibilities exist for future work to achieve this goal. First, remaining positive clones isolated

from nucleotide screening should be characterized. An insert longer than 875 bp may be identified as a result of the screening. A longer insert would provide more information for characterization of the lectin protein. Second, the 875 bp insert, when expressed in the same frame as the original 300 bp insert, should be antigenic and react more strongly with the anti-GHA antibody. This would confirm its correlation to a GHA protein.

If these experiments do not confirm the identity of the cDNA, screening of remaining λgt11 libraries with the purified antibody may be done. The purpose of the screening would be to search for a new and different λgt11 clone that is antigenic for the GHA antibody. This new clone may contain the proper cDNA which corresponds to GHA A search for a new antigenic clone may also, in the extreme, require construction of a new fat body cDNA library.

CONCLUSIONS

Thesis research carried out and presented here represents the successful completion of the major goal: determination of the nucleotide sequence of an 875 bp cDNA cloned from a grasshopper fat body λgt11 library. Major milestones achieved during the work include:

(1) Isolation of a λgt11 positive clone done by means of immunoscreening with grasshopper agglutinin polyclonal antibody.

(2) Excision, sequencing and nick translation of the initial 300 bp cDNA insert.

(3) Polymerase chain reaction amplification of cDNA inserts of several λgt11 positive clones.

(4) Detection by probe-hybridization, and nucleotide sequencing of an 875 bp cDNA insert.

(5) Computer based open reading frame analysis and database alignment of the 875 bp cDNA.

Some additional confirmatory sequence analysis and acquisition of a larger cDNA will be necessary to verify that the isolated cDNA clone corresponds to the grasshopper agglutinin molecule.

REFERENCES CITED

REFERENCES CITED

Aviv, H., P. Leder (1972) Purification of biologically active globin messenger RNA by chromatography on oligo-thymidylic acid-cellulose. Proc. Natl. Acad. Sci. **69**:1408

Bayne, C.J. (1990) Phagocytosis and non-self recognition in invertebrates. BioScience **40**:723

Benton, W.D., R.W. Davis (1977) Screening λgt11 recombinant clones by hybridization to single plaques in situ. Science **196**:180

Birnboim, H.C., J. Doly (1979) A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucleic Acids Res. **7**:1513

Bradley, R.S., G.S. Stuart, B. Stiles, K.D. Hapner (1989) Grasshopper haemagglutinin: Immunochemical localization in haemocytes and investigation of opsonic properties. J. Insect Physiol. **35**:353

Chirgwin, J.M., A.E. Przybyla, R.J. Macdonald, W.J. Rutter (1979) Isolation of biologically active ribonucleic acid from sources enriched in ribonuclease. Biochem. **18**:5294

Cohen, S.N., A.C.Y Chang, L. Hsu (1972) Nonchromosomal antibiotic resistance in bacteria: Genetic transformation of Escherichia coli by R-factor DNA. Proc. Natl. Acad. Sci. **69**:2110

de Wet, J.R., H. Fukushima, N.N. Dewji, E. Wilcox, J.S. O'Brien, D.R. Helinski (1984) Chromogenic immunodetection of human serum albumin and α-L-fucosidase clones in a human hepatoma cDNA expression library. DNA **3**:437

Dunn, P.E. (1986) Biochemical aspects of insect immunology. Ann. Rev. Entomol. **31**:321

Dunn, P.E. (1990) Humoral immunity in insects. BioScience **40**:738

Edmonds, M., M.H. Vaughan, Jr., H. Nakazato (1971) Polyadenylic acid sequences in the heterogenous nuclear RNA and rapidly-labeled polyribosomal RNA of HeLa cells: Possible evidence for a precursor relationship. Proc. Natl. Acad. Sci. **68**:1336

Giga, Y., A. Ikai, K. Takahashi (1987) The complete amino acid sequence of Echinoidin, a lectin from the coelomic fluid of the sea urchin Anthocidaris crassispina. J. Biol. Chem. **262**:6197

Giga, Y., K. Sutoh, A. Ikai (1985) A new multimeric hemagglutinin from the coelomic fluid of the sea urchin Anthocidaris crassispina. Biochemistry **24**:4461

Gubler, V., B.J. Hoffman (1983) A simple and very efficient method for generating cDNA libraries. Gene **25**:263

Hanahan, D. (1983) Studies on transformation of Escherichia coli with plasmids. J. Mol. Biol. **166**:557

Hapner, K.D. (1983) Haemagglutinin activity in the haemolymph of individual Acrididae (grasshopper) specimens. J. Insect Physiol. **29**:101

Hapner, K.D., M.R. Stebbins (1986) Biochemistry of arthropod agglutinins. In A.D. Gupta (ed.) Hemocytic and Humoral Immunity in Arthropods, John Wiley & Sons, New York

Higuchi, R.G., H. Ochman (1989) Production of single stranded DNA templates by exonuclease digestion following the polymerase chain reaction. Nucleic Acids Res. **17**:5865

Ish-Horowicz, D., J.F. Burke (1986) Rapid and efficient cosmid cloning. Nucleic Acids Res. **9**:2989

Jurenka, R., K. Manfredi, K.D. Hapner (1982) Haemagglutinin activity in Acrididae (grasshopper) haemolymph. J. Insect Physiol. **28**:177

Karp, R.D. (1990) Cell-mediated immunity in invertebrates. BioScience **40**:732

Karp, R.D. (1990) Inducible humoral immunity in insects: does an antibody-like response exist in invertebrates? Res. in Immunology **141**:932

Kim, B.S., C. Jue (1990) Direct sequencing of lambda-gt11 recombinant clones. Biotechniques **8**:156

Komano, H., D. Mizuno, S. Natori (1980) Purification of lectin induced in the haemolymph of Sarcophaga peregrina larvae on injury. J. Biol. Chem. **255**:2919

Lackie, A.M. (1981) Immune recognition in insects. Dev. Comp. Immunol. **5**:191

Lackie, A.M. (1988) Haemocyte behaviour. Adv. Insect Phys. **21**:85

Laemmli, U.K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4 Nature **227**:680

Lis, H., N. Sharon (1986) Lectins as molecules and as tools. Ann. Rev. Biochem. **55**:35

Mierendorf, R.C., C. Percy, R.A Young (1987) Gene isolation by screening λgt11 libraries with antibodies. Meth. Enzymol. **152**:458

Muramoto, K., H. Kamiya (1986) The amino-acid sequence of a lectin of the acorn barnacle Megabalanus rosa. Biochem. Biophys. Acta **874**:285

Muramoto, K., H. Kamiya (1990) The amino-acid sequence of multiple lectins of the acorn barnacle Megabalanus rosa and its homology with animal lectins. Bioc. Biophys. Acta **1039**:42

Natori, S.C. (1990) Dual functions of insect immunity proteins in defence and development. Res. in Immunology **141**:938

Ogden, R.C., D.A. Adams (1987) Electrophoresis in agarose and acrylamide gels. Meth. Enzymol. **152**:61

Pearson, W.R., D.J. Lipman (1988) improved tools for biological sequence comparison. Proc. Natl. Acad. Sci. **85**:2444

Ratcliffe, N.A., A.F. Rowley, S.W. Fitzgerald, C.P. Rhodes (1985) Invertebrate immunity: Basic concepts and recent advances. International Review of Cytology **97**:183

Saiki, R.K., D.H. Gelfand, S. Stoffel, S.J. Scharf, R. Higuchi, G.T. Horn, K.B. Mullis, H.A. Erlich (1988) Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science **239**:487

Sambrook, J., Fritsch, E.F., T. Maniatis (1989) Molecular cloning: A laboratory manual 2nd ed. Cold Spring Harbor Laboratory, New York

Sanger, F., S. Nicklen, A.R. Coulson (1977) DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. **74**:5463

Sharon, N., H. Lis (1989) Lectins as cell recognition molecules. Science **246**:227

Southern, E.M. (1975) Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. **98**:503

Stebbins, M.R., K.D. Hapner (1985) Preparation and properties of haemagglutinin from haemolymph of Acrididae (grasshoppers). Insect Biochem. **15**:451

Stiles, B., R.S. Bradley, G.S. Stuart, K.D. Hapner (1988) Site of synthesis of the haemolymph agglutinin of Melanoplus differentialis (Acrididae:Orthoptera). J. Insect Physiol. **34**:1077

Suzuki, T., T. Takagi, T. Furukohri, K. Kawamura, M. Nakauchi (1990) A calcium-dependent galactose-binding lectin from the tunicate Polyandrocarpa misakiensis. J. Biol. Chem. **265**:1274

Takahashi, H., H. Komano, N. Kawaguchi, N. Kitamura, S. Nakanishi, S, Natori (1985) Cloning and sequencing of cDNA of Sarcophaga peregrina humoral lectin induced on injury of the body wall. J. Biol. Chem. **260**:12228

Takahashi, H., H. Komano, S. Natori (1986) Expression of the lectin gene in Sarcophaga peregrina during normal development and under conditions where the defence mechanism is activated. J. Insect Physiol. **32**:771

Towbin, H., T. Staehelin, J. Gordon (1979) Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: Procedures and some applications. Proc. Natl. Acad. Sci. **76**:4350

Young, R.A., R.W. Davis (1983) Efficient isolation of genes by using antibody probes. Proc. Natl. Acad. Sci. **80**:1194

THE APPENDICES

APPENDIX I

Recipes for Buffers and Microbiological Media

APPENDIX I

LB medium
10g Bacto-tryptone
5g Bacto-yeast
5g NaCl
To 1 liter, pH 7.5

LB Agar Plates
LB medium with
1.5% Bacto-agar

Top Agarose
LB medium with
0.6% agarose

M-9 minimal media
8.0mM NaCl
42mM Na_2HPO_4
22mM KH_2PO_4
19mM NH_4Cl
1.5% agarose
2.0mM $MgSO_4$
0.2% glucose
0.1mM $CaCl_2$
1.0mM thiamine-HCl

SOC medium
20g Bacto-tryptone
0.5g Bacto-yeast
0.5g NaCl
to one liter
2.5mM KCl
10mM $MgCl_2$
20mM glucose
pH 7.0

TBS
10mM Tris-HCl
0.15M NaCl
pH 8.0

TBST
TBS with
0.05% Tween-20

TE
10mM Tris-HCl (pH 8.0)
1mM EDTA

TBE
89mM Tris
89mM boric acid
2.5mM EDTA
pH 8.3

SM
0.1M NaCl
10mM $MgSO_4$
20mM Tris-HCl
0.1% gelatin
pH 7.4

Coupling buffer
67mM sodium borate
0.67M NaCl
pH 8.0

Solution I
50mM glucose
25mM Tris-HCl (pH 8.0)
10mM EDTA (pH 8.0)

Solution II
0.2N NaOH
1% SDS

Solution III
3M potassium acetate
11.5% acetic acid

SSPE
0.15M NaCl
1.0mM EDTA
10mM $NaH_2PO_4 \cdot H_2O$
pH 7.4

Gel sample buffer
15% Ficoll
0.2M EDTA
6X TBE
0.25% xylene cyanol

Hybridization buffer
50% formamide
5X Denhardt's
100μg/ml sheared, denatured salmon sperm DNA
5X SSPE
0.5% SDS

Denhardt's reagent
0.02% Ficoll
0.02% polyvinylpyyrolidone
0.02% BSA (Sigma-frac.V)

Alkaline phosphatase buffer
0.1M Tris-HCl
0.1M NaCl
5mM $MgCl_2$
pH 9.5

APPENDIX II

Figure 16. Comparison of Sarcophaga Lectin cDNA with Grasshopper 875 bp cDNA

APPENDIX II

Gap Weight:	5.000	Average Match:	1.000
Length Weight:	0.300	Average Mismatch:	0.000
Quality:	277.9	Length:	1003
Ratio:	0.313	Gaps:	6
Percent Similarity:	35.706	Percent Identity:	35.706

```
 51  CAATGAAGAACGTAGAAGGCTTCGTTATATTTTTAGTAATTTTTACGTCT  100
                                                 |   |
  1                                      GGGAGAGCACGGCTGCAG   18

101  ACAGCGGCAGTGCCCCAATTACAAAAGGCTTTAGATGGCAGAGAATATCT  150
       ||| |||          | |   | |||       |   |    | ||    | |
 19  CTCGCGCCAGGAGGAGGAGTTGATAAGCAGGATGCAGCTGGTGACGGTGT   68

151  TATTGAAACAGAACTTAAGTACAATTGGCATCAAGCCTGGCATGAATGTG  200
       |   | |       |   ||| |   |  ||   |  | |       |
 69  GCGCGGCGCTGGTGGCAGCGACAGTACCCTGCACCCTGGCCGCCGTGGAC  118

201  CCCGCCATGATCAACAACTTGTAACAATCGAAAGTGCTGATAAGAATAAT  250
     |    |   | | | | |  ||   ||     |  |     ||  |      |
119  CTGTTCTGCAGCTGCCAGGTGCGCCACCACAGGGACTCGACGACGGCCGT  168

251  GCCATTATCGATCTGGTGAAACGGGTCGTTGGAAAATCTCATAATTTATG  300
     ||    |  | || || |||  | || |    ||||       | ||  ||
169  GC....ACTGCTCAGGGGAACAGAGTGGGAACAAAA....CGATTTCTTG  210

301  GTTGGGCGGCAATGATGAGTATAGTTCAAGTCGTGACTATGGCAGACCCT  350
           |   | |     | | | | || ||     || |   |    ||
211  CCAAAAAGCTCAAGTGCCGGACATTCCACGTGACTACCACTACGTGCCAG  260

351  TTTTCTGGTCACCCACCGGTCAAGCATTCTCCTTTGCCTACTGGTCGGAA  400
       | |   ||  | | | ||   ||||  |  |          ||  |  |
261  GCTACGCCTCGTCAAGCTGTACCGCATAATGATGACATGGGAGGAAGCCA  310

401  AACAATCCCGATAATTATAAGCATCAAGAACATTGTGTCCATATATGGGA  450
     || |  || |  ||     | | | ||| |  |       |||        |
311  AAAAGGCCTGCGAAGCCGAGGGAGCAAAATTAGCAGTCCCA.....A  352

451  TACAAAGCCCTTATATCAATGGAACGATAATGATTGTAATGTTAAAATGG  500
       | | | ||        | ||||   ||    |||  | | ||||   |
353  GAGACAACCACGCCTACGATGGCCTGAAGCAGATCTTCAAGTTAGGGTTT  402
```

```
501  GTTACATATGTGAACCCAATCATTTCCGGGAAACATATGATCAAGCACTC  550
403  GGGGTGTACTGGGCCAACATCGGAATCACAGATCATGAGAGCGAGGGATA  452

551  AAGCAAAAATGCGAAGCAATTAAGATAACAAATTCAAAAATTTCAACAGA  600
453  TTCAGCGGAGTGGATGGTCATCCAGTGTC..GTTCCTGCCATGGAATCCT  500

601  ATTTGATCAATTGCATGCCAAACAATCATTGGAATTTGATAGTATAACGC  650
501  AATGAACCCAACAACCCGGAGGCAACGAGAACTGTGTTAACGTCAACGAC  550

651  AAAATGTAGCAAAAGTGAATGAAGATTGGAAAATTGAAATCCAAAAACTA  700
551  AAAGGACAGCTGAACGACTGGCATTGCGGGATACAGCGCCATTCTTCTGC  600

701  CAGAATGCCACACAAATTGCCATACAACAGATTATGGAAAATCATGAGAA  750
601  GAGCGCCGGCCCTCGGTGGGCATACCACCCTCCTATGTGTGGCTGAAGGA  650

751  GAAGATAAGAGATTTAAGTGATAATCTACTTAAGCAGCTACAAGATTCCA  800
651  CGCGAGCCGCTTCTACAAGGTGCACGCCGAGAAGCACGTGTACGCGGAGG  700

801  ATGAACAACTGAAACAGTCCACTGACCATATGAATGCATCGTTTGGTGAG  850
701  CAGCCAGGGTGTGCCGATCCGAGAACGCGACGCTCGC.......TGTGCC  743

851  AAATTGAAAGGCCAACAAGCAGAAAATAATGAAATTTGTTAAGC..AATT  898
744  CGACACCTGGGACCGTGTCGAGACCCTGCTGCGACTCCTCGAGCCGAAAG  793

899  CCCGGAAAGCCCTTAAAGGATGGCAGTAAATCTTGTGATGTGATATCTTC  948
794  AAGAGTTCTACCTGACAGGATTCACAGATGAGGCTGTGGAAGGTGACTTC  843

949  TGTTTTGTAATTTTACTAGAATTCATGTGAAAATAAAACCAAGTCATTAA  998
844  GTTACCGAAACAGGAAGACACCTAAAAGGCAT                    875
```

Figure 16. Comparison of Sarcophaga Lectin cDNA with Grasshopper 875 bp cDNA. The sequence on top is the Sarcophaga cDNA nucleotide sequence. This alignment was produced by running ALIGN software (in the GCG program package) at the Pittsburgh Supercomputing Center.