1994

CLONING, EXPRESSION, PURIFICATION AND CHARACTERIZATION OF DOMAIN B OF WHEAT GERM AGGLUTININ

Ann C. Rice

Virginia Commonwealth University
School of Medicine

This is to certify that the dissertation prepared by Ann C. Rice entitled **"Cloning, Expression, Purification and Characterization of Domain B of Wheat Germ Agglutinin"** has been approved by her committee as satisfactory completion of the dissertation requirement for the degree of Doctor of Philosophy.

[redacted]

Christine S. Wright, Ph.D., Director of Dissertation

[redacted]

Robert B. Harris, Ph.D., School of Medicine

[redacted]

LaVerne G. Schirch, Ph.D., School of Medicine

[redacted]

Gregory A. Buck, Ph.D., School of Medicine

[redacted]

Paul S. Swerdlow, M.D., School of Medicine

[redacted]

Robert K. Yu, Ph.D., Chairman

[redacted]

Hermes A. Kontos, M.D., Ph.D., Dean, School of Medicine

[redacted]

William L. Dewey, Ph.D., Dean, School of Graduate Studies

June 28, 1994

Date

CLONING, EXPRESSION, PURIFICATION AND CHARACTERIZATION OF DOMAIN B OF WHEAT GERM AGGLUTININ

A dissertation submitted in partial fulfillment of
the requirements for the degree of Doctor of Philosophy
at Virginia Commonwealth University

By

Ann C. Rice

B.S. Virginia Polytechnic Institute and State University, 1981
M.S. Virginia Commonwealth University, 1987

Director: Christine S. Wright, Ph.D.
Research Associate Professor
Department of Biochemistry and Molecular Biophysics

Medical College of Virginia
Virginia Commonwealth University
Richmond, Virginia
August, 1994

ACKNOWLEDGEMENTS

I would like to take this opportunity to thank those who helped me during this learning experience. My daughter, Sara Driskell, deserves special thanks for being so supportive and tolerant. Others I would like to thank for their support and friendship are Paloma de Prada, Christine Kell and her family, Lesley Harris, Ramesh Raghupathi and Jian Zheng. I would like to thank my parents, Ben and Kathy Rice for always believing in me. There are many others at this university who assisted me in providing equipment, supplies, technical expertise and fruitful discussions. Two at the top of this list which deserve special thanks are Dr. Robert B. Harris and Dr. Darrell L. Peterson. In addition, Ms. Esther Branch has been very understanding and very helpful.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

amp - ampicillin

BA - Domain B of WGA with Ala28

BS - Domain B of WGA with Ser28

BSA - bovine serum albumin

cDNA - complementary DNA

core - HBV E_{Ac} truncated core antigen

dATP - deoxy-adenosine triphosphate

EDTA - ethylene diaminetetraacetic acid

EtBr - ethidium bromide

FXa - Factor Xa

GlcNAc - N-Acetyl-D-Glucosamine

HAP - hydroxylapatite

H-bond - Hydrogen bond

HBV - hepatitus B virus

HPLC - high pressure liquid chromatography

HRP - horseradish peroxidase

IPTG - isopropylthio-ß-galactoside

MBP - maltose binding protein

mRNA - messenger RNA

N-Ac - N-Acetyl

NeuLac - N-Acetly Neuramin Lactose

NeuNAc - N-Acetly Neuraminic Acid

nHev - native hevein

PAGE - polyacrylamide gel electrophoresis

rHev - recombinant hevein

SDS - sodium dodecyl sulfate

TFA - trifluoroacetic acid

TLC - thin layer chromatography

WGA - Wheat Germ Agglutinin

Cloning, Expression, Purification and Characterization of Domain B of Wheat Germ Agglutinin

ABSTRACT

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at Virginia Commonwealth University.

Ann C. Rice

Virginia Commonwealth University

Advisor: Christine S. Wright, Ph.D.

Wheat germ agglutinin (WGA) belongs to a family of dimeric chitin binding lectins specific for N-acetly-D-glucosamine (GlcNAc) and N-acetyl-D-neuraminic acid (NeuNAc). The polypeptide chain consists of a tandem repeat of four conserved 4.3 kDa domains (A, B, C, D) stabilized by four disulfide bonds. Saccharide binding occurs in the dimer interface where domains of different subunits are in contact (A with D, B with C). An aromatic amino acid rich pocket on one domain (sufficient for saccharide binding) and a polar region on the contacting domain constitute a complete binding site. Saccharide binding affinities may differ among the four unique sites (eight/dimer) due to sequence divergence. Two equivalent sites/monomer were detected in solution. However, conclusive evidence is lacking as to their locations on the

dimer. To delineate individual specificities and dimerization requirements, it was desirable to isolate and characterize each site independently.

This thesis describes an expression system by which single WGA domains can be efficiently generated as functional proteins. The B-domain was cloned first, because binding was observed at this site in several WGA-oligosaccharide crystal complexes. Two B-domain sequences were cloned varying at residue 28 (Ala→Ser). In a putative domain dimer the polar Ser28 would mimic an H-bond observed to stabilize NeuNAc in the WGA B-site from the contacting C-domain. The domains were expressed as fusion proteins from which they were proteolytically separated and isolated in high yields. The recombinant domains were shown to associate with chitin (poly-GlcNAc). The correct tertiary structure was inferred by saccharide binding ability and antibody recognition. All cysteines were found to be in disulfide linkages. Isothermal titration calorimetry showed that $(GlcNAc)_{3\&4}$ binding to both B-domain mutants is seven-fold weaker than to WGA ($K_d \approx 3.5 \times 10^{-4}$M versus 0.54×10^{-4}M). Binding of N-acetylneuraminyl-lactose was undetectable. Gel filtration, Mass spectral and NMR analysis indicated that the recombinant domains exist as monomers in solution. Thus, the complete WGA binding site was not reproduced and the low affinity reflects only the interactions of the saccharide with the aromatic pocket.

INTRODUCTION

PROTEIN CARBOHYDRATE INTERACTIONS

Protein-carbohydrate interactions mediate a variety of physiological functions. In animals, these include lymphocyte migration, neuronal cell adhesion throughout development (Rademacher et al., 1988), hepatic removal of asialo serum glycoproteins (Drickamer, 1991; Lee, 1992), metastasis (Sharon and Lis, 1989), and glycoprotein hormone-receptor activation (Sairam, 1989; Kobata, 1992). Cell surface glycosylation patterns have been identified which are species specific, tissue specific, and immunogenic. Viruses with surface glycoproteins acquire the species and tissue specific glycosylation patterns of the cell from which they bud (Rademacher et al., 1988). The virus can use these glycosylation patterns to evade the host immune system and to determine cell types it can infect due to carbohydrate pattern recognition by specific cell surface receptors. Several diseases have been identified in which abnormal glycosylation patterns occur on certain membrane bound or secreted glycoproteins. Ongoing research consists of determining whether this is the cause or the result of the disease (Rademacher et al., 1988; and Kobata, 1992). Understanding

the molecular basis of the protein-carbohydrate interaction is important for many areas of research such as cancer, embryology, aging and immunology.

LECTINS

Non-enzymatic carbohydrate binding proteins are classified into two categories, those of immune and those of non-immune origin. Lectins represent the carbohydrate binding proteins of non-immune origin. The protein - carbohydrate interaction is highly specific in which the lectin recognizes a particular saccharide or polysaccharide including the anomeric conformation and linkage (Rademacher et al., 1988). Although lectins have been identified in all forms of life, plant lectins are the best characterized (Sharon and Lis, 1989; Drikamer, 1991).

Plant lectins were discovered more than 100 years ago, yet their physiological function is still speculation. They were initially classified based on the source of erythrocytes they agglutinated. It was eventually determined that the interaction leading to agglutination depended on the presence of specific saccharides, and that agglutination could be blocked or reversed by addition of the saccharide. Plant lectins are now extensively used as a tool to identify specific saccharides on various cell surfaces or glycoproteins. Lectins have been identified in a wide variety

of plants. Those within related families of plants usually have similar sequences and saccharide specificities (Etzler, 1985).

Chitin Binding Lectins

The chitin binding lectins are a well characterized family of plant lectins. Chitin is a polymer of ß 1-4 linked N-acetyl-D-glucosamine (GlcNAc) residues. Members of this family contain one, two or four chitin binding domains. The chitin binding domain encompasses 30 - 43 amino acids with many highly conserved residues, including cysteine, glycine, aromatic and acidic residues (Wright et al., 1991; Raikhel et al., 1993). The tertiary structures have been determined for wheat germ agglutinin (WGA), which has 4 copies of this domain (Wright, 1977) and for hevein, which consists of a single domain (Rodriguez-Romero et al., 1991; Andersen et al., 1993). The chitin binding domain has an irregularly folded structure which is held together by 4 intra-chain disulfide bridges. The α-carbon backbone of the chitin binding domain, derived from the wheat germ agglutinin (WGA) crystal structure (Wright, 1977; Wright et al., 1991) is illustrated in Figure 1.

The chitin binding domain sequence exists in a wide variety of plant proteins. Rice (*Oryza sativa*), barley (*Hordeum vulgare*), rye (*Secale cereale*), couch grass

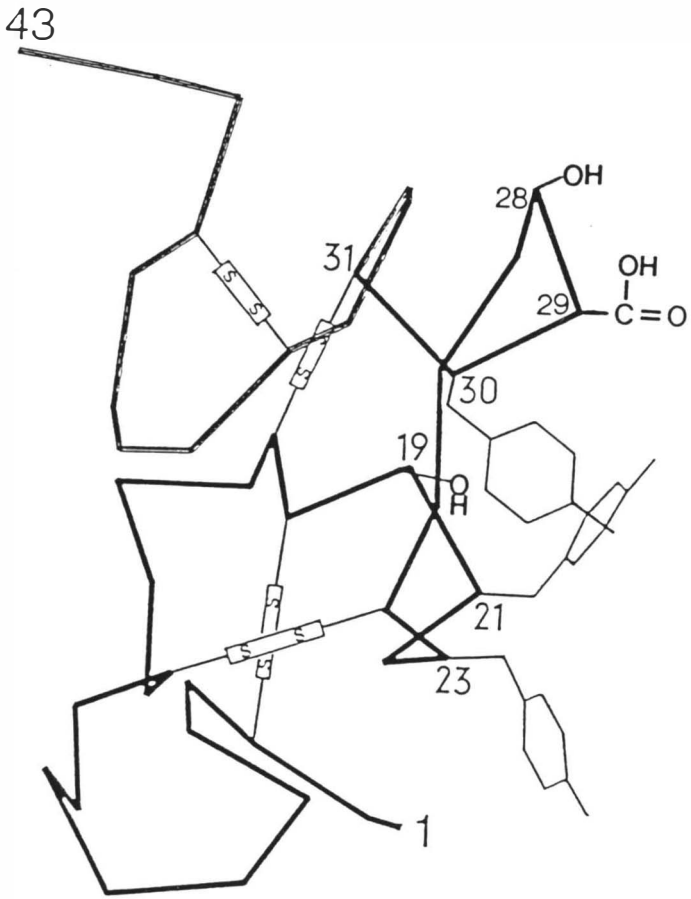


Figure 1. α-Carbon Backbone of the Chitin Binding Domain. The S-S boxes represent the disulfide bridges. The three tyrosines, which function in saccharide binding are shown on the right side of the molecule in the orientation they appear in the domain B binding site. An -OH at position 19 represents a Ser which is part of the domain B aromatic pocket. Amino acids at positions 28 and 29 contribute to the binding site in an opposing domain in WGA.

(Agropyrum repens), false brome grass (*Brachypodium* sylvaticum) and WGA (*Triticum aestivum*) lectins all contain 4 tandem repeats of the domain. The stinging nettle lectin (*Urtica dioica agglutinin*, UDA) consists of two tandem chitin binding domains. Proteins which contain a single chitin binding domain with an unrelated carboxy-terminal sequence are two wound inducible proteins from potato, basic chitinases from kidney bean (*Phaseolus vulgarus*), tobacco (*Nicotiana tabacum*), potato (*Solanum tuberosum*), poplar (*Populus trichocarpa*) and rice, and other lectins from potato, thorn apple (*Datura stramonium*) and tomato (*Lycopersicon esculentum*). Hevein from the latex of rubber trees (*Hevea brasiliensis*) and Ac-AMP1 and Ac-AMP2 from *Amaranth caudatus* consist of only one chitin binding domain sequence (Raikhel, et al., 1993). The amino acid sequences of some of these chitin binding domains are aligned in Figure 2. The high degree of sequence similarity over a broad spectrum of plants indicates an extremely important physiological function of the chitin binding domain (Raikhel et al., 1993). Possible evolutionary conservation of the structure leads to the speculation that the two and four domain proteins arose from gene duplication events of an original single chitin binding domain (Wright et al., 1991).

Figure 2. Amino Acid Alignments of Chitin Binding Domains. The top line of each section is the sequence of WGA I. The sections correspond to each domain of the 4 domain proteins. The disulfide linkages are indicated only for the first sequence listed. The amino acids for the other proteins are shown only if they differ from the WGA I sequence. Dots indicate spaces skipped to optimize alignment. WGA is wheat germ agglutinin, BL is barley lectin, RL is rice lectin, UDA is stinging nettle, RC is rice chitinase, PVC is kidney bean chitinase, WIN 1 & 2 are wound inducible proteins from potato, STC is potato chitinase, TC is tobacco chitinase, WIN 8, 6 2B & 6 2C are wound inducible proteins from poplar, and HEV is hevein (Parsons et al., 1989; Shinshi et al., 1990; Wright et al., 1991 and references therein; Davis et al., 1991; Huang et al., 1991; Van Burren et al., 1992; Raikhel et al., 1993 and references therein).

Amino Acid Sequence Homologies

WGA I - A	QRCGEQGSNM ECPNNLCCSQ YGYCGMGGDY CGKGCQNGACWTS
WGA II - A	
WGA III - A	G
BL - A	Y
RL - A	T K NDG I H F L R T S CS
UDA - A	S GG T ALR I W W ASSP Y R SG

WGA I - B	KRCGSQAGGA TCTNNQCCSQ YGYCGFGAEY CGAG..CQGGPCRAD
WGA II - B	P H H
WGA III - B	K P H H
BL - B	T K P H W
RL - B	Q G S S S N
UDA - B	E A V NP GQLR V H W G ND
RC 1	EQ L P CL W STSD .SQ SGG
RC 2	EQ V P CL F W STSD SQCS LR
PVC	EQ R L PGGN F W STTD P S.Q GGP
WIN 1	Q R K L SG L F W STP F SPSQG S.R TGT
WIN 2	Q R R L G L F W STP SPSQG S.Q TGS
STC	QN G K A ASG K F W NTND S N S.Q PG.
TC B	EQ R ASGL K F W NTND P N S.Q PWG
TC A	EQ R PSGL K F W NTND P N .SQ PGG
TC 3	EQ K R PSGM N F W NTQD P K .SQ PSG
POP	AQ N P DL S G LTVA ... A .. VSS
WIN 8	AQ N P DL S G LTVA C .VSQ NC
WIN 6 2B	EQ Q D L PGGL S W TTAD D .SQ DGG
WIN 6 2C	EQ R D L PGGL F W TTVD D .SQ DGG

WGA I - C	IKCGSQAGGK LCPNNLCCSQ WGFCGLGSEF CGGG..CQSGACSTD
WGA II - C	S
WGA III - C	Y E N
BL - C	Y E G
RL - C	RN N E M Y N CPE
HEV	EQ R W STD SPDHN .N KDS

WGA 1 - D	KPCGKDAGGR VCTNNYCCSK WGSCGIGPGY CGAGCQSGGCDGD
WGA II - D	A
WGA III - D	V
BL - D	A K V
RL - D	R Q D K P F A G Y L GN S YKG

WHEAT GERM AGGLUTININ (WGA)

Background

The most extensively characterized chitin binding protein is WGA. Originally, a mouse tumor cell agglutinating activity was identified in a wheat germ lipase extract (Aub et al., 1963). Wheat germ agglutinin was isolated from this extract in 1967 as the agglutinating agent. N-Acetyl-D-Glucosamine monomers and dimers could inhibit or reverse the agglutination (Burger and Goldberg, 1967; Burger, 1969). Wheat germ agglutinin was also found to bind N-Acetyl-Neuraminic acid (NeuNAc) (Peters et al., 1979; Monsigny et al., 1980). Amino acid compositional analysis of WGA revealed a high Gly and half-Cystine content (Allen et al., 1973; Privat et al., 1974a; Nagata and Burger, 1974; Rice and Etzler, 1975). Molecular weight determination by sedimentation studies and polyacrylamide gel electrophoresis (PAGE) showed that WGA exists as a 35,000 Da dimer under physiological conditions and as a 17,500 Da monomer at low pH or in the presence of denaturants and reducing agents (Allen, et al., 1973; Privat et al., 1974a; Nagata and Burger, 1974; Rice and Etzler, 1974; Rice and Etzler, 1975). The binding constants for various GlcNAc and NeuNAc derivatives were measured by a variety of techniques (Lotan and Sharon, 1973; Nagata and Burger, 1974; Rice and Etzler, 1974; Privat et al., 1974a; Privat et al., 1974b; Privat et al., 1976; Midoux et al., 1980; Kronis and

Carver, 1982). In 1977, the crystal structure of WGA was solved at 2.2 Å resolution (Wright, 1977), prior to determination of the complete amino acid sequence (Wright et al., 1984). Crystallographic studies have defined the sugar binding interactions (Wright, 1980; Wright, 1984; Wright, 1992).

Functional Characteristics

The physiological function of WGA is still obscure. Wheat germ agglutinin from the wheat *Triticum aestivium* exists as three almost identical isoforms, I, II & III (Figure 2). These correspond to genes A, D & B, respectively, with each isoform being encoded on a different diploid genome of this hexaploid wheat (Raikhel and Wilkins, 1987). The WGA mRNA and protein were localized in the tissues of the wheat embryo and seedling that are in direct contact with the external environment (Triplett and Quatrano, 1982; Mishkind et al., 1983; Mansfield et al., 1988; Raikhel et al., 1988; Mansfield and Raikhel, 1990). Certain insect larvae fed large quantities of WGA took almost twice as long to reach maturity (Chrispeels and Raikhel, 1991). It was proposed that WGA binds to the epithelial lining in the gut of these larvae and interferes with nutrient absorption. Based on these observations and the fact that chitin is not found in plants, although it is endogenous to many plant pathogens, it has been

suggested that WGA functions in plant defense (Chrispeels and Raikhel, 1991).

Structural Characteristics of WGA

Under physiological conditions WGA exists as a dimer of identical monomers. Each monomer is composed of four tandem repeats of the chitin binding domain. The domains are referred to as A, B, C & D, starting from the amino-terminus as schematically illustrated in Figure 3A. The two monomers align head to tail to form the dimer. Domain A of one monomer opposes domain D of the other monomer and domain B of one opposes domain C of the other. Saccharide binding occurs in the dimer interface between two opposing domains (Wright, 1980).

A typical saccharide binding site consists of an aromatic pocket on one domain and a polar region on the opposing domain. The saccharide lies in the aromatic pocket and is stabilized by H-bonds contributed from the opposing domain. Each domain has an aromatic pocket and is capable of saccharide occupancy (Wright, 1984; Wright, 1992). Only the A, B and C domains contain polar regions. Comparison of the amino acid residues from all 4 domains which interact with the saccharide indicates that the aromatic region is more conserved than the polar region (Figure 2). The deviations from complete sequence conservation result in 4 different

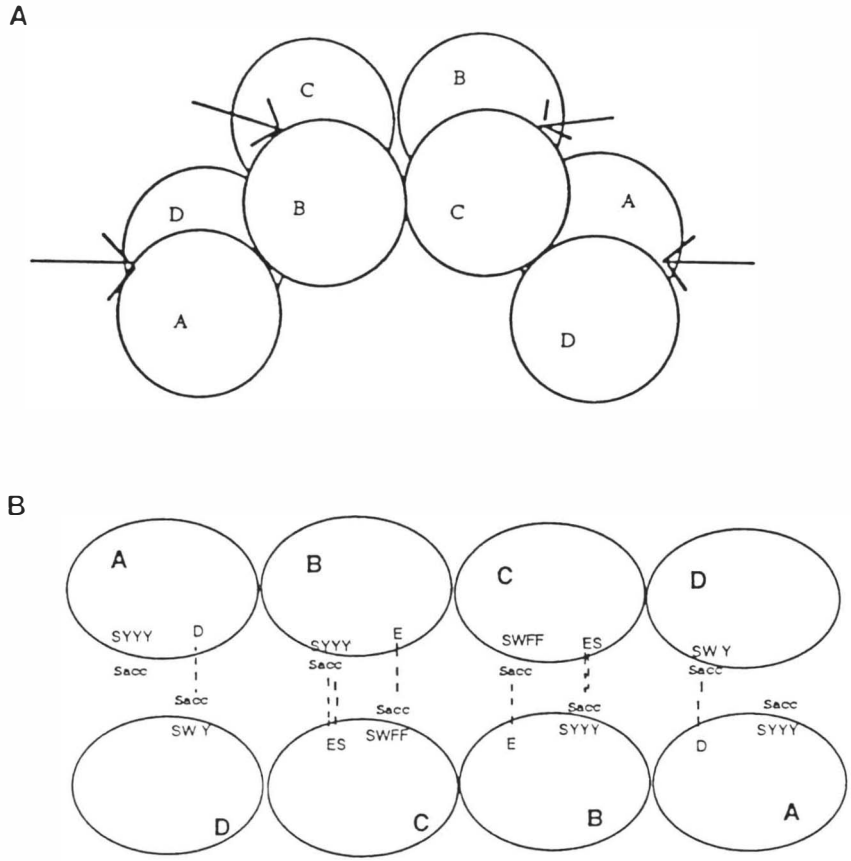


Figure 3. Schematic Representation of the Dimerization in WGA. A, two WGA monomers are depicted aligned head to tail with the domains A - D identified and arrows indicating the location of saccharide binding. B, the two monomers are depicted looking down the dimer interface. All potential saccharide binding sites are represented by 'sacc' The residues in the binding site are designated by the single letter amino acid code. The residues adjacent to 'sacc' constitute the aromatic pocket. The polar residues from the opposing domain are linked to the sacc by dashes which represent H-bonds.

types of binding sites in WGA. The WGA dimer contains two of each of the four types of binding sites yielding a total of eight potential saccharide binding sites (Figure 3B). The aromatic pocket alone is sufficient for saccharide binding as indicated by the fact that the monomeric single domain protein hevein binds to chitin (Van Parijs et al., 1991) and the terminal NeuNAc of a glycopeptide binds in the WGA domain A aromatic pocket, which lacks the polar region from the opposing domain (Wright, 1992). Here, to simplify the terminology, each binding site is referred to by the domain which contributes its aromatic pocket. For example, the B site refers to the site composed of the domain B aromatic pocket and the domain C polar region; the C site is composed of the C aromatic pocket and the B polar region.

Crystal structure studies of several types of WGA-saccharide complexes reveal saccharide binding at each of the potential sites, but not in all 8 sites simultaneously. For instance, occupancy by $(GlcNAc)_2$ was observed in both of the B and D sites (Wright, 1984). The B aromatic pocket makes more van der Waals contacts and was speculated to possess a higher ligand affinity than the D pocket, which lacks one of the aromatic residues. The terminal NeuNAc of neuraminyl lactose (NeuLac) was observed in the B sites only (Wright, 1990). However, binding of the terminal NeuNAc residues from a bivalent sialic acid containing glycopeptide was detected in

an A, B and C site (Wright, 1992). The crystal lattice structure may interfere with occupancy of some of the 8 potential sites in these WGA-saccharide crystal complexes.

Extensive van der Waals and H-bond contacts contribute to the saccharide-WGA interactions (Wright, 1984; Wright, 1987; Wright, 1990). The contacts in the B site (domain B aromatic pocket and domain C polar region) are described here and are similar for the other binding sites. Figure 4 depicts a stereo view of the domain B binding site occupied by $(GlcNAc)_2$. Binding contacts on the GlcNAc and NeuNAc occur at the N-Acetyl group, the pyranose ring and an equatorial -OH on the ring carbon adjacent to the N-Acetyl ring carbon. The aromatic pocket consists of 3 tyrosine residues (domain positions 21, 23 and 30) and a Ser (position 19). The aromatic residues at positions 23 and 30 make van der Waals contacts with the saccharide ring and contribute a hydrophobic surface for the N-Acetyl methyl group, respectively. The Ser interacts through an H-bond with the N-Acetyl carbonyl oxygen. The opposing C-domain contributes H-bonds from polar residues at domain positions 28 and 29. A Glu in domain position 29 forms an H-bond with the N-Acetyl amido group of both GlcNAc and NeuNAc. A Ser in domain position 28 contributes an additional H-bond to the carboxylate of NeuNAc (Wright, 1990).

Solution binding studies of GlcNAc and NeuNac oligosaccharides indicate that there are 2 independent but

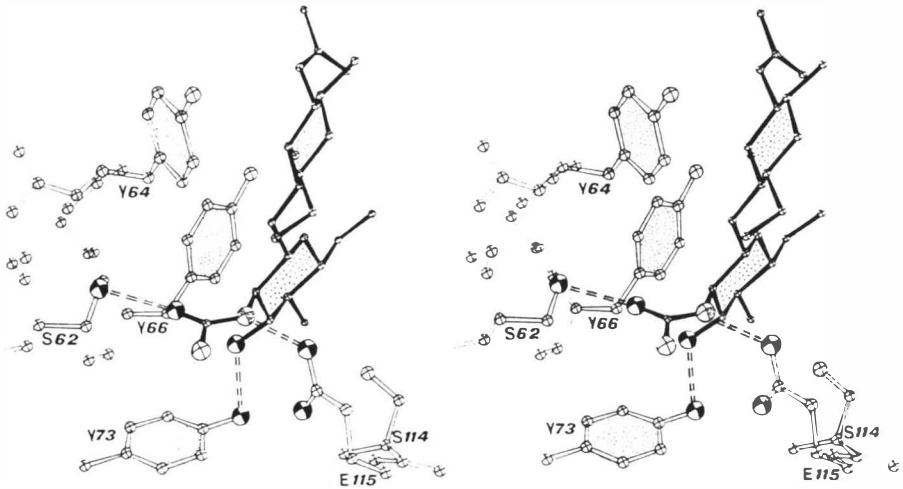


Figure 4. Stereo view of the WGA domain B binding site occupied by $(GlcNAc)_2$. The GlcNAc disaccharide is shown with solid bonds. Atoms involved in the protein/saccharide interactions are enlarged. The partially solid circles represent oxygen atoms, the partially shaded circles represent nitrogen atoms. Hydrogen bonds are depicted by double dashed lines. The amino acid residues are identified by the single amino acid letter code; numbering is according to the domain position. (B) indicates the residue is from domain B of one monomer. (C) indicates the residue is from domain C on the opposing monomer.

equivalent binding sites per WGA monomer (Privat et al 1974b; Nagata and Burger, 1974; Mideux et al, 1980; Kronis and Carver, 1985). These studies do not provide any evidence for which of the 8 potential sites are occupied. However, they do suggest involvement of tyrosine, tryptophan and acidic residues. Domains A, B and D have tyrosines and domains C and D have tryptophans in their aromatic pockets.

RATIONALE

In order to determine which of the 8 possible domain sites delineated in the crystal structure correspond to the 4 sites that bind saccharide with equivalent affinities in solution, each of the 4 types of sites would have to be characterized independently. Based on crystal binding studies, one can speculate that the sites occupied in solution are the B and C sites, because these are the two highly occupied sites in the bivalent sialoglycopeptide-WGA complex (Wright, 1992). Since fewer saccharide contacts are observed at the A and D sites, they are likely to be weaker binding sites, and binding at these latter sites may not be detectable in solution.

The long term project goal is to generate each domain binding site using recombinant techniques and quantitatively characterize the saccharide affinities and specificities. Due to the unavailability of specific proteolytic cleavage sites

between domains in the intact protein, and due to the protease resistant nature of WGA, it is improbable that the intact protein could be proteolytically fragmented into single domain units (Wright, et al., 1984).

Since the B site is the best characterized and the most highly occupied site (as determined by crystallographic analysis; Wright, 1984; Wright, 1992), domain B was cloned first in the present study. In the event that the recombinant domains self-associate in the same fashion as the opposing B and C domains do in WGA, two identical saccharide binding sites should be generated in the interface. The likelihood that domain B dimers could form in solution is substantiated by the observation that several H-bonds and numerous van der Waals contacts exist in the WGA B/C interface (Figure 5). These contacts include an H-bond between Tyr30 -OH (domain B) and Glu29 COO^- (domain C), a water mediated H-bond between Tyr30 (domain B) and the amide nitrogen of Ser28 (domain C), and two reciprocal H-bonds between the backbone α-NH and α-CO groups of Phe26 (domain B) and Leu26 (domain C) (Wright, 1987). The latter two backbone H-bonds will only form if residue 27 of both domains is a glycine. Replacement of this residue by a threonine in monomeric hevein prevents the same type of domain-domain contact. In WGA, many other contacts stabilize the association between opposing monomers.

To recreate the WGA domain B site accurately, we

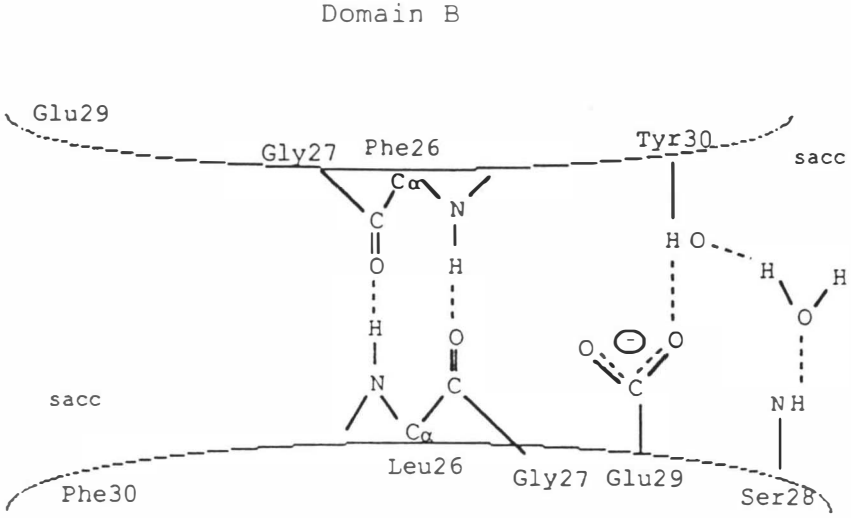


Figure 5. <u>Schematic Representation of the Domain-Domain H-bond Contacts in the WGA B/C Interface.</u> Numbering of the residues is according to the domain position. Dashed lines represent H-bonds.

postulated that an Ala to Ser mutation would have to be introduced at domain position 28 (Figure 3B). This would allow the formation of an additional H-bond from the polar region to the carboxylate of NeuNAc, if NeuLac were to bind. If our hypothesis was correct, this additional H-bond would yield a tighter complex resulting in a better binding constant compared with the unmodified B domain. If, however, the binding site was not faithfully reproduced by dimerization of the B-Ala28 (BA) or B-Ser28 (BS) domain molecules, the magnitude of the binding constants should reflect only the binding contributions from the aromatic pocket residues.

We speculated that we would be able to obtain correctly folded and functional proteins produced in E. coli because functional recombinant barley lectin has been isolated from E. coli (Schroeder and Raikhel, 1992). Barley lectin is highly homologous to WGA with the 32 Cys residues in identical locations, and with similar saccharide specificity and cross-reactivity with anti-WGA antibodies.

This report describes the cloning, expression and purification systems of recombinant BA and BS domain proteins. In this work, the 43 residue monomeric protein hevein was used as a model system because its cDNA was made available to us (Broekaert et al., 1990). Hevein is very similar to the WGA domains in sequence and structure and hevein also binds chitin (Rodriguez-Romero et al., 1989; Van Parijs et al., 1992;

Andersen et al., 1993). The protocols developed to successfully generate functional recombinant hevein were subsequently applied to the BA and BS sequences. The modified BS clone was generated by synthesizing its cDNA and using PCR to generate the site directed change for the native BA sequence. Chitin binding and antibody recognition assays were used to indicate that the domains were correctly folded. The oligomeric state of the recombinant domains were assessed using gel filtration chromatography and mass spectral analysis. Finally, isothermal titration calorimetry (ITC) was used to assess binding between the domain B proteins and GlcNAc and NeuNAc.

MATERIALS AND METHODS

DNA METHODOLOGY

Standard Techniques

All general DNA cloning techniques used are adapted from standard protocols (Maniatis et al., 1982; Berger and Kimmel 1987; Sambrook et al., 1989). Restriction enzymes, ligases, polymerases and nucleotides were purchased from New England Biolabs (NEB), United States Biochemical (USB), or GIBCO BRL, and were used according to the manufacturer's recommendations.

Bacteria, Media & Antibiotics

E.coli type TB1 was used for all cloning work. The E.coli were grown in modified LB media, in which casamino acids were used in place of bactotryptone (10g casamino acids, 5g yeast extract, 5g NaCl, 1g glucose/liter). Bacteria were grown at 37°C with rapid shaking to ensure aeration. All plasmids contained the ß-lactamase gene which renders cells containing the plasmid resistant to ampicillin. Therefore, bacteria containing plasmids were grown in the presence of 50μg/ml ampicillin (amp).

Competent cells

E.coli cells were made competent to take in exogenous DNA by the $CaCl_2$ method (Maniatis et al., 1982; Berger and Kimmel 1987; Sambrook et al., 1989). Briefly, cells were grown to mid log phase, chilled on ice for 10 min, pelleted at 2500 x g, and resuspended in half the original volume with ice cold $CaCl_2$ buffer (10mM Tris, 50mM $CaCl_2$, pH 8.0). They were then incubated on ice for 10 min and repelleted. The cells were resuspended in 1/15 original volume with ice cold $CaCl_2$ buffer and 100μl were used for each transformation experiment.

Transformation of Competent Cells

Freshly prepared competent cells were used for all transformations. DNA from ligation reactions or closed circular plasmids was incubated with the competent cells for 30 min on ice. The cells were subjected to heat shock (42°C) for 2 min and then 0.5ml warm media was added. They were then incubated at 37°C for 1 hr. Aliquots were spread on modified LB media/amp plates and grown overnight at 37°C.

Plasmid Isolation

Bacteria containing plasmids were lysed using the alkali-lysis method (Maniatis et al., 1982; Berger and Kimmel 1987; Sambrook et al., 1989). For mini-plasmid preps (from 10ml overnight cultures) the plasmid was isolated by extracting the

cell lysate with phenol/chloroform and ethanol precipitating the DNA. Qiagen columns (Qiagen Inc.) were used according to the manufacturer's protocol to isolate plasmids in the cell lysate from larger (100-500ml) overnight bacterial cultures.

Electrophoresis

High melting (0.8 - 2.0%) DNA grade agarose gels were used to separate DNA fragments greater than 150 base pairs in length. Electrophoresis was done in a mini submarine rig using a TBE buffer system (89mM Tris, 89mM borate, 25mM EDTA) at a constant voltage (80V) until the bromophenol blue dye band had migrated more than half way across the gel. Smaller fragments of DNA were identified on 8% polyacrylamide gels. These were run in the TBE buffer system at a constant current of 35mAmps. Bands were visualized by ethidium bromide staining under UV light.

Isolation of DNA Fragments

DNA fragments were isolated from the electrophoresis gels by electroelution. The ethidium bromide stained bands were sliced out of the gel and placed in 12,000 - 14,000 Molecular Weight Cut Off (Fisher Scientific) dialysis tubing with TBE buffer. The DNA was electroeluted from the gel into the buffer in the dialysis tubing. The DNA was precipitated from the buffer with cold ethanol.

Synthetic DNA

DNA was synthesized in the DNA Core Facility of the Department of Microbiology and Immunology, Virginia Commonwealth University. The synthesized oligonucleotides are listed in Table 1. They were deblocked in NH_4OH at 55°C overnight in a tightly sealed tube and desiccated in a speed vac evaporator (Savant). The DNA was resuspended in 0.25ml sterile deionized water. DNA less than 35 nucleotides was isolated using Water's Sep-pak C-18 columns following the manufacturer's protocol. DNA longer than 35 nucleotides was purified by gel electrophoresis using 8M urea/12% polyacrylamide gels in TBE buffer. The DNA was visualized by UV shadowing on TLC plates containing a fluorescent indicator. The band representing the full length DNA chain was extracted from the acrylamide by soaking the gel slice in 3M ammonium acetate at 4°C overnight.

DNA Sequencing

The USB Sequenase 2.0 kit was used for DNA sequencing according to the manufacturer's protocol. This protocol follows the Sanger dideoxy method for sequencing. The $^{35}S\alpha dATP$ used for radiolabelling the sequencing reactions was purchased from New England Nuclear. Reaction products were resolved on 10% acrylamide/8M urea gels and visualized by autoradiography. Synthetic DNA oINSSS and the pUC 19 universal forward primer

Table 1. Synthesized Oligonucleotides

oBWH.90	CCC AAG CTT TTA TTA GTC AGC TCT GCA GGG ACC ACC CTG GCA ACC GGC GCC GCA ATA TTC GGA ACC GAA TCC GCA GTA ACC GTA CTG GGA
oBWB.90	GGG GGA TCC AGG GGG GAG CAA TGC GGT TCC CAG GCT GGT GGT GCT ACC TGC ACC AAC AAC CAG TGC TGC TCC CAG TAC GGT TAC TGC GGA
oBWH.18	CCC AAG CTT TTA TTA GTC
oBWB.18	GGG GGA TCC AGG GGG GAG
oINSSS	GAA TTA ATT CGA GCT CGG
o3.210	ATT CGG CGC CGA ATC CGC AGT AAC C

Oligos are listed 5′ to 3′.

were used as the primers for the sequencing reactions.

Plasmids

The plasmid pMal-P was purchased from NEB. This plasmid encodes the maltose binding protein (MBP) with its leader sequence which targets the protein to the periplasmic space. The MBP protein coding region is followed by the Factor Xa (FXa) recognition sequence and a multicloning site. The plasmid has the tac promoter, so protein production is induced with isopropylthio-ß-galactoside (IPTG). HEV N was the gift of Dr. Natasha Raikhel, Michigan State University, and contains the sequence encoding the 43 amino acid hevein domain (Broekaert, et al., 1990). The plasmids pUC 18 and pFS14NSD were the gifts of Dr. Darrell Peterson, Virginia Commonwealth University and Dr. Jian Zheng, Ortho Labs (Zheng, et al., 1992). pUC 18 is a high copy number cloning plasmid. pFS14NSD encodes the E_{AC} truncated core antigen (core ≈ 17,000 Da) from Hepatitis B Virus (HBV). This construct expresses the core antigen in E.coli after the bacteria have reached stationary phase without requiring IPTG induction and contains convenient restriction enzyme sites for cloning to the 3' end of the core sequence to express proteins as core fusion proteins.

GENERAL PROTEIN METHODOLOGY

Growing Bacteria

Cells for MBP-fusion protein production were grown in 300ml modified LB media to an absorbance at 600nm ≈ 0.6. Then IPTG was added to a final concentration of 0.3mM and the cells were grown for another 1.5 - 2 hrs. Core-fusion protein producing cells were grown in TNY media (10g casamino acids, 10g yeast extract, 5g NaCl, 1g dextrose/liter). A 0.5ml aliquot of a 10ml overnight culture was added to 300ml of media. This culture was grown for ≈ 4hrs and then added to 20 liters of media. The 20 liter culture was grown in a well aerated fermenter maintained at 37°C. Optimal fusion protein expression was obtained after 24-30 hrs.

Osmotic Shock

Osmotic shock was performed on the E.coli producing the MBP-hevein fusion protein to release the periplasmic space proteins. Briefly, the cells were harvested by centrifugation (≈ 4000 x g, 4°C, 10 min) were resuspended at 1/15th volume in the sucrose solution (20% sucrose, 10mM Tris, 150mM NaCl, pH 8.0) and incubated for 30 min on ice. The cells were pelleted, resuspended in the same volume with sterile ice cold water, and incubated for 30 min on ice. The cells were again pelleted and the supernatant containing periplasmic space proteins was further processed to obtain the fusion protein.

Lysing Bacteria

The cells harvested from 20 liter cultures of E. coli producing the HBV core antigen-recombinant domain fusion proteins were resuspended in 300-500 ml water. The cells were lysed at 1200 psi in an Aminco French Pressure Cell using the continuous flow adaptation. The particulate matter was pelleted by centrifugation (≈ 8000 x g, 4°C, 30 min) and the fusion proteins were purified from the supernatant.

SDS-PAGE

Polyacrylamide gels were run using the Laemmeli system (Laemmeli, 1970). Gels were prepared as 12, 15, or 18% acrylamide resolving gel and 6% acrylamide stacking gel. A Hoeffer system with 8cm x 8cm size plates and 0.75mm spacers was used. Generally, 5μg of protein were loaded per lane. Running conditions were constant current at 35mAmps until the bromophenol blue dye band was at the bottom of the resolving gel. The gel was stained with 1% Coommassie Brilliant Blue R in 10% acetic acid and 50% methanol for several hours and destained overnight in 7.5% acetic acid and 5% methanol.

MBP-Hevein Fusion Protein Purification

MBP-Hevein fusion protein present in the periplasmic space protein preparations extracted by osmotic shock was purified by amylose affinity chromatography. The amylose column

(prepared as indicated by NEB in Protocols for MBP Vectors) was equilibrated in 10mM Tris, 150mM NaCl, 1% Tween 20, pH 8.0 buffer. The first flow through fraction, as monitored by absorbance at 280nm, was reapplied to the column. The column was washed well with buffer and then the fusion protein was eluted with equilibrating buffer containing 10mM maltose. This yielded relatively pure fusion protein as indicated by a single band on SDS-PAGE.

HBV Core Antigen-Recombinant Domain Fusion Protein Purification

The HBV core antigen (core)-recombinant domain fusion proteins made were core-hevein (CH), core-BS (CBS), which is the core-WGA domain B Ser28 mutant fusion protein and core-BA (CBA), which is the core-native WGA domain B Ala28 fusion protein. They were purified according to the protocol of Zheng et al (1992) for purification of core protein with slight modifications subsequently made in the procedure by Dr. Darrell Peterson (personal communication). The cells were lysed under 1200 psi, the cell lysate supernatant was brought to 45% $(NH_4)_2SO_4$ saturation and the protein pellet obtained after centrifugation was dialyzed against 50mM sodium phosphate, pH 6.8 overnight at 4°C. The dialysate was applied to a hydroxylapatite column (made according to Jenkins, 1962) equilibrated in the same buffer. The column was washed with

1.5 - 2 liters of buffer, until the absorbance at 280nm dropped below 0.15. The flow through fraction was collected and precipitated again at 45% $(NH_4)_2SO_4$ saturation. This latter $(NH_4)_2SO_4$ pellet was resuspended in 50mM sodium phosphate buffer, pH 6.8 and applied to a Sepharose CL4B (Pharmacia) column equilibrated in the same buffer. The core-fusion protein containing fractions (identified by elution profile position as the second peak and the major peak reported by Zheng et al., 1992) were pooled and concentrated again by precipitation at 45 %$(NH_4)_2SO_4$ saturation and dialyzed against 10mM Tris, 150mM NaCl, 2mM $CaCl_2$, pH 8.0. This purification protocol optimally yielded ≈1g fusion protein, which consisted primarily of the core antigen-recombinant domain fusion protein with some fusion protein degradation products as determined by SDS-PAGE and western blot analysis.

Trypsin Digestion of the Core-Recombinant Domain Fusion Proteins

To separate core from the recombinant domain under investigation, trypsin digestion was performed at a 1:100 molar ratio of trypsin to fusion protein, in 10mM Tris, 150mM NaCl, 2mM $CaCl_2$, pH 8.0. Trypsin digestion was either carried out in buffer alone or in the presence of 2M guanidine-HCl as described in the Results section.

Reduction/Alkylation of MBP-Hevein Fusion Protein

A 50 fold molar excess of dithioerythretol (Sigma Chemical Co.) to expected cysteine residues was used to achieve complete reduction of all disulfide bonds in the MBP-hevein fusion protein. The reaction was performed in 10mM Tris buffer, pH 8.8 at 50°C for 30 min. A 100 fold molar excess of iodoacetamide to cysteine residues was subsequently added and the sample was incubated at 50°C for another hour.

Western Blot Analysis

The protocol of Burnette (1981) was followed for western blot analysis. Following SDS-PAGE the acrylamide gel was soaked in transfer buffer (20mM Tris, 150mM glycine, 20% methanol) for ≈ 10 minutes. One sheet of 0.45μm nitrocellulose (Kodak) and 4 pieces of #1 Whatman filter paper were cut to the size of the resolving gel and wetted in the transfer buffer. The transfer layers were stacked on the bottom electrode of the Millipore Polyblot "semi-dry" transfer apparatus taking care to exclude air bubbles. Two sheets of filter paper were laid down first, followed by the sheet of nitrocellulose, the gel, two more sheets of filter paper and then the top electrode. The transfer was accomplished using constant current at 2.5mAmps per cm^2 of gel, but not exceeding 200mAmps for 30 - 45 minutes. Following transfer the gel was soaked in Coommassie Brilliant Blue stain and the

nitrocellulose sheet was soaked in 25ml blocking buffer (10mM Tris, 150mM NaCl, 1% w/v BSA, pH 8.0). The solution was shaken gently at room temperature for 2 hours, then an appropriate dilution of antiserum (1:2500 polyclonal rabbit anti-WGA (Sigma Chemical Co.); 1:1000 polyclonal rabbit anti-core the generous gift of J. Zheng and D. Peterson) was added to the blocking buffer. Incubation overnight at room temperature was followed by removal of the first antibody buffer and washing 3 times with wash buffer (10mM Tris 0.5M NaCl, 0.1% v/v Tween 20, pH 8.0). The blot was then incubated with the second antibody in 10mM Tris, 150mM NaCl, 1% w/v BSA, pH 8.0 buffer and 10μl horseradish peroxidase (HRP) conjugated goat-anti-rabbit antisera was added. After 2 hours at room temperature the blot was washed again 3 times with buffer, then developed with a solution of 8ml of TBS buffer (10mM Tris, 150mM NaCl, pH 8.0) and 2ml 4-chloronaphthol (25mg/ml in ethanol) and 5μl H_2O_2. Immunoreactive bands developed within 5 minutes.

Protein Concentration Determination

The Pierce BCA protein assay kit was used for most protein concentration determinations according to the manufacturer's protocol. BSA provided with the kit was the protein standard used. The molar extinction coefficients

(WGA≈29,320 $M^{-1}cm^{-1}$; hevein≈13,600 $M^{-1}cm^{-1}$; BS & BA≈4320 $M^{-1}cm^{-1}$) and absorbance measurements at 280nm were used to determine protein concentrations of purified proteins.

Fluorometry

A Shimadzu fluorometer was used to detect the intrinsic tryptophan fluorescence intensity for the chitin binding assay and the shifts in λ_{max} of WGA, nHev and rHev after incubation with tri- and tetra-GlcNAc oligosaccharides. One ml of sample was used for each measurement. The samples were excited at 290nm, at the maximum absorbance of tryptophan. Fluorescence intensity values were determined at the emission λ_{max} for WGA and hevein (349 & 347nm, respectively). Shifts in λ_{max} were monitored by scanning between 320 - 370nm for hevein and WGA.

Chitin Binding Assay

A chitin binding assay was developed to screen for the presence and function of the recombinant domain proteins (hevein, BS and BA). Chitin was weighed into an Eppendorff tube and 1.2ml of the test protein solution was added. Amounts of protein used were nhev≈8.3μg/ml, rhev≈25μg/ml, WGA≈42μg/ml, BA≈420μg/ml and BS≈330μg/ml. The tube was laid on its side to maximize surface exposure and incubated at room temperature for 1 hour. The chitin was pelleted in a microfuge and the relative amounts of protein remaining in the

supernatant were determined by either intrinsic tryptophan fluorescence or protein concentration determination using the Pierce BCA kit.

Amino Acid Compositional Analysis

Amino acid compositional analyses were performed in the Protein Core Facility in the Department of Biochemistry and Molecular Biophysics Department, Virginia Commonwealth University, Richmond, VA. Single time point acid hydrolyses were performed on samples of hevein, BS and BA at varying stages of purity. Most samples were reduced and alkylated before hydrolysis.

N-Terminal Sequencing

N-terminal sequence analysis of the recombinant proteins was done by automated Edman degradation at the Protein Core Facility in the Department of Biochemistry and Molecular Biophysics, Virginia Commonwealth University, Richmond, VA.

Mass Spectral Analysis

Matrix Assisted Laser Desorption Ionization - Time of Flight (MALDI-TOF) mass spectral analysis was performed at the Dana Farber Cancer Institute, Boston, MA on samples of the BA protein. Samples were prepared in 50 mM sodium phosphate buffer/pH 6.8, with 35 fold molar excess $(GlcNAc)_3$, and at

pH<2.0 that would yield between 10-20 pmoles in 0.5μl. Mass spectral (MALDI-TOF) analysis was performed by Commonwealth Biotechnologies, Inc., Richmond, VA on samples of the BS protein in water. The analysis were performed using two types of matrices, α-cyano and sinapinic acid.

Free Sulfhydryl Determination

The free sulfhydryl determination for rHev was performed by Dr. Darrell Peterson, Dept. of Biochemistry, Virginia Commonwealth University, using 5,5'-dithiobis(2-nitrobenzene). The free sulfhydryl content of the BA and BS proteins was determined by alkylation with 80 molar excess of iodoacetamide over protein in the presence of 6M guanidine-HCl. The moles of carboxymethyl cysteine was quantitated by amino acid compositional analysis.

HPLC Gel Filtration

The HPLC SEC 2000 GF column purchased from Phenomenex was used to estimate molecular weights of BS and BA. The column was equilibrated in 50mM sodium phosphate, 150mM NaCl, pH 6.8 and operated at a flow rate of 1 ml/min. The elution profile was monitored at 280nm. Molecular weight standards used were bovine serum albumin (BSA) at 64,000 Da, soybean trypsin inhibitor (SBTI) at 21,500 Da, ribonuclease A (RNAsa A) at 13,700 Da and bovine pancreatic trypsin inhibitor (BPTI) at

6,400 Da. A linear regression of retention time vs log molecular weight yielded a regression coefficient of 0.998.

HPLC Reverse Phase (C-18)

Reverse-phase HPLC was used to help identify the recombinant domain proteins and determine their purity. The column (Spherisorb RP-18, 4.6mm x 25cm, 5 micron, Column Resolution Inc.) was equilibrated in 0 or 10% (v/v) acetonitrile, 0.1% (v/v) trifluoroacetic acid in water (TFA). The column was developed at 1ml/min using a gradient of 0 or 10% solvent B to 50% solvent B in 25 or 20 min, respectively. The elution profile was monitored at an absorbance of 220 or 280nm.

Isothermal Titration Calorimetry (ITC)

Isothermal titration calorimetry experiments were done to determine the binding constants and thermodynamic parameters of the saccharide binding to BS, BA, hevein and WGA. Samples were prepared in 50mM sodium phosphate, pH 6.8 and degassed prior to each experiment. Protein concentrations were 0.24 - 0.7mM for BA and BS, 0.087mM for native hevein and 0.17mM for WGA. Protein concentrations were determined by absorbance measurements using calculated molar extinction coefficients. Theoretical saccharide concentrations in the dropping syringe were either 10 or 20mM. The protein solution was loaded into

the sample cell of the MicroCal Omega instrument, taking care to remove all air bubbles. The solution was stirred at 400 rpm throughout the experiment. For each experiment 15-20 10μl injections (of 10 sec duration) were made at 3 minute intervals. The heat (Q, area under each peak) was corrected for injection of saccharide into buffer alone and the cumulative heat released (Q_T) vs the ligand concentration was plotted to generate a binding isotherm, using a non-linear fit according to the binding equation:

$$Q=V[M](n\Delta HK_a[L]/1+K_a[L]) \quad (1)$$

where V is the cell volume, [M] is the molar concentration of protein, n is the number of ligand binding sites per macromolecule, ΔH is the enthalpy of binding, K_a is the association constant and [L] is the molar concentration of free ligand (Freire, 1990; Wiseman et al, 1989).

RESULTS

PLASMID CONSTRUCTS GENERATED

We chose to express the recombinant proteins as fusion proteins to facilitate purification and isolation of the 43 amino acid residue domains. Although two types of fusion protein systems were employed, we could isolate the recombinant domains from only one. Hevein was the first protein cloned and expressed in order to develop the protocols to be applied to the domain B proteins. A total of five constructs were generated to produce the three recombinant domains. The amino acid sequences of these domains (rhev, BS and BA) are shown in Figure 6.

pMN

The sequence encoding the mature hevein protein was initially cloned into the pMal-P expression vector to be expressed as a maltose binding protein (MBP) fusion protein (MBP-Hev). The pMal-P vector is a commercially available vector, which is designed to express high levels of fusion protein. Ideally, the fusion protein is purified on an amylose affinity column and the fusion partners separated by cleavage with FXa.

```
        1         10        20        30        40    43
BS   GEQCGSQAGGA TCTNNQCCSQ YGYCGFGSEY CGAG..CQGGPC RAD
BA                              A
RHEV      R    K L P  L     W W  STD    SPDHNCQSN.C KD
```

Figure 6. Amino Acid Sequences of the Recombinant Domains. BS, WGA domain B-Ser28 mutant; BA, native WGA domain B-Ala28 sequence; rHev, recombinant hevein sequence. Only residues in the BA and rHev sequences differing from the BS sequence are indicated.

The Hev-N plasmid contains the hevein cDNA. Figure 7 depicts the cloning protocol utilized to generate the MBP-hev fusion construct. The DNA sequence encoding the 43 amino acid mature hevein protein is flanked by a 5′ Sma I restriction enzyme site and a 3′ Hind III restriction enzyme site in the Hev-N plasmid. The multi-cloning site (site with numerous unique restriction enzyme recognition sequences) of pMal-P has a Stu I restriction enzyme site located 5′ of a Hind III site. Stu I and Sma I both produce blunt ended products. The hevein sequence was isolated from Hev-N by cleavage with Sma I and Hind III and ligated into pMal-P at the Stu I and Hind III sites, yielding the pMN vector. Positive clones were identified by the Bgl II restriction enzyme digestion pattern (Figure 8) and verified by Western blot analysis described later. Bgl II cleaves once in the vector and once in the insert to release a 1 kilobase fragment (lane 6). The MBP-Hev fusion protein was expressed and purified according to the protocol described in Materials and Methods. The extent of digestion with FXa was monitored by a shift in molecular weight size on SDS-PAGE from MBP-Hev fusion protein (46 kDa) to MBP (42 kDa) alone. Applying the protocol recommended by NEB, digestion was found to be inefficient, although the FXa activity was verified by its ability to hydrolyze a colorimetric substrate (N-Benzoyl-I-E-G-R-p-nitroanilide, Sigma Chemical Co.). Thus, the MBP-fusion protein system was

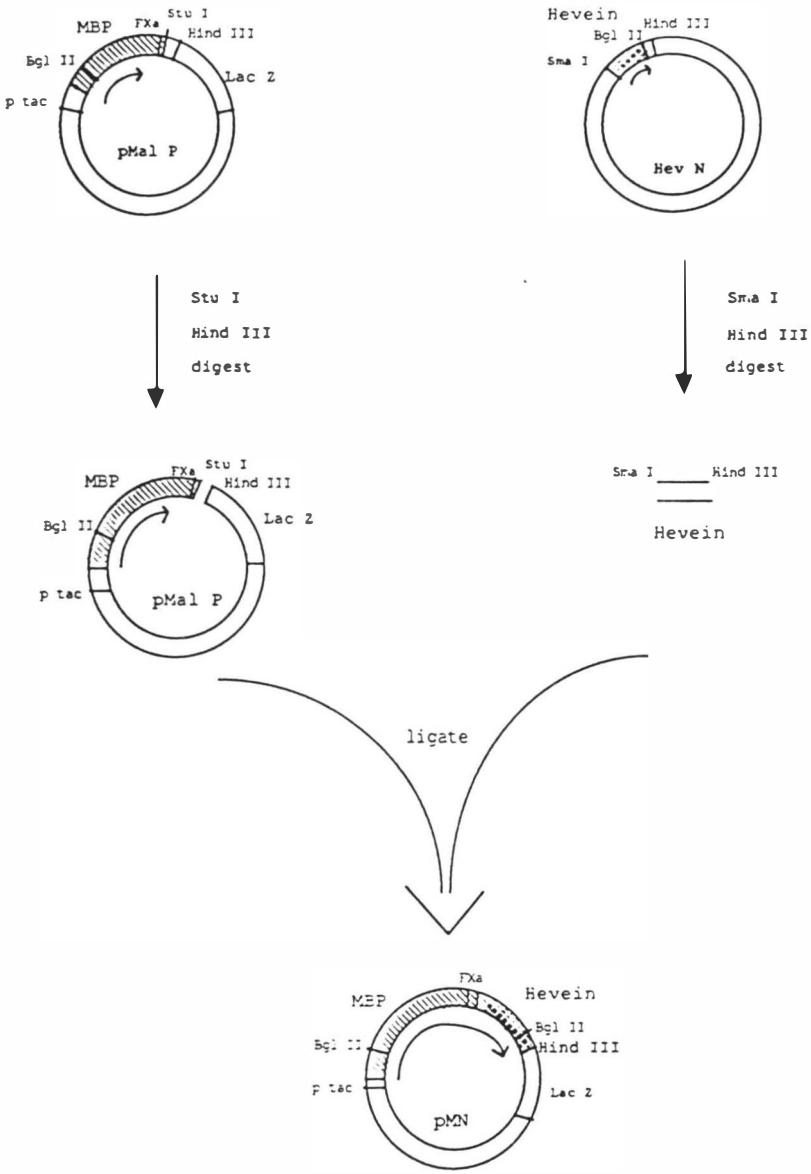
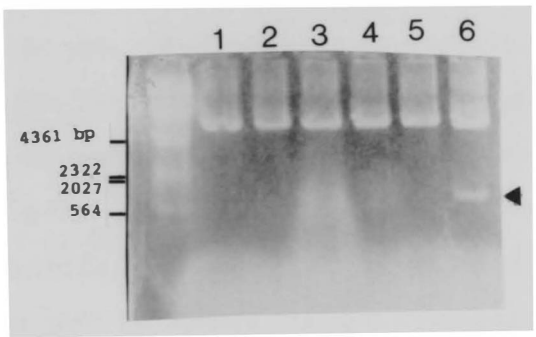


Figure 7. pMN Construction.

Figure 8. Bgl II Digestion of pMN. The first lane on the left shows molecular weight markers. Lanes 1-6 are plasmids extracted from different colonies and digested with Bgl II. The arrow indicates the excised 1 kilobase band from the positive clone.


1
2
3
4
5
6
4361 bp
2322
2027
564

abandoned as the fusion protein expression system for this project.

p18N

To place the hevein cDNA into another expression vector required subcloning into an intermediate vector with useful restriction enzyme cleavage sites. The protocol utilized for cloning the hevein cDNA into pUC18 is depicted in Figure 9. The pUC 18 vector was chosen because of its extensive multi-cloning site. Hence, pMN and pUC 18 were cleaved with Sst I and Hind III. In pMN, this releases a 180 base pair fragment containing the FXa recognition sequence followed by the hevein cDNA. In pUC 18 both sites are in the multi-cloning site. The 180 base pair hevein fragment was then isolated as described in Materials and Methods and ligated using T4 DNA ligase into the pUC 18 vector at the compatible restriction enzyme sites. This cloning protocol produced p18N in which the FXa-hevein coding sequence is flanked by numerous restriction enzyme sites. The FXa site was maintained in this construct due to location and availability of restriction enzyme sites. Positive clones were identified by Ssp I restriction enzyme digestion patterns (data not shown). There is one Ssp I site within hevein and one site in the vector, which releases a 640 base pair fragment.

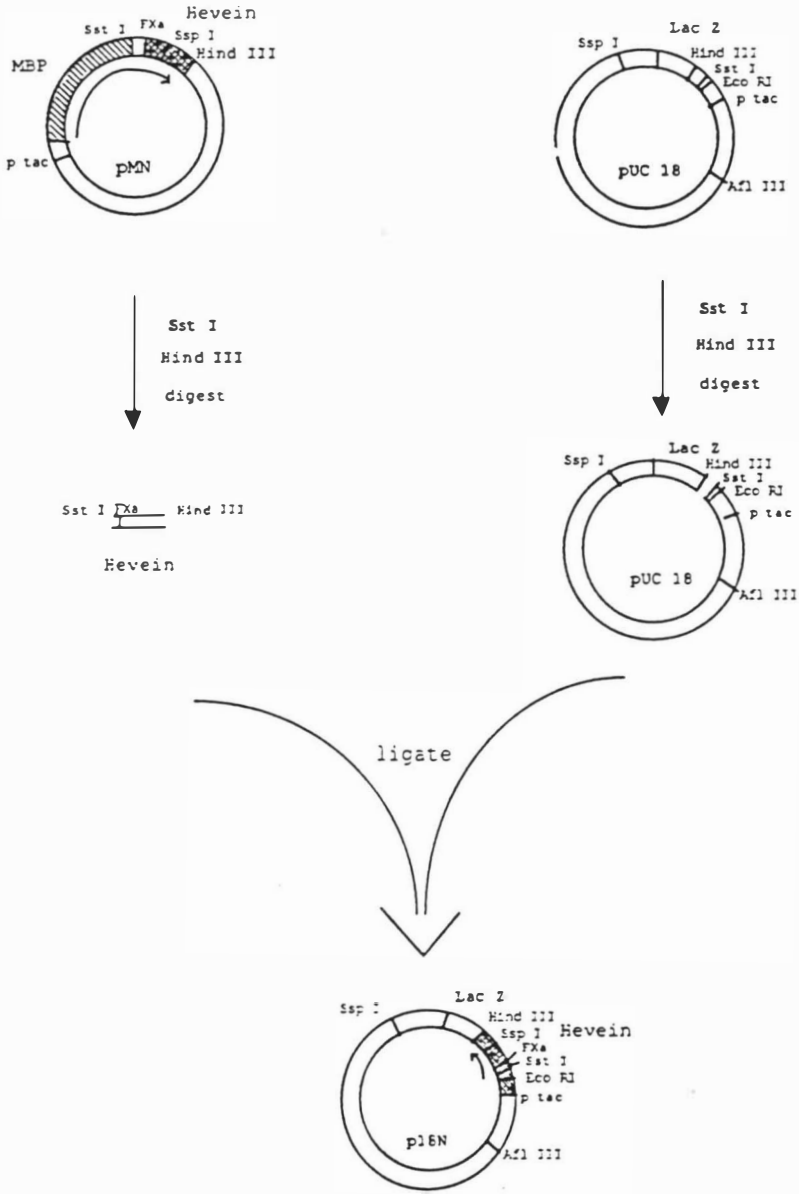


Figure 9. p18-N Construction.

pCH

In the next expression vector hevein was fused to the C-terminus of the HBV core antigen (core). This system was chosen because the core antigen is expressed at moderate levels without requiring induction. The core protein is purified fairly easily from E.coli extracts (Zheng, et al., 1989 and D. Peterson personal communication). Figure 10 depicts the cloning protocol utilized to generate the core-hevein fusion protein construct. The plasmid pFS14NSD encodes the core antigen with a 3' Eco RI site. There is an Eco RI site at the 5' end of the FXa recognition sequence in p18N. To fuse the hevein cDNA to the C-terminus of the core antigen in the correct reading frame required digestion of both plasmids with Eco RI and filling in the resulting 5' overhangs with the Klenow fragment of DNA polymerase. The Eco RI and Klenow were heat inactivated. Both plasmids were subsequently cleaved with Afl III. The core antigen sequence was isolated as an 800 base pair fragment from this digestion. In p18N a 500 base pair fragment 5' of the hevein cDNA was released and the remaining 18N plasmid was isolated. The core antigen sequence was then ligated 5' of the hevein gene into the p18N vector through the Afl III and blunt ended sites. Positive clones were identified by cleaving plasmids with Bgl II, which cleaves once in hevein and once in core to release a 350 base pair fragment (Figure 11), and verified by western blot

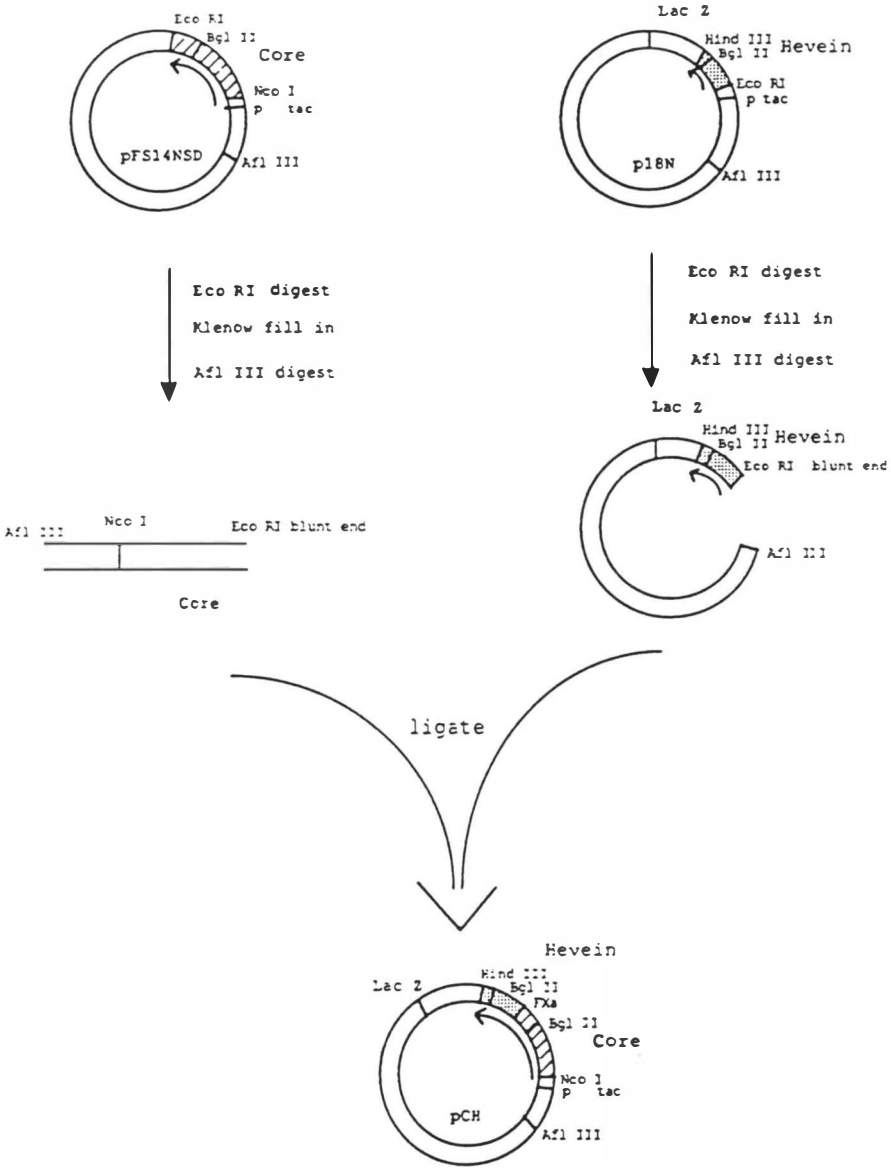
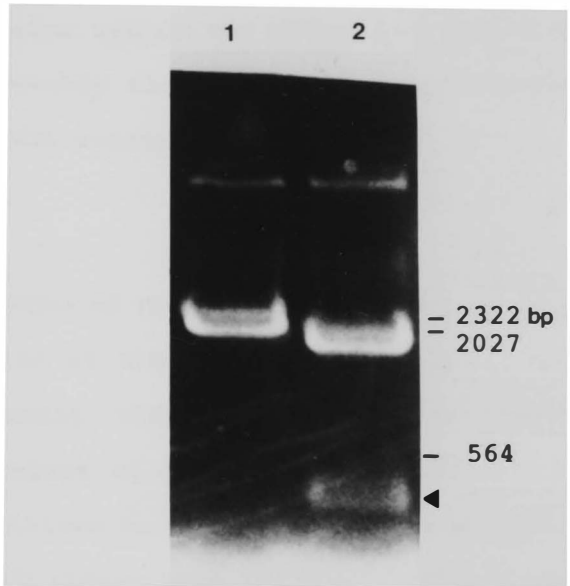


Figure 10. pCH Construction.

Figure 11. pCH Digestion with Bgl II. Lane 1 is the plasmid linearized with Nco I. Lane 2 is the plasmid digested with Bgl II. The arrow indicates the 350 base pair fragment released from a positive clone.


1
2
2322 bp
2027
564

analysis described later. This clone produces a core-hevein fusion protein, which was purified as described in Materials and Methods. The core-hevein fusion protein retains the FXa recognition sequence between core and hevein. The FXa never efficiently released hevein from core, but trypsin was found to be effective for the proteolytic separation as described later under purification of the recombinant domains. Since this expression system was effective, the domain B sequences were subsequently cloned in a similar fashion into the core antigen fusion construct.

pCBS

Generation of the WGA domain B Serine (BS) clone involved DNA synthesis of the complete sequence. This sequence was amplified using the polymerase chain reaction (PCR) and ligated in place of hevein in pCH. Figure 12 shows the PCR strategy utilized to obtain the domain BS cDNA. Two synthetic 90 nucleotide chains were prepared, oBWH.90 and oBWB.90, which encode approximately half of the desired protein and which contain Hind III and Bam HI restriction enzyme sites on their 5' ends, respectively. They overlap in 21 base pairs, which were annealed and elongated to generate the full length double stranded clone. The elongation reaction was carried out using each 90-oligomer at 250nM and the thermostable Vent polymerase (NEB) according to the manufacturer's recommendations. This

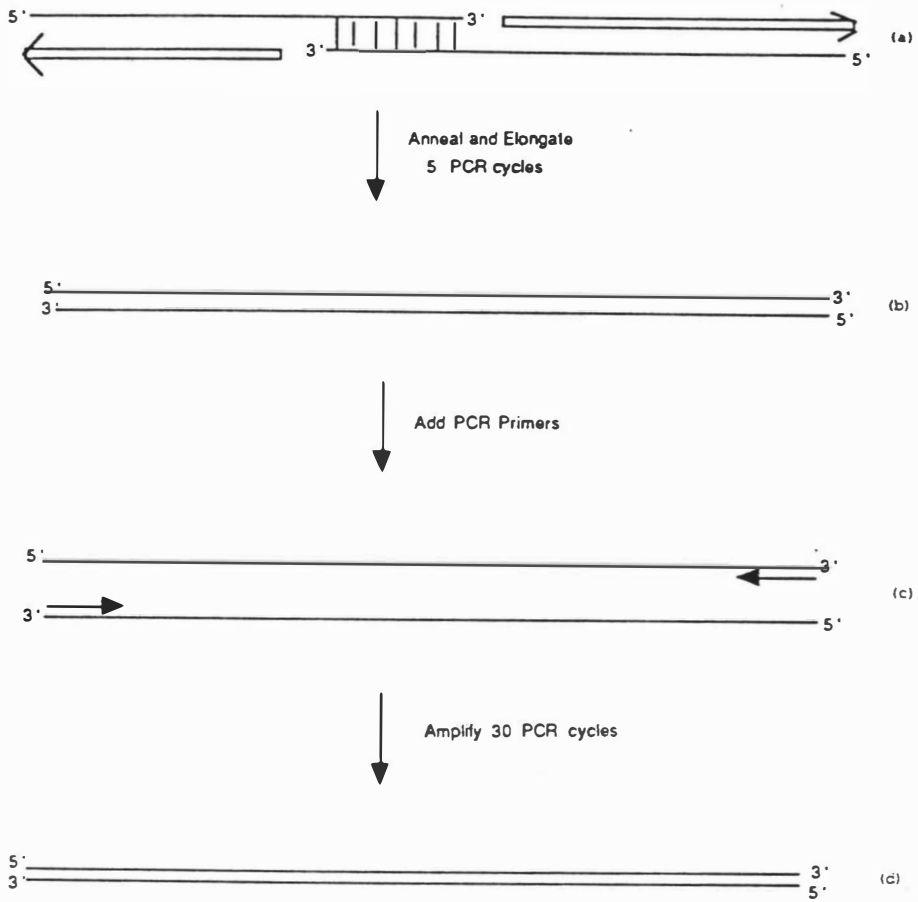
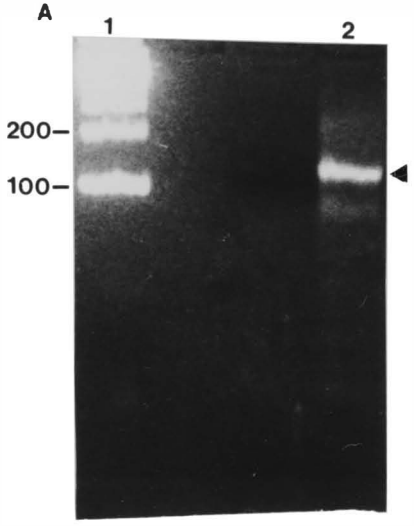
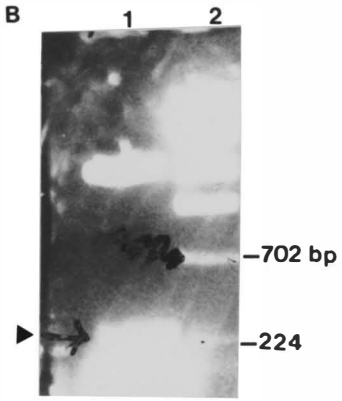


Figure 12. pCBS PCR. (a) The solid lines represent the synthetic oligonucleotide primers, oBWH.90 and oBWB.90. The vertical lines represent the 21 bp overlap of the 90-mers, which can anneal to each other. The hollow arrows indicate the elongation products of the oligonucleotide primers. (b) represents the product of the elongation reaction. (c) depicts the second set of shorter primers, oBWH.18 and oBWB.18 as small arrows, which are used to PCR amplify the elongation product. (d) depicts the product of the PCR amplification reaction.

reaction consisted of 5 PCR cycles of 1 minute at 94°C, 1 minute at 52°C, and one-half minute at 72°C. Two synthetic 18 nucleotide chains, oBWH.18 and oBWB.18, identical to the 5′ ends of each 90-oligomer, were used for the amplification reaction. The amplification reaction used 5 μl of the elongation reaction as the template, 1μM concentration of each 18-oligomer and Vent polymerase at 94°C for 1 minute, 37°C for 1 minute, and 72°C for one-half minute for a total of 35 cycles. Figure 13A shows the products of the PCR elongation-amplification reaction. The PCR product (BS cDNA) was isolated and cleaved with Bam HI and Hind III so it could be ligated into the plasmid at compatible restriction sites. Figure 14 shows the cloning strategy utilized to obtain the core-BS fusion protein construct. The plasmid pCH was cleaved with Bam HI and Hind III. The BS sequence was ligated 3′ of the core antigen in place of hevein at the compatible restriction enzyme sites. The complete FXa recognition sequence was omitted in this clone since FXa digestion was never effective. However, Arg was retained to allow trypsin digestion to cleave the core-BS fusion protein. Positive clones were identified by Pst I restriction enzyme digestion pattern, which cleaves once in the BS sequence and once in core releasing a 310 base pair fragment (Figure 13B). Positive clones were verified by DNA sequencing through the entire inserted region. Fusion protein production was verified

Figure 13. pCBS PCR and Digestion with Pst I. A shows the PCR product from the elongation and amplification reactions. Lane 1 is molecular weight markers. Lane 2 is 15μl of the 100μl PCR reaction. The arrow indicates the ≈160 base pair PCR product. B shows the Pst I digestion pattern of pCBS. Lane 1 is the products of the Pst I digestion of pCBS. Lane 2 is molecular weight markers. The arrow indicates the 310 base pair excised fragment in the correct clone.


A
1
2
200–
100–


B
1
2
–702 bp
–224

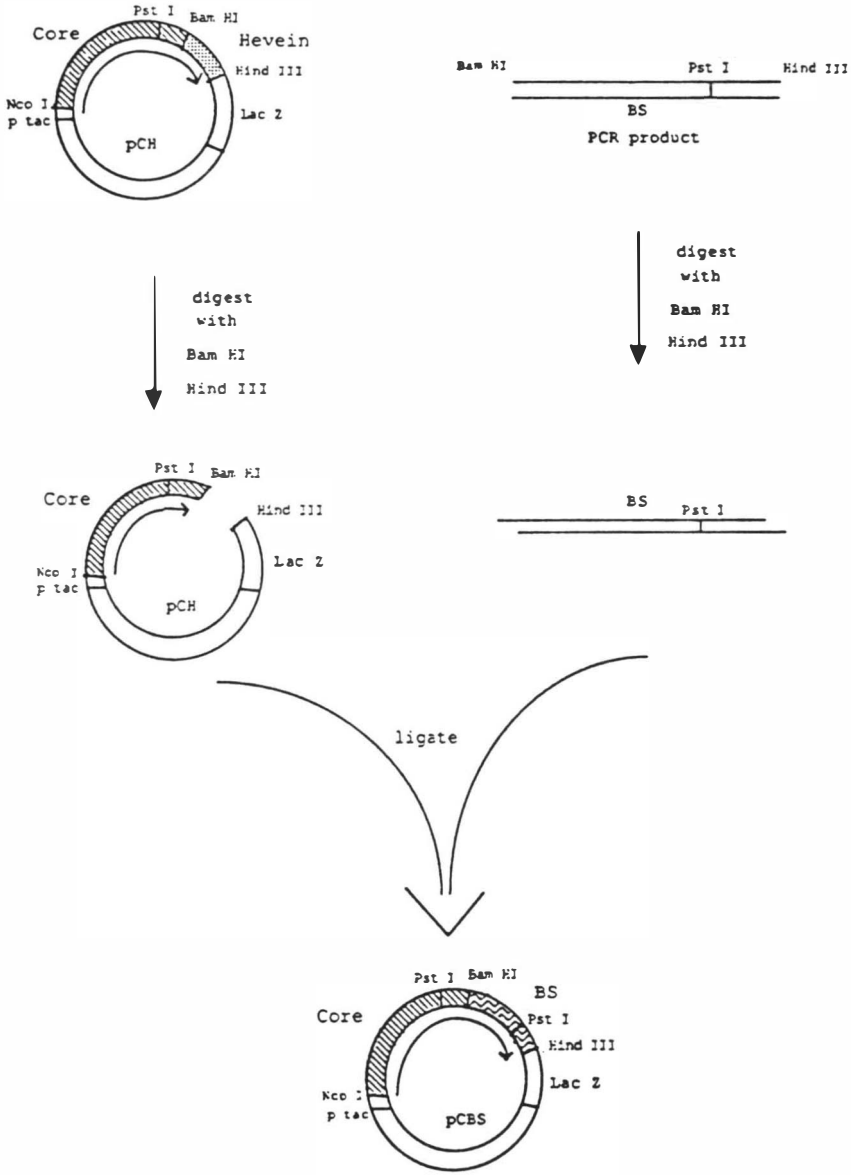


Figure 14. pCBS Construction

by western blot analysis. The clone produces a core-WGA domain B Serine (CBS) fusion protein, which was purified as described in Materials and Methods for the core-recombinant domain fusion proteins.

pCBA

Generation of the WGA domain B Alanine (BA) clone involved PCR amplification of the 5′ half of the BS sequence with 2 nucleotide changes in one primer. This primer encodes an alanine instead of a serine at domain position 28. The PCR product is ligated in place of the 5′ half of the BS sequence. Figure 15 shows the cloning strategy utilized to obtain the core-BA fusion protein construct. The end of one PCR primer, o3.210, contains the 3′- half of the Ssp I recognition sequence (Ssp I produces blunt ended products), the 2 nucleotide changes and 17 nucleotides that can anneal to the sequence being amplified. The other PCR primer, oINSSS, has a Bam HI recognition sequence and can anneal 5′ of the BS sequence. Amplification was carried out using pCBS as a template and oINSSS and o3.210, at 3μM, as the PCR primers for 25 cycles at 90°C 1.5 min, 30°C 1 min and 72°C 1 min. The PCR product was isolated and cleaved with Bam HI. pCBS was partially digested with Ssp I, since there are 2 sites in the plasmid. One Ssp I site is located in the desired cloning site and the other is located 600 base pair away in the

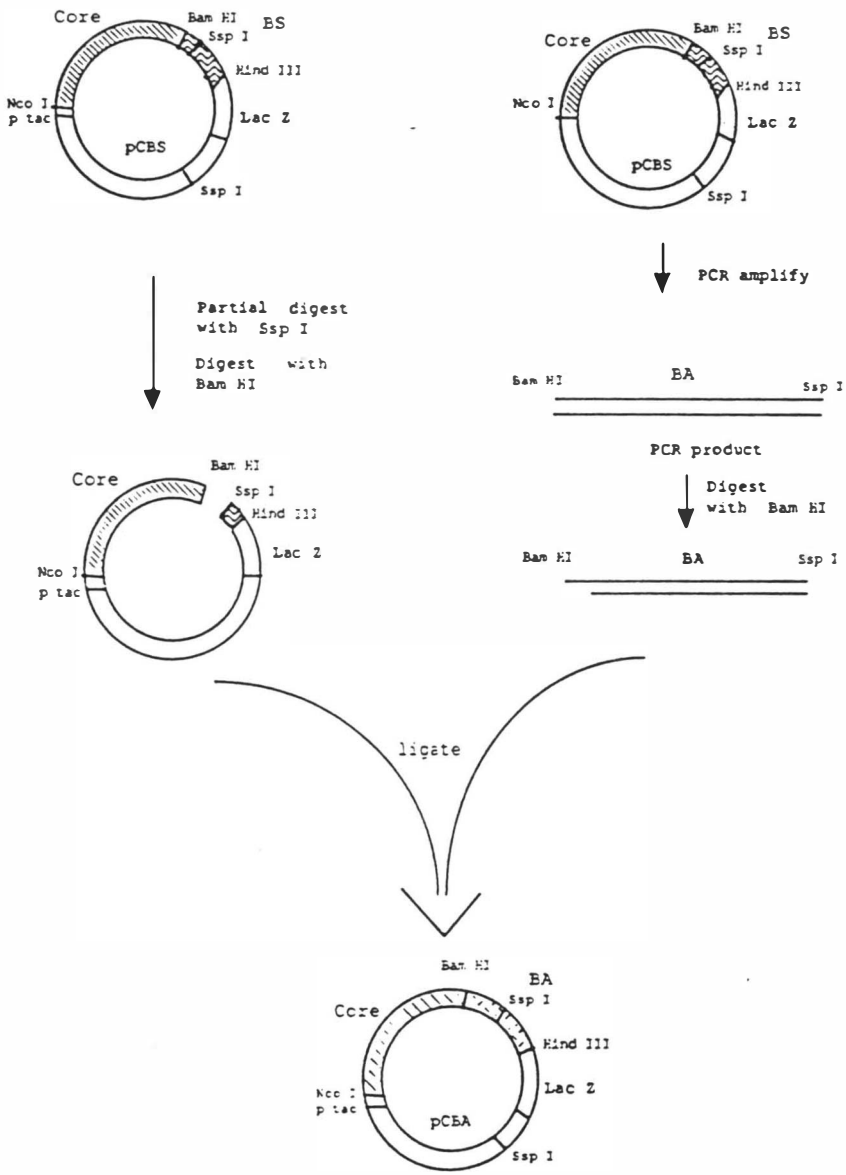
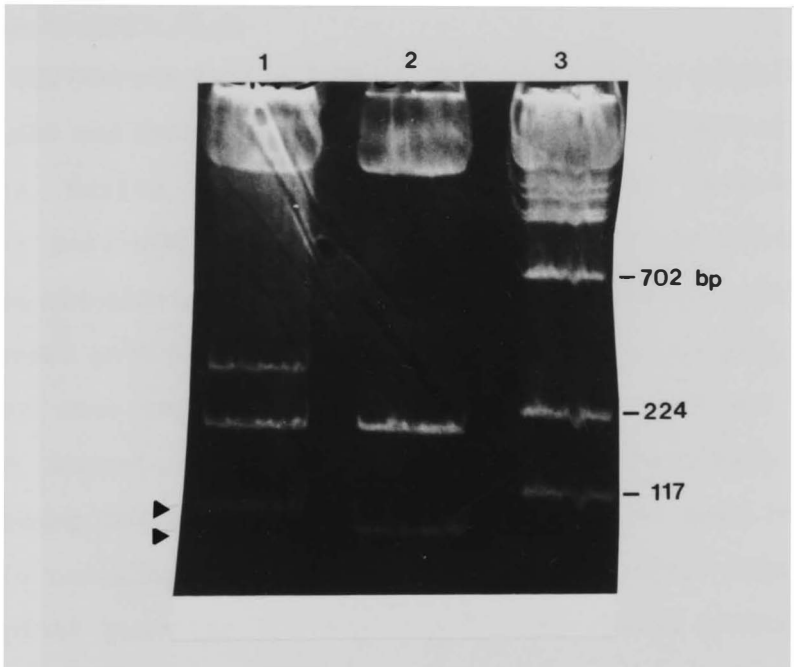


Figure 15. pCBA Construction.

vector. The linearized plasmid band was isolated and cleaved with Bam HI to allow discrimination of the products cleaved at different Ssp I sites and to produce the appropriate end for ligation purposes. The higher molecular weight plasmid band, which contains the restriction enzyme digested ends at the desired locations, was isolated. The PCR product was ligated into the plasmid at the compatible sites. Clones with the 2 nucleotide changes were identified by digesting with Nar I, which releases an 87 base pair fragment in BA compared to a 103 base pair fragment in BS (Figure 16). The 2 nucleotide changes were verified by DNA sequencing. This clone makes a core-WGA domain B Alanine (CBA) fusion protein, which is purified as described in Materials and Methods for the core-recombinant domain fusion proteins.

Figure 16. Digestion of pCBA and pCBS with Nar I. Lane 1 is pCBS digested with Nar I. Lane 2 is pCBA digested with Nar I. Lane 3 is molecular weight markers. The upper arrow indicates the 103 base pair band released from pCBS upon Nar I digestion. The lower arrow indicates the 87 base pair band released from pCBA upon digestion with Nar I.


1
2
3
702 bp
224
117

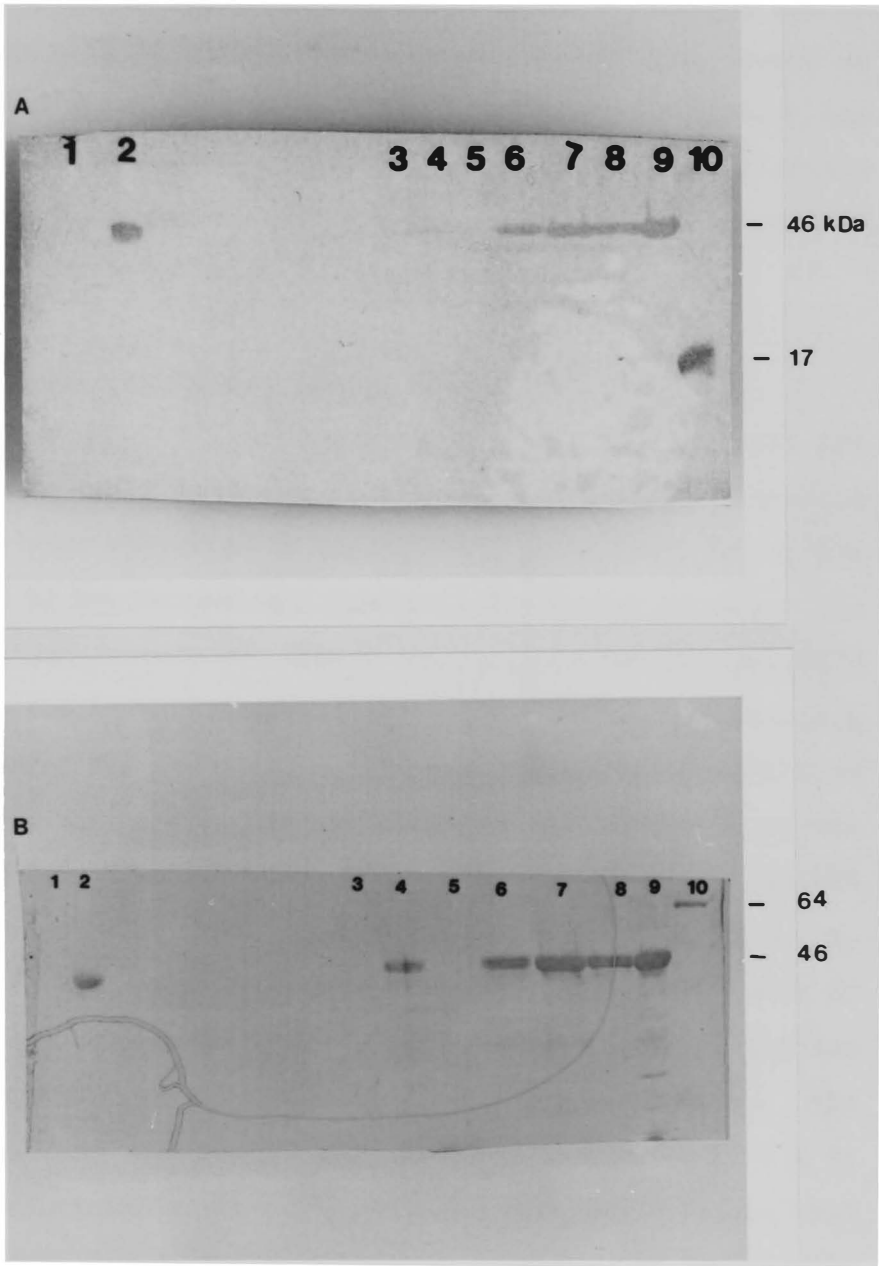
WESTERN BLOT ANALYSIS OF FUSION PROTEINS

Western blots analyses were performed on both the MBP- and core-fusion proteins using anti-WGA antisera to identify the presence of the recombinant domains and provide some evidence supporting a correctly folded domain.

MBP-Hevein Western Blot

The MBP-Hevein fusion protein was purified as described in Materials and Methods. A western blot was performed on the MBP-Hevein fusion protein using commercially available polyclonal anti-WGA antisera. Cell extracts from cultures containing the plasmid without the hevein cDNA insert, pMal-P, did not react with this antisera indicating that the anti-WGA antibodies were recognizing the hevein domain and not MBP (data not shown). Figure 17 depicts a western blot and corresponding SDS-PAGE gel in which some samples were never exposed to reducing agents (lanes 1-2), some samples were run under typical reducing SDS-PAGE conditions (lanes 5-10) and some samples were reduced and alkylated prior to electrophoresis (lanes 3-4). The antisera recognized the non-reduced and the reduced samples equally well (lanes 2 & 6), but the reduced and alkylated minimally (lane 4). Native hevein (nHev) treated under the same conditions (lanes 1, 3 & 5) is also shown on this blot and gel. The nHev did not stain with Coommassie or cross-react with the anti-WGA antisera. We

Figure 17. MN Western Blot. A is a western blot probed with anti-WGA. B is the corresponding SDS-PAGE gel. Lanes 1, 3 & 5 are native hevein (≈4800 Da). Lanes 2, 4 & 6 - 9 are MBP-hevein fusion protein (≈46,000 Da). Lane 10 is BSA (≈64,000 Da) and WGA (≈17,000 Da) as molecular weight markers. Lanes 1 & 2 are non-reduced. Lanes 3 & 4 are reduced and alkylated. Lanes 5 - 9 are reduced.

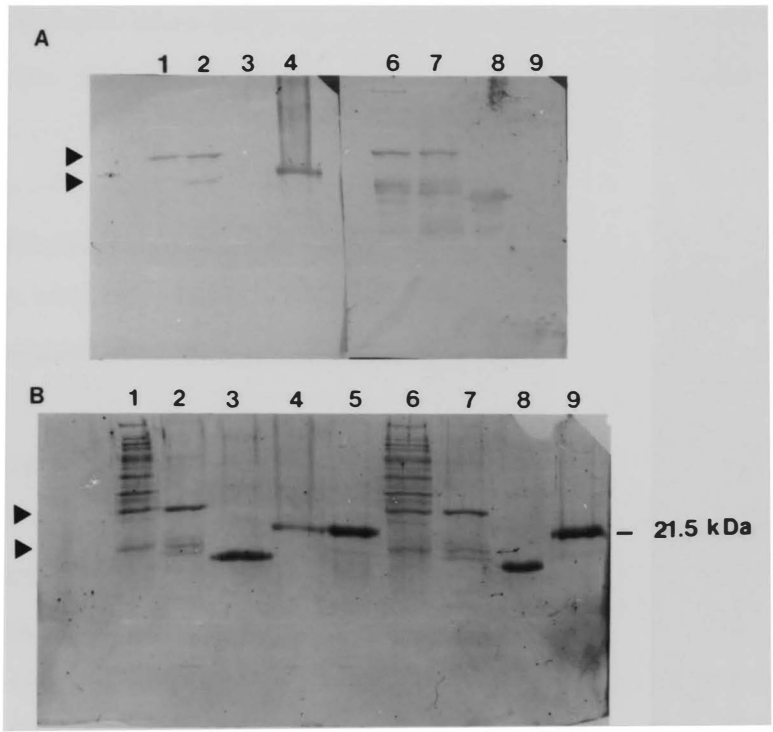

A
1 2 3 4 5 6 7 8 9 10
46 kDa
17
B
1 2 3 4 5 6 7 8 9 10
64
46

suspect that the lack of antibody cross-reactivity is due to the small size (4800 Da) of hevein and the minimal amount of material loaded on the gel. The inability to detect the single domain protein by SDS-PAGE or western blot analysis indicated the single recombinant domain proteins were going to be difficult to identify during purification.

Core Antigen-Recombinant Domain Western Blot Analysis

Fusion proteins produced by core-Hevein, CBS and CBA clones that were positive by plasmid restriction digestion pattern, were subjected to western blot analysis to verify the presence of the recombinant domain in the fusion protein. The core antigen-recombinant domain fusion proteins were purified as described in Materials and Methods. Core-BA cell extracts were assayed for recognition with both anti-WGA and anti-core antisera (Figure 18). Bands in identical locations at the expected molecular weight size for the fusion protein (23,000 Da) reacted with both antibodies (lanes 1, 2, 6 & 7), indicating that the fusion protein was produced. Lack of recognition of core by anti-WGA and WGA by anti-core implied that there was no cross-reactivity of the antibodies. The lower bands in the fusion protein lanes (lanes 1, 2, 6 & 7) react predominately with anti-core antibodies indicating that these lower bands are degradation products of the fusion protein in which the recombinant domain is one of the first

Figure 18. CBA Western Blot. A shows a western blot and B is the corresponding SDS-PAGE gel. Lanes 1 and 6 are CBA whole cell extracts. Lanes 2 and 7 are the CBA C14B pooled fractions. Lanes 3 and 8 are core. Lanes 4 and 9 are WGA. Lane 5 is a molecular weight marker (21,500 Da). Lanes 1 - 4 were probed with anti-WGA. Lanes 5 - 9 were probed with anti-core. The upper arrow indicates the fusion protein (≈23,000 Da). The lower arrow indicates core (≈17,000 Da).


A
1 2 3 4 6 7 8 9
B
1 2 3 4 5 6 7 8 9
21.5 kDa

fragments lost. Similarly to hevein, the liberated recombinant domain (BA) is not detected with either anti-WGA antibodies or Coommassie stain (lanes 1 & 2). Core-hevein and Core-BS western blots were also probed with anti-WGA antibodies (data not shown). Both reacted with a 23,000 molecular weight size band as expected for the fusion protein (≈17,000 Da for core plus ≈5000 Da for the recombinant domain).

RECOMBINANT DOMAIN PURIFICATION

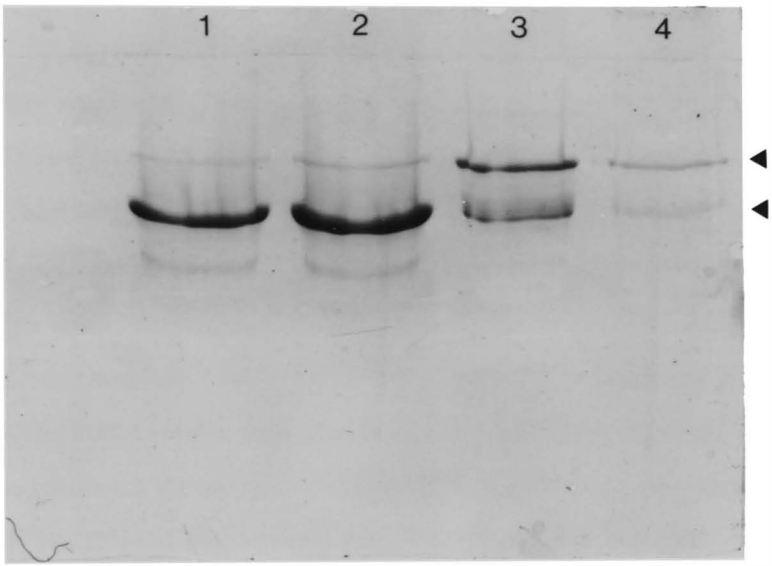
This section describes the separation, isolation and identification of the recombinant domains.

Trypsin Digestion

Since FXa cleavage did not successfully liberate recombinant hevein and the FXa site (I-E-G-R-X, X any amino acid except P) contains an Arg, trypsin was used to cleave the core-hevein fusion protein. The FXa recognition sequence was maintained in the core-hevein fusion protein construction because of the presence of restriction enzyme cloning sites. Only the Arg was present in the CBA and CBS constructs, since they were derived from completely synthetic DNA. The extent of cleavage was monitored by a shift from fusion protein molecular weight size (23,000 Da) to core molecular weight size (17,000 Da) by SDS-PAGE. In developing a protocol for

efficient trypsin digestion several difficulties were experienced. In 10mM tris, 150mM NaCl, 2mM $CaCl_2$, pH 8.0 buffer, trypsin digestion of the core antigen-recombinant domain fusion proteins was found to be incomplete. The liberated recombinant domain proteins were isolated in soluble form in the supernatant after core and the remaining fusion protein were precipitated at 45% $(NH_4)_2SO_4$ saturation. In an effort to improve the yield, an additional trypsin digestion was performed in the presence of 2M guanidine-HCl. Some difficulties were experienced when the fusion protein precipitated in the presence of 2M guanidine-HCl. Subsequently it was determined that precipitation could be avoided, if the protein concentration was kept below ≈2.5 mg/ml during trypsin digestion in 2M guanidine-HCl. SDS-PAGE indicated a nearly complete shift of the fusion protein band to the core band. Figure 19 shows the digestion pattern for CBS and CBA with the molecular weight size band shifts from partially digested fusion protein in physiological buffer (lanes 3 and 4) to nearly total cleavage in 2M Guanidine-HCl (lanes 1 and 2). The recombinant domain was recovered from this trypsin digest by precipitation of core antigen as a result of lowering the pH to ≈4.5 with 10% acetic acid. The supernatant containing the recombinant domain protein was desalted by dialysis. The dialysate was lyophilized or concentrated in a speed vac evaporator and yielded ≈20mg of

Figure 19. CBS & CBA Trypsin Digest Gel. Lanes 1 & 2 are the CBS and CBA trypsin digests in 2M guanidine. Lanes 3 & 4 are the CBS and CBA trypsin digests in physiological buffer. The upper arrow indicates the fusion protein (≈23,000 Da). The lower arrow indicates core (≈17,000 Da).


1
2
3
4

rhev and ≈20-50mg of the domain B proteins.

Further characterization (described below) revealed three major products of the trypsin digestion, core, an arginine rich peptide from the C-terminus of core and the recombinant domain protein. The core protein is extremely stable and not significantly proteolytically degraded, which facilitates isolation of the recombinant domain proteins. The BS and BA proteins have only one other potential trypsin cleavage site 3 residues from the C-terminus. Loss of this short piece would not be expected to affect the domain's structure and function. Rhev has 3 potential trypsin sites at positions 6, 10 and 42. N-terminal sequencing (described below) indicated a single cleavage site at the Arg in the FXa site. Thus, it is unlikely that a significant amount of trypsin cleavage occurred at either of the potential trypsin cleavage sites, Arg6 or Lys10. Although the possibility of cleavage at Lys42 one residue removed from the C-terminus was not investigated, cleavage at this residue would not be expected to affect the hevein structure or function. This result is consistent with the finding that WGA was found to be resistant to enzymatic proteolysis in the native state (Wright, et al., 1984).

Isolation and Identification of the Recombinant Domain Proteins

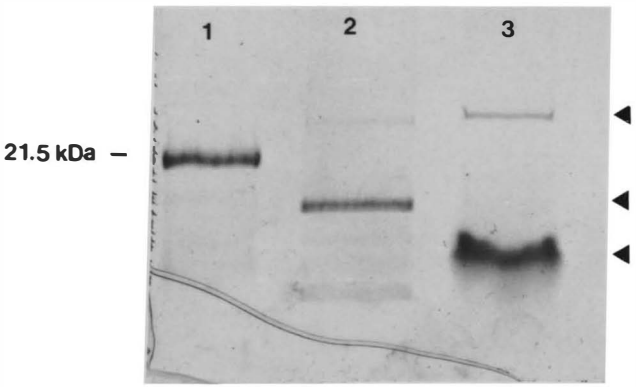
To assess the presence of remaining core antigen in the

supernatant after trypsin digestion, aliquots of rhev, BS and BA samples were analyzed by SDS-PAGE. Direct evidence for the presence of the recombinant domains at various stages of purification was obtained from amino acid compositional analyses, which revealed high Cys contents.

In the case of rhev, undigested core fusion protein was present. Therefore it was further purified by P30 gel filtration column. Figure 20 depicts an 18% SDS-PAGE of the precipitate (lane 2) and supernatant (lane 3) of the trypsin digested material precipitated with acid. The P30 column was equilibrated with 50mM sodium phosphate buffer, pH 6.8, and eluted fractions were monitored by absorbance at 280nm (Figure 21). Fractions which corresponded to elution profile peaks were pooled and analyzed by SDS-PAGE. The third and major peak showed only a low molecular weight band on SDS-PAGE. Amino acid compositional analysis of the pooled third P30 peak yielded a high Cys content indicating that this fraction probably contained the chitin binding domain protein. A high Arg content indicated the presence of other digestion products. Those fractions were further resolved on a reverse phase HPLC C-18 column (Figure 22A).

Fractions which corresponded to absorbance peaks at 220nm were eluted from the C-18 column, dried and their amino acid composition determined. The composition of the peak which eluted at 22 min retention time indicated a Cys rich peptide

Figure 20. Products of the Acid Precipitation of Trypsin Digest. Lane 1 is 21,500 Da molecular weight marker. Lane 2 is precipitated material from the acid precipitation of the trypsin digest. Lane 3 is material from the supernatant from the acid precipitation of the trypsin digest. The top arrow indicates fusion protein (≈23,000 Da). The middle arrow indicates core (≈17,000 Da). The lower arrow indicates the low molecular weight material, which was Cys rich and Arg rich.


1
2
3
21.5 kDa

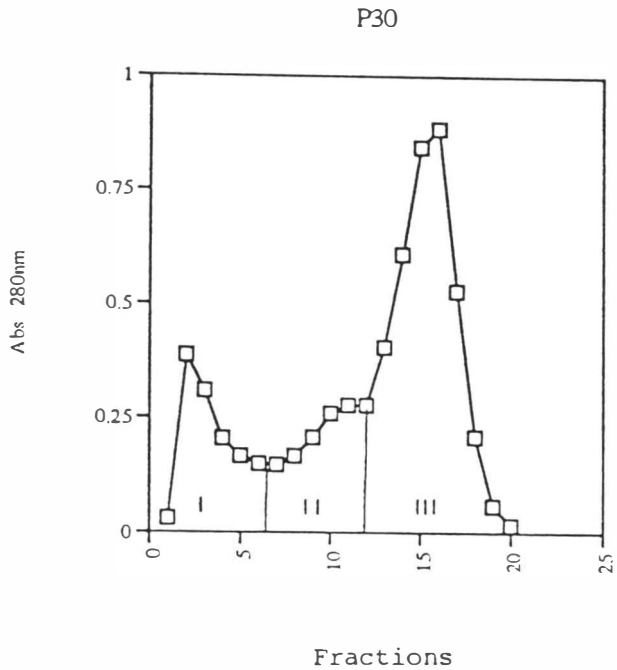
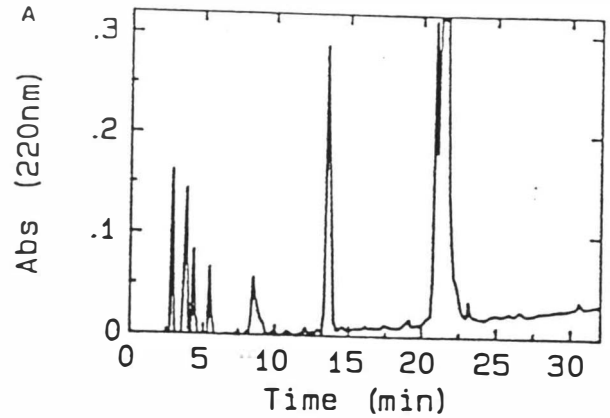
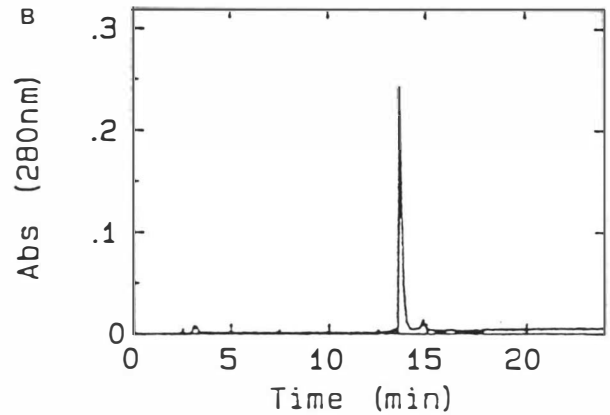
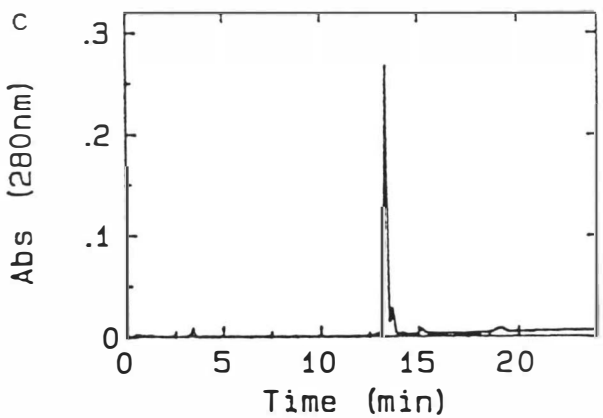


Figure 21. P30 Elution Profile. Elution profile of the supernatant from the acid precipitation of the trypsin digest applied to the P30 column.

Figure 22. C-18 Elution Profiles. A is the elution profile of rHev using a gradient of 0 - 100 % CH_3CN and monitoring elution at 220 nm. B is the elution profile of BA. C is the elution profile of BS. The gradient for BA and BS was 10 - 50 % CH_3CN and was monitored at 280 nm.


A
Abs (220nm)
.3
.2
.1
0
0 5 10 15 20 25 30
Time (min)


B
Abs (280nm)
.3
.2
.1
0
0 5 10 15 20
Time (min)


C
Abs (280nm)
.3
.2
.1
0
0 5 10 15 20
Time (min)

similar to that of hevein. A smaller Arg rich peak eluted at 13 min. The composition of this peak was consistent with the Arg rich C-terminal region of core.

In the case of BS and BA, SDS-PAGE analysis showed that there was no evidence of contamination by core antigen. These samples were directly applied to the reverse phase HPLC column to further assess their purity (Figure 22B and C). The fractions corresponding to the major peaks were submitted for amino acid compositional analysis. An amino acid compositional analysis representative of this peak in each C-18 profile is shown in Table 2. They correlate well with the expected Cys-rich compositions for the recombinant domains. However, the amino acid residues which oxidize easily, such as cysteine, showed a low yield. The glycine content was lower than expected in this particular amino acid composition for BS and BA, but it gave the expected value in other compositional analyses.

N-terminal sequence analysis was performed to determine the purity of the rhev and BS samples (Table 3). Both rhev, purified by C-18, and BS, purified by gel filtration chromatography, yielded the N-terminal sequence G-E-Q. This was the expected sequence, if trypsin cleaved at the desired Arg.

Table 2. Amino Acid Compositional Analysis

	rHEV		BS		BA	
	obs	exp	obs	exp	obs	exp
Asx	7.3(7)	7	3.5(4)	3	3.6(4)	3
Glx	4.6(5)	6	6.8(7)	7	6.9(7)	7
CMC	6.3(6)	8	6.5(7)	8	6.8(7)	8
Ser	3.6(4)	4	3.1(3)	3	2.4(2)	2
Gly	6.0(6)	6	5.7(6)[a]	11	5.7(6)	11[a]
His	0.8(1)	1	-	0	-	0
Thr	1.0(1)	1	1.8(2)	2	2.0(2)	2
Ala	-[b]	1	4.0(4)	4	4.9(5)	5
Arg	0.5(1)	1	1.0(1)	1	1.0(1)	1
Tyr	0.5(1)	1	1.8(2)	3	2.1(2)	3
Val	0.2(0)	0	-	0	-	0
Met	-	0	-	0	-	0
Trp	not done	3	-	0	-	0
Phe	-	0	0.9(1)	1	1.0(1)	1
Ile	0.2(0)	0	1.3(1)[c]	0	1.4(1)[c]	0
Leu	2.2(2)	2	-	0	-	0
Lys	1.8(2)	2	-	0	-	0
Pro			1.1(1)	1	1.2(1)	1

a Gly was unexpectedly low in these samples; probably the standards were abberrantly high.

b The Ala peak in this sample was not resolved from the Gly peak.

c The value for BS and BA represents an abnormal peak which is probably an artifact and not an amino acid.

Table 3. N-Terminal Sequencing

rHev	G-E-Q
BS	G-E-Q

Free Sulfhydryl Determination

In WGA and nhevein, all the Cys residues exist in disulfide bonds (Wright, 1977; Rodriguez-Romero, et al., 1991; Andersen, et al., 1993). Thus, the free cysteine content was determined for the recombinant domains to provide evidence supporting a correctly folded structure. For rhev, no free sulfhydryls were detected using 5,5'-dithio-bis(2-nitrobenzoic acid). The number of free sulfhydryls in the BS and BA domains was determined by alkylation of the protein in the presence of denaturant. As apparent from Table 4, no carboxymethyl-Cys was detected by amino acid compositional analysis. Thus, all the cysteines present in our recombinant domains are in disulfide bonds.

MOLECULAR WEIGHT DETERMINATION

This section describes the experiments performed to determine the oligomeric state of the domain B proteins (BS and BA). Conclusions drawn will help to assess whether the domain B binding site is generated.

HPLC-Gel Filtration Chromatography

The recombinant domain B proteins were analyzed by HPLC gel filtration chromatography to estimate their molecular weights. The predicted molecular weights of BS and BA based on their cDNA sequences, are 4375 and 4358, respectively. A

Table 4. Free Sulfhydryl Determination of the Recombinant Domains

	Moles of Cys/Mole of Protein		
	acid hydrolysis	alkylation acid hydrolysis	reduction & alkylation acid hydrolysis
BA	0	0	3.2(3)[a]
BS	0	0	6.8(7)

a low values were the result of anomalous distribution of the carboxymethyl cysteine peaks

molecular weight of ≈4,800 was estimated for BS from the retention time by graphical analysis (See Figure 23). This result indicates that BS is a monomer under the experimental conditions employed. We presume that the BA domain would also migrate ≈5,000 Da as the two proteins only differ by one amino acid. In experiments performed at lower ionic strength (50mM sodium phosphate, pH 6.8) the molecular weight estimate for BS was ≈10,000 (data not shown). This large discrepancy suggests that there is an appreciable error associated with this type of determination for low molecular weight proteins. Since the retention times appear to be critically dependent on the ionic strength of the buffer used, it is important to test a range of buffer conditions and select molecular weight standards in a suitable range in order to deduce reliable molecular weight estimates by this method.

Mass Spectral Analysis

Mass spectral analyses were also performed to assess the oligomeric state of the domains as a more sensitive technique. Samples of BA submitted for mass spectral analysis contained between 10 and 20 pmoles of BA in 0.5μl and were prepared at pH 6.8, pH < 2 and in the presence of $(GlcNAc)_3$ (Table 5A). The results indicate that under all conditions the primary species present was the monomer (≈4500 Da). Small amounts at the dimeric, trimeric and tetrameric masses were also

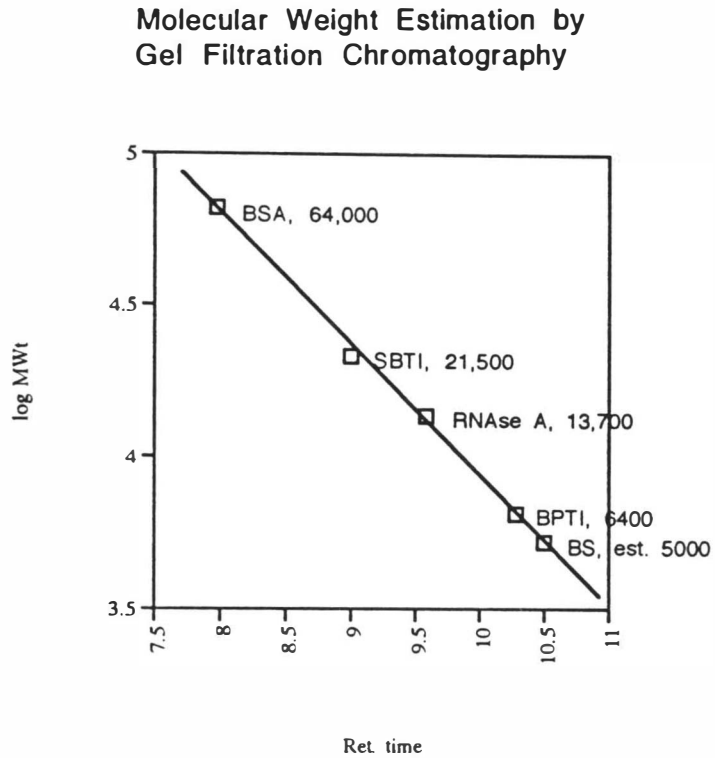


Figure 23. Gel Filtration Chromatography. The molecular weight standards used are BSA, SBTI, RNAse A and BPTI with their estimated molecular weights indicated.

Table 5. Mass Spectral Analysis of BA and BS

A.

Relative Mass Amounts of BA (%)[a]

Mass	pH 6.8	pH<2	+ligand
4330	85	80	95
4625	95	92	94
8692			17
8965	18	23	26
9245	17	26	20
13,425		17	17
13,980		16	17
18,770		13	
22,370		12	

B.

Relative Mass Amounts of BS[b]

Mass	
4153	100
4342	73

a. MALDI-TOF experiment performed by the Dana Farber Cancer Institute, Boston, MA

b. MALDI-TOF experiment performed by Commonwealth Biotechnologies, Inc., Richmond, VA

observed. Two peaks were observed in approximately equal ratios in all spectra, because this particular BA sample contained two species differing by 3 amino acids on the N-terminus (G-S-R). Subsequently, a sample of BS with a single N-terminus was submitted for mass spectral analysis (Table 5B). This sample gave two peaks which correspond to masses of the full length BS sequence and the BS sequence truncated by a dipeptide lost from the C-terminus (A-D).

SACCHARIDE BINDING STUDIES

Saccharide binding studies were performed to assess the functional characteristics of the recombinant chitin binding domains. Saccharide binding capabilities would infer that the domains are correctly folded.

Association With Chitin

A crude chitin binding assay was performed to screen the saccharide binding function of the recombinant domains. Chitin is an insoluble GlcNAc polymer and the only known natural ligand of hevein. This assay takes advantage of the insoluble nature of chitin. The presence of unbound hevein, was monitored by the intrinsic tryptophan fluorescence of its three tryptophans. A fluorescence scan of the supernatant, after precipitation of the chitin, revealed the relative amounts of unbound protein (Figure 24). WGA, nHev and rHev

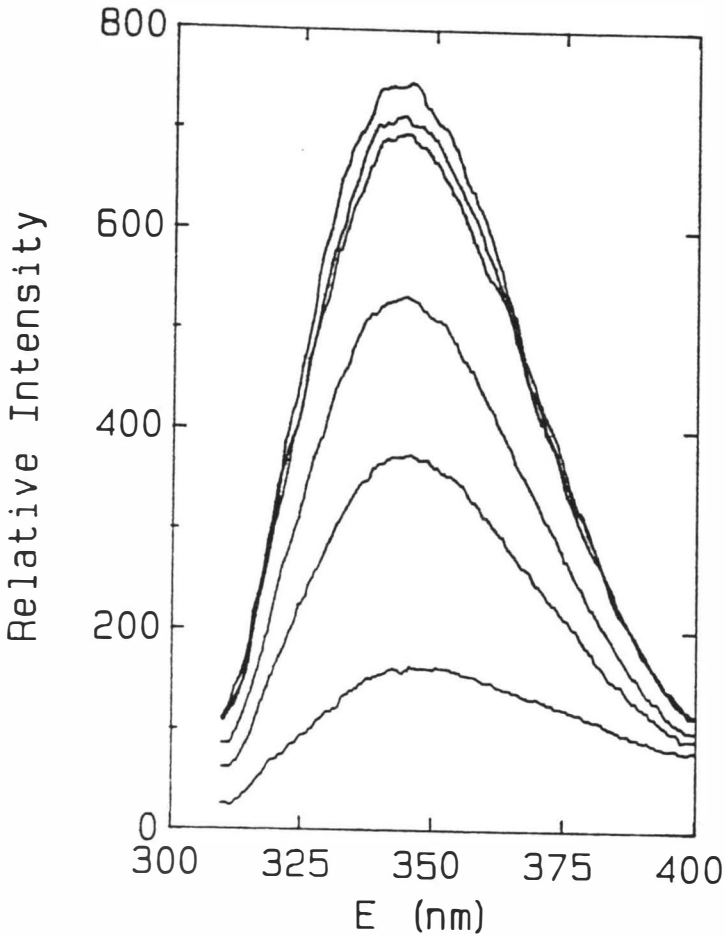


Figure 24. Intrinsic Tryptophan Fluorescence after Incubation with Increasing Amounts of Chitin. Fluorescence emission scan of nhev in the supernatant after incubation with increasing amounts of chitin for 1 hr, then precipitating the chitin. Excitation λ was 290nm. Emission λ was scanned from 310-400nm. The top curve was incubated with no chitin. Each successive decreasing curve was incubated with 5, 10, 25, 50 or 100mg chitin, respectively.

showed a decreasing trend in fluorescence intensity upon incubation with increasing amounts of chitin indicating binding (Figure 25). To test the possibility of nonspecific protein-chitin association, a parallel study with the core antigen indicated a complete lack of binding.

The amounts of unbound recombinant domain B protein were assessed using the Pierce BCA protein determination assay (Table 6). Although the changes were small compared with those seen in the fluorescent studies for rhev, a reproducible decrease in protein concentration as a function of increasing chitin was observed for both BS and BA indicating specific binding.

$(GlcNAc)_n$ Binding Assessed by Intrinsic Tryptophan Fluorescence

The two tryptophans in the hevein binding site proved useful to study the binding of $(GlcNAc)_n$ (n=1-4) oligomers by fluorometry. For this assay, a shift in the λ_{max} upon addition of saccharide is indicative of a change in polarity of the environment of the tryptophan residues (Privat et al., 1974a). With rHev, nHev and WGA a blue shift was observed in λ_{max} with addition of $(GlcNAc)_{3,4}$ (Table 7). A similar shift had previously been reported for WGA (Privat, et al., 1974a; Lotan and Sharon, 1973). Since the aromatic residues are involved in van der Waals contacts with the saccharide as seen in the

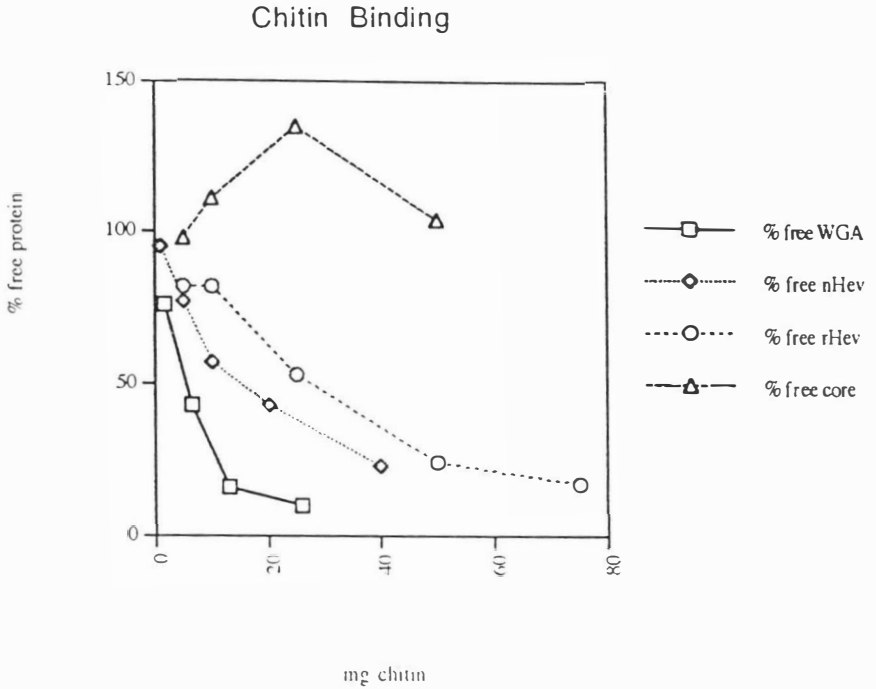


Figure 25. Percent Relative Intrinsic Fluorescence. The intrinsic tryptophan fluorescence intensity of unbound WGA, nHev, rHev and core (in the supernatant from the chitin binding assay), with increasing chitin relative to the intensity with no chitin.

Table 6. Chitin Association of the Domain B Proteins

prot	chitin (mg)	prot(mg)
BS	0	0.405
	50	0.334
BA	0	0.516
	50	0.452

Protein concentration in the supernatant after incubation with and without chitin.

Table 7. Fluorescence Emission λ_{max} Shift upon Saccharide Binding.

prot	saccharide	Emiss. λ_{max} nm
WGA	-	349
	$(GlcNAc)_3$	342
	$(GlcNAc)_4$	342
nhev	-	347
	$(GlcNAc)_3$	342
	$(GlcNAc)_4$	342
rhev	-	347
	$(GlcNAc)_3$	342
	$(GlcNAc)_4$	342

Fluorescence λ_{max} of WGA nHev and rHev with no saccharide, $(GlcNAc)_3$ and $(GlcNAc)_4$.

crystal structure of WGA, it was expected that the same would be true for hevein. This is again supportive evidence that the protein binds saccharide and is therefore probably folded correctly.

Saccharide Binding Assessed by Isothermal Titration Calorimetry

The binding of GlcNAc oligomers and NAc-Neuramin lactose (NeuLac) to BS, BA, nhev and WGA was studied using isothermal titration calorimetry (ITC). Figure 26 shows typical binding isotherms obtained with the BS and BA proteins titrated with $(GlcNAc)_3$. The purity of the BS and BA samples was confirmed by reverse phase HPLC and amino acid compositional analysis. The BS sample was subjected to N-terminal sequencing to verify the location of trypsin cleavage. The BS and BA proteins that were titrated with $(GlcNAc)_4$ had 2 N-termini differing by 3 residues. Native hevein was assayed for $(GlcNAc)_4$ binding only, due to limited amounts of protein.

Table 8 shows binding constants and other thermodynamic parameters for these binding events. These data indicate that the recombinant domain B proteins bind the WGA specific saccharide. However, the binding is ≈7 times weaker (K_d ≈4.0 x 10^{-4}M) compared with that in the four domain lectin WGA (K_d ≈0.54 x 10^{-4}M), and ≈20 times weaker compared with native hevein (K_d ≈0.19 x 10^{-4}M). Binding of $(GlcNAc)_2$ and NeuLac to

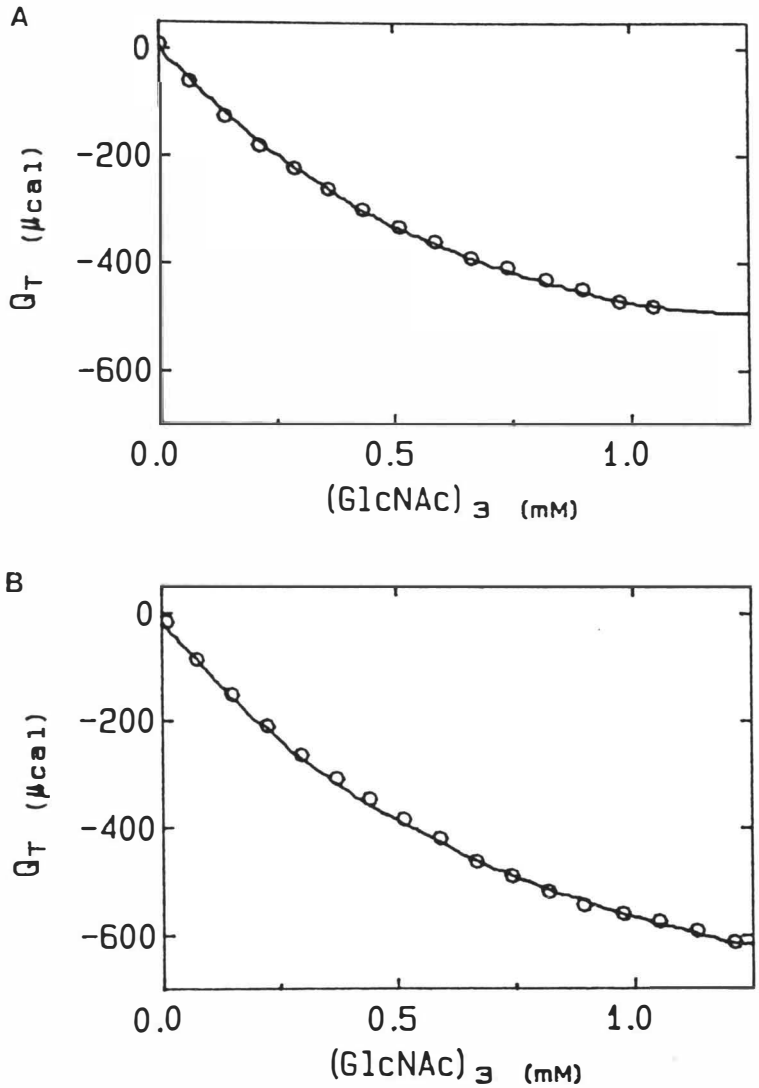


Figure 26. ITC Binding Isotherms. A is the binding isotherm obtained for BA titrated with $(GlcNAc)_3$ at 25°C and 0.44mM protein concentration. B is the binding isotherm obtained for BS titrated with $(GlcNAc)_3$ at 25°C and 0.36mM protein concentration.

Table 8. BS, BA, nHev and WGA ITC Results

	K_d	ΔH	TΔS(°C)	ΔG	n	C	SD
	x10^4(M)	kcal/mol	kcal/mol	kcal/mol			
BA							
$(GlcNAc)_3$	2.98	-1.11	3.60(25)	-4.7	1	1.5	3.39
	3.27	-1.41	3.17(30)	-4.7	1	0.7	3.09
	2.42	-0.87	4.05(30)	-4.9	1	1.8	3.58
	3.52	-0.83	3.96(35)	-4.8	1	1.3	3.33
$(GlcNAc)_4$	4.76	-1.50	2.94(30)	-4.4	1	1.1	
$(GlcNAc)_2$	No binding detected						
NeuLac	No binding detected						
BS							
$(GlcNAc)_3$	5.25	-1.87	2.43(25)	-4.3	1	0.7	3.08
	5.12	-1.63	2.75(30)	-4.4	1	0.7	2.41
	2.14	-1.62	3.29(30)	-4.9	1	1.2	2.86
	4.72	-1.31	3.22(35)	-4.5	1	0.8	2.32
$(GlcNAc)_4$	3.45	-3.90	0.48(30)	-4.4	0.5	2.0	
$(GlcNAc)_2$	No binding detected						
NeuLac	No binding detected						
nHev							
$(GlcNAc)_4$	0.19	-3.1	3.09(30)	-6.19	0.6	4.6	0.93
WGA							
$(GlcNAc)_3$	0.54	-8.23	-3.21(30)	-5.02	2.2	1.4	11.86
α2-3 Neulac	3.88	-0.39	4.29(30)	-4.68	2.5	0.5	3.93

Values were determined using ITC. 'n' values are per monomer. For BS and BA, n was held constant during analysis. For WGA and nHev all variables were allowed to float during analysis. C is defined as K_a x molarity of the protein solution. SD refers to the standard deviation of fit of the theoretical binding isotherm to the actual data points.

BS and BA was not detectable under the conditions used in these ITC experiments. Table 9 compares observed binding constants for WGA with those of other investigators.

Table 7. Comparison of WGA Dissociation Constants

	$K_d \times 10^4$ (M)				
	Rice ITC	Bains ITC	Privat[a] W fluor	N&B eq.dial.	K&C nmr
GlcNAc		25.0	14.5	7.6	
(GlcNAc)2		1.9	2.2	0.5	
(GlcNAc)3	0.54	0.90	0.50	0.12	
(GlcNAc)4		0.81	0.435		
(GlcNAc)5		0.53			
α2-3 Neulac	3.88				8.7
α2-6 Neulac					53
n/monomer	2.2	2	1		2

Comparison of K_d values obtained using different methods. Rice is data presented here. Bains is data from Bains, et al, 1992. Privat is data from Privat, et al, 1974a. N&B is data from Nagata and Burger, 1974. K&C is data from Kronis and Carver, 1982. ITC is Isothermal Titration Calorimetry, W fluor is intrinsic tryptophan fluorescence, eq.dial. is equilibrium dialysis, nmr is nuclear magnetic resonance.
a One binding site per monomer was assumed to derive these values.

DISCUSSION

RESTATEMENT OF OBJECTIVES

The goal of the project was to develop a protocol to express, purify and characterize domain B of WGA, and ultimately to study the WGA domain B binding site. This protocol will eventually be applied to the other domains and their binding sites. A WGA binding site consists of contributions from two opposing domains. The saccharide lies in an aromatic pocket on one domain and is stabilized by polar residues on the opposing domain. Since each domain contains an aromatic pocket and most of the domains contain the polar region, two potential saccharide binding sites exist between a pair of opposing domains (see Figure 3B). A complete WGA binding site would be generated in a dimer of single domains, if they associated as opposing domains do in the WGA dimer.

Domain B of WGA was cloned and expressed in two forms varying at position 28. One has the native sequence with Ala28 (BA) and the other has the mutated sequence with Ser28 (BS). In WGA, the two saccharides GlcNAc and NeuNAc bind identically in the B-site except that an additional H-bond exists between bound NeuNAc and the Ser28 -OH of domain C. A putative dimer of domain B with a Ser at position 28 would

allow this same H-bond to be formed and thus would faithfully reproduce the WGA B-site. Characterization of the recombinant domains includes evidence supporting a correctly folded structure, oligomeric state determination and saccharide binding ability.

CLONING, EXPRESSION AND PURIFICATION OF THE RECOMBINANT DOMAIN PROTEINS

Cloning of the Domain B cDNA Sequences

The BS cDNA was generated by completely synthesizing the desired sequence as this was the most expedient approach for making the exact sequence desired. Subsequently PCR mediated site directed mutagenesis was used to generate the point mutation to make the BA clone (Vallette et al., 1989; Hemsley et al., 1989). In designing the primers for the Ser→Ala mutation, we intentionally altered the Nar I restriction enzyme site. Changing this site allowed positive clones to be identified by the restriction enzyme digestion pattern, alleviating the need to screen colonies by DNA sequencing. An alternate, less efficient approach to generate the BS sequence by PCR mediated site directed mutagenesis, would have required the incorporation of many changes in the hevein cDNA that we had available (a WGA cDNA was not available). Using PCR mediated site directed mutagenesis techniques, would have required at least three independent changes in the DNA

sequence to modify the residues involved in saccharide binding, and 12 or more to completely change from the hevein sequence to the domain B sequence. Each change would be extremely labor intensive, requiring screening colonies for the desired clone and verifying the clone by DNA sequencing.

Fusion Protein Expression

By following the protocol described, three different chitin binding domain proteins were successfully expressed in E.coli as fusion proteins and isolated in yields sufficient for characterization. The primary reason for choosing a fusion protein system was to facilitate purification of the recombinant domains. Since the fusion proteins could be easily purified in large quantities, we only had to develop a system to separate the fusion partners and isolate the recombinant domain.

Expression of the recombinant domains without a fusion partner would have required developing a purification protocol. However, it was uncertain if the single domain proteins could be tracked by their saccharide binding ability or by antibody recognition. The latter could be ruled out as we demonstrated that anti-WGA antibodies did not detect the single domain proteins by Western blot analyses (Figures 17 and 18). In addition, hevein antibodies were not commercially available and we could not generate our own antibodies due to

a lack of sufficient amounts of native hevein. It was also not possible to express the domain without an N-terminal methionine. The presence of this extra residue may conceivably have affected the folding or function of the domain. Conversely, expression of the recombinant domains without fusion partners would have provided some information as to the ability of the domains to spontaneously fold correctly. However, whether the disulfide rich WGA would fold spontaneously or require assistance in folding by chaperonins or disulfide exchange proteins was unknown.

Both types of fusion proteins were easily purified in large quantities. The moderate levels of protein expression in the core-fusion protein expression system was advantageous, because many recombinant proteins expressed at high levels become sequestered in inclusion bodies (Tsuji, et al., 1987; Schroeder and Raikhel, 1992). The protein in the inclusion bodies generally has to be solubilized under denaturing conditions. Removal of the denaturant allows the protein to refold. It was deemed better to avoid inclusion bodies, since there is no assurance of obtaining a functional protein.

Fusion Protein Cleavage

The major obstacle encountered with isolating the recombinant domains was proteolytic cleavage between the fusion partners. Trypsin was used to carry out the cleavage

reaction in the HBV core antigen-fusion protein system because Factor Xa proved ineffective. This may be due to the fact that a disulfide bridge exists three residues removed from the cleavage site. The proximity of this disulfide bond may restrict the flexibility of this peptide to attain the conformation necessary for recognition by the protease.

In addition, problems were experienced with determination of the optimal conditions for trypsin digestion. In Tris buffered saline, pH 8.0, trypsin released only a small amount of recombinant hevein, and the BS and BA domain proteins from their respective HBV core antigen-fusion partners. However, in 2M guanidine-HCl trypsin digestion yielded almost complete separation of the recombinant domain. This indicates that some of the recombinant domain proteins may be on the inside of the core particle and, therefore, are inaccessible to trypsin digestion until the core particle is dissociated. It is known that the core antigen produced recombinantly forms a native-type viral core particle (Zheng, et al., 1992). The individual core antigen monomers in the construct which was used are disulfide cross-linked into dimers. In 2M guanidine-HCl the core particle dissociates into disulfide bound core dimers (Dr. Darrell L. Peterson, personal communication), which was verified in our lab using gel filtration chromatography (data not shown). Another problem experienced with the trypsin digestion was precipitation of the core-

fusion proteins in 2M guanidine-HCl. However, if the concentration of the fusion protein was kept below 2.5 mg/ml, this problem was alleviated. Yields of recombinant protein obtained were dependent on the efficiency of the proteolytic cleavage of the fusion protein. Optimal yields obtained were between 10-50mg/ 20 liter culture.

TERTIARY STRUCTURAL CHARACTERIZATION

One would not predict that a highly disulfide rich protein, known to lack other stabilizing secondary structural elements, would fold correctly when produced in a bacterial expression system (reducing environment). However, Schroeder and Raikhel (1992) have recently succeeded in purifying active recombinant barley lectin (homologous to WGA) from E. coli without using a fusion protein system. This is remarkable, because the four domains of the polypeptide chain have to fold independently and form only intra-domain disulfide bridges. In this study, recombinant barley lectin was solubilized from inclusion bodies by incubation in a redox buffer, which promotes isomerization of incorrectly folded proteins. Active protein was isolated by GlcNAc affinity chromatography. Thus it was demonstrated for this four domain protein that significant amounts of correctly folded protein could be prepared when allowed to reoxidize. As additional structural evidence, this recombinant material could be crystallized

under identical crystallization conditions and with the same crystal packing as found for WGA (Wright et al., 1993). This encouraged us in our belief that correctly folded single domains could be successfully cloned and expressed.

While we have no unequivocal evidence which allows us to conclude that correctly folded recombinant domains were produced, the observed data are highly suggestive of a hevein-like tertiary structure. The most convincing evidence is that the domains can bind oligosaccharides. This is based on several different types of experiments. Other experimental evidence addressing the conformational aspects of the recombinant domains, also supports a correct tertiary structure.

Role of the Disulfide Bridges

The experimental results presented above indicated that all the Cys residues are in disulfide linkages in the three recombinant domains produced (rhev, BA and BS). The presence of free Cys residues would have precluded a "native WGA-like domain" structure.

The four disulfide bridges in each domain of WGA are crucial to maintaining the tertiary structure. This was concluded from the studies of Erni et al (1980), who showed that WGA loses activity upon reduction and alkylation in the absence of a denaturant. No intermediate states of partially

reduced protein were observed when less than molar ratios of reductant to Cys residues was used. This evidence implies that partially reduced stable species are not formed. Another interesting observation made by Erni et al (1980) was that some activity could be retained when the protein was reduced and alkylated in the presence of saccharide. However, there was no evidence to suggest which disulfides bridges remained intact and which were unprotected and thus susceptible to reduction.

To assess whether alternate disulfide pairings could exist and maintain the intact aromatic pocket, the structure of WGA domain B was examined on computer models. Model building showed that at most three alternate disulfide pairings could exist, which would not dramatically perturb the position of the residues involved in saccharide binding. As illustrated in Figure 27, the native disulfide pairings are between Cys3 and Cys18, Cys12 and Cys24, Cys17 and Cys31, and between Cys35 and Cys40. Inspection of the distances between the $S\gamma$ atoms of Cys3 to Cys24 and Cys12 to Cys18 indicated that they approach within 2Å, a normal distance for a disulfide bond. Cys17 and Cys35 could also form a disulfide bridge. However, this would leave Cys31 and Cys40 unpaired and most likely reduced (Figure 27). For the purpose of this discussion, a functional structure implies that the domains possess identical disulfide pairings as in WGA. It is

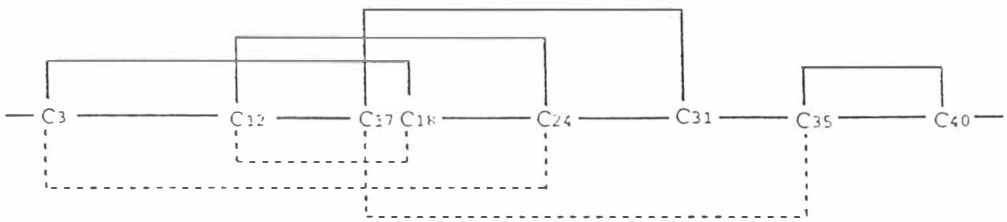


Figure 27. Alternative Disulfide Pairings. The solid lines above the sequence represent the native disulfide pairings. The dashed lines below the sequence represent the potential alternate disulfide pairings, which would not dramatically disturb the general fold of the Cα backbone.

possible, however, that other tertiary structures also may exhibit sugar binding activity.

Antibody Recognition

Western blot analysis probed with anti-WGA confirmed the presence of the recombinant domains as fusion protein partners. Antibody recognition of the MBP-hevein fusion protein implied that hevein had enough sequence or structural similarity to cross-react with anti-WGA antibodies.

Strong antibody recognition of the non-reduced and reduced hevein fusion protein samples indicated that rhev in the MBP-fusion protein could possibly be folded correctly as purified and possibly refold before transfer to the western blot membrane. Evidence which corroborates our hypothesis that disulfide rich domains can refold correctly when the reductant and SDS are diluted during Western blot analysis, comes from two other independent techniques. It has been observed that active disulfide rich proteins can be recovered after reduction and reoxidation (Yang, 1967; Schroeder and Raikhel, 1992). As another example, the Southwestern assay was employed to identify functional DNA binding proteins (Farrell, et al., 1990). The protein refolds when the gel is equilibrated in the transfer buffer before blotting to a membrane and is identified by binding to a radiolabelled DNA ligand.

Conversely, the antibodies could recognize sequence, which would be identical in both the reduced and non-reduced samples. We also have no direct evidence whether the polyclonal anti-WGA antibodies recognize sequence or structure. Their polyclonal nature suggests that they probably recognize both. According to Sigma, the antibodies are raised against non-reduced WGA. This would suggest that at least some of the antibodies recognize structure.

The lack of antibody recognition of the reduced and alkylated hevein fusion protein could be due to a loss of structural epitopes, which would suggest that structure is recognized. If only the sequence was recognized by the antibodies, the loss of recognition of the alkylated domain may be due to the additional bulky carboxy-methyl groups.

Unfortunately, the native hevein domain could not be detected by the anti-WGA antibodies on the Western blot. This may have several explanations. Due to its small compact fold, the domain may pass through the membrane upon transfer, or it may not adhere to the hydrophobic membrane because it is deficient in aliphatic residues. It is also possible that it eluted out of the gel during equilibration in the transfer buffer; or, perhaps, it reoxidized to a structure not recognizable by the antibody. Alternatively, there may not be sufficient amino acid sequence recognition by the anti-WGA antibodies, since hevein is only 65% identical to WGA in

sequence. Recognition of the core fusion proteins, but not of the degradation products by anti-WGA antibodies (Figure 18) argues for the view that the recombinant domain B proteins, like hevein, are not detectable by western blot analysis. This may simply be a characteristic of this unique disulfide rich protein.

Preliminary Structural Studies

Preliminary efforts to investigate the tertiary structure of BS and BA using NMR and crystallographic techniques, have shown that the recombinant B-domain proteins behave much like hevein. The 1-Dimensional NMR spectrum of the BA domain was found to compare well with a spectrum of native hevein in terms of chemical shifts and line widths (Dr. Hans C. Siebert, unpublished results). Initial crystallization attempts showed that both the recombinant BS and BA domain proteins crystallize under the same conditions as nhevein. Crystallization requires the presence of Ca^{++} ions.

In summary, every technique employed has provided strong evidence that the recombinant domains are structurally very similar to native hevein. Although we suspect that the recombinant domains are folded correctly, definitive evidence will have to await crystallographic or NMR structural analysis.

QUATERNARY STRUCTURE DETERMINATION

The oligomeric state of the recombinant domain B proteins was investigated in view of the fact that only dimers can recreate the complete WGA B-binding site. The results based on gel filtration chromatography and mass spectrometry techniques have suggested that single domains behave as monomers in solution.

Our initial assumption that there is a high probability that single B-domains could dimerize was based on structural evidence. In WGA the contacts between opposing domains B and C are stabilized by four H-bonds and many van der Waals interactions. Two of the H-bonds occur between the backbone carbonyl and amide groups of residue 26 on both domains (see Figure 5). The residues involved can only come sufficiently close to make these H-bonds if the carboxyl linked residue (27) is a glycine in both domains. The presence of a side chain at residue 27 would obstruct dimerization. This is consistent with the finding, that all proteins of the chitin-binding family thus far characterized as dimers, have a conserved Gly in position 27. On the other hand, in proteins known to function as monomers, residue 27 is not conserved (Wright et al., 1991). For instance, monomeric hevein has a threonine in position 27, which precludes formation of the same two inter-domain H-bonds (van Parijs, et al., 1991). Since the recombinant domain B sequences include a Gly in

position 27, the same interactions present in the B/C interface of WGA are possible. However, it is difficult to estimate whether these contacts are sufficient for dimerization. In WGA many other inter-subunit interactions are present that stabilize the dimer. These include ≈20 H-bonds, that occur not only between two opposing domains, but also between non-opposing domains and extensive van der Waals inter-subunit contacts (Wright, 1987b).

Several pieces of evidence have also suggested that dimerization is not essential for saccharide binding. Privat et al (1976) have shown that oxidation of a single Trp residue resulted in dissociation of the WGA dimer and approximately three fold weaker binding to the saccharide. The three fold weaker binding could be attributed to the loss of the polar contributions from the opposing domain. The fact that a number of monomeric chitin binding proteins bind GlcNAc (chito-oligosaccharides), provides additional evidence that a single aromatic pocket is sufficient for saccharide binding (Wright et al., 1991). This evidence has suggested that dimerization is not a requirement for saccharide binding, but it can increase the binding affinities as a result of additional stabilizing H-bond interactions.

Mass Spectral Analysis

Mass spectral analysis of the BA domain provided evidence

for the presence of primarily monomers (Table 5). Diminishing amounts of dimers, trimers and tetramers were observed in some spectra indicating that under some conditions aggregation may occur. The particular sample investigated was from a trypsin digest which had produced two N-termini (see above). However, there is no reason to believe that three additional residues at the N-terminus would affect the solubility, aggregation or dimerization potential of the domain. Since the mass spectrometry sample was dried on a screen, there is a possibility that aggregates formed in the drying process. Mass spectral analysis, carried out at a later date on a sample of BS which was homogeneous with respect to a single N-terminus, revealed only monomeric species (Table 5).

The weak tendency to aggregate may be explained considering the solubility limits of the single domain proteins. WGA is a relatively insoluble protein having solubility limits ≈2.5 mg/ml at neutral pH. Similar low solubility may apply to the recombinant domains. Although, the maximum solubility of the domains was not determined, it was estimated to be 5-10 mg/ml. Precipitates which formed upon storage were analyzed for their amino acid compositions. These were found to be identical to those of the soluble protein (data not shown).

NMR

Unpublished NMR data of the BA domain are consistent with our results. The 1-Dimensional NMR spectrum was very similar to the spectrum of native hevein and did not suggest the presence of aggregates (Dr. H. Siebert, personal communication).

In summary, the combined evidence presented here, which is based on gel filtration, mass spectrometry and NMR data, supports our conclusion that the recombinant domains are monomeric. The functional studies should therefore reflect only the saccharide interactions that occur in the aromatic pocket. It is thus likely, that the four H-bonds predicted to stabilize a dimeric recombinant domain structure are not sufficient to maintain a dimeric species in solution.

FUNCTIONAL CHARACTERIZATION

Saccharide binding of the recombinant domains was assessed by several independent techniques. On one hand we chose a simple chitin binding assay to qualitatively establish binding. In addition, fluorescence spectroscopy was used to monitor GlcNAc binding to WGA and hevein. On the other hand, ITC (a more sensitive technique) was used to quantitatively assess binding constants and thermodynamic parameters. The results have allowed us to conclude that the GlcNAc-oligosaccharides bind at the aromatic binding pocket.

However, binding was found to be weak as compared to WGA and hevein. Measurable binding is only observed for larger GlcNAc oligomers and not for the disaccharide or NeuLac. Independent evidence, suggesting GlcNAc binding at the aromatic site, comes from preliminary chemically induced dynamic polarization NMR (CIDNP-NMR) studies by H.C. Siebert (personal communication). These experiments have shown that the signals pertaining to two of the three tyrosines in the binding site (presumably the two more highly exposed Tyr21 and Tyr23), could be suppressed in the presence of GlcNAc-oligomers.

Carbohydrate-lectin interactions are relatively weak, because the binding sites are located on the surface and are exposed to solvent (Imberty et al., 1993). For example, estimates of binding constants (K_d) for the binding of GlcNAc-oligomers and NeuLac to WGA range from 10^{-3} to 10^{-5}M using a variety of techniques (Table 9 and references within). Similarly, weak binding has been observed for other lectins such as Con A ($K_d \approx 10^{-4}$ to 10^{-6}M; Mandal et al., 1994), pea lectin ($K_d \approx 10^{-3}$ to 10^{-6}M; Stubbs et al., 1986) and Winged Bean Lectin ($K_d \approx 10^{-4}$M; Schwarz et al., 1991). For WGA, disruption of the dimer results in approximately three fold weaker binding (Privat et al., 1976). Thus, in the case of the graminea lectins, the affinity is strongest when the molecule is a dimer. In contrast, lectins like Con A and pea lectin (legumes) and Winged Bean Lectin have one saccharide binding

site per monomer and each site functions independently for the multimeric proteins.

The forces driving the saccharide binding to WGA, Con A and Winged Bean Lectin are enthalpic in nature ($\Delta H \approx -6$ to -19 kcal/mol; $T\Delta S \approx -2$ to -14 kcal/mol) (Munske et al., 1984; Kronis and Carver, 1985; Schwarz et al., 1991; Bains et al., 1992; Mandal et al., 1994). Changes in the magnitudes and signs of the enthalpy and entropy terms are sensitive to structural changes that accompany ligand binding. The patterns observed in all these studies suggest that the binding interactions are essentially of a polar nature (H-bonds and van der Waals contacts dominate) (Eftink and Biltonen, 1980; Imberty et al., 1993).

To explain why our binding affinities are so low, several questions can be addressed: 1) Is the identity of the binding site the same as that in WGA? 2) To what extent is dimerization necessary for binding? 3) Is there any evidence that the binding site is extended? 4) Is there a preference for certain saccharides? These questions will be addressed in the discussion below.

Chitin Binding

The chitin binding experiments were only useful as an initial screening test to assess saccharide binding function. In the case of hevein, the presence of the three tryptophan

residues allowed detection of the protein. Binding was demonstrated as a function of increasing amounts of chitin in the sample (Figures 24 & 25). Comparison of rhev with nhev revealed weaker binding for the recombinant material.

Direct comparisons of BS and BA with rhev could not be made in the chitin binding assay because of the different methods used to quantitate the protein. Since the binding sites of the domain B proteins contain tyrosine residues, binding was monitored by a protein concentration determination assay (Table 6). The protein concentrations used in the BS and BA assays were ten fold higher as compared with those used for hevein. The low degree of binding observed for BA and BS may be a result of the higher protein concentration used, and thus could be rationalized in terms of saturation of the available sites on the chitin. Tyrosine fluorescence spectroscopy would have provided a more viable alternative method. In such an assay protein concentrations similar to those used in the hevein assay may have yielded more realistic results.

Intrinsic Tryptophan Fluorescence λ_{max} Shift of WGA, nHev and rHev in the Presence of Oligosaccharides

It was known from previous studies on WGA, that the intrinsic tryptophan fluorescence spectrum is perturbed as a function of saccharide binding (Wright, 1984; Privat et al.,

1974a). This suggested the presence of Trp residues in the saccharide binding site. Similar behavior is observed here (Table 7) when WGA, rHev and nHev were incubated with $(GlcNAc)_{3,4}$. The observed blue shift indicates that the saccharide is likely to bind and shield the tryptophan sidechains from the aqueous environment. Other investigators have used intrinsic tryptophan fluorescence of WGA to determine binding constants for GlcNAc oligosaccharides (Lotan and Sharon, 1973; Privat et al., 1974a). These calculations were not made with the recombinant hevein, because even though a slight increase in fluorescence intensity was observed concurrent with the blue shift, the values were not very consistent. This may have been due to inaccuracy in pippeting. The increase in fluorescence intensity upon saccharide binding is presumably due to exclusion of water around the tryptophan residues, which quenches the fluorescence emission. The BS and BA domains, which have tyrosines instead of tryptophans in the binding site, were not subjected to this assay. As an alternative technique, which was not pursued, O-(4-methyl umbelliferyl)-glycosides could have been used to carry out a direct comparison between the binding affinities of all three recombinant proteins.

ITC BINDING STUDIES

Isothermal titration calorimetry is a technique by which

both binding constants and thermodynamic parameters can be determined in the same experiment. This technique measures binding directly in terms of the heat absorbed or evolved for each stepwise addition of ligand. From the binding isotherm the energy components, ΔS and ΔH, are derived. The magnitude and sign of ΔH and ΔS provide information as to the interactions that dominate the binding event. Other methods traditionally used for determining binding constants were less suitable for this study. For example, equilibrium dialysis posed difficulties, because of the small differences in molecular weight between the domain proteins and the oligosaccharides. This technique would have required availability of radiolabelled ligands and dialysis tubing with a very narrow molecular weight range in the case of the tri- and tetra-saccharides. Fluorescence spectrometry, using fluorescence-labelled saccharides, would have provided an alternative method for determination of the binding constants (Privat et al., 1974b; van Landschoot et al., 1977). However, both of these techniques do not yield direct information as to the energetics of the reaction (ΔH & ΔS).

Consideration of Experimental Conditions and Data Analysis

The proper choice of experimental conditions is crucial to obtain reliable and meaningful results. Conditions have to be chosen such that the heat released (or absorbed) is greater

than the background heat of dilution, and such that it detectably decreases with each additional injection of ligand over 10-20 injections. One important limitation of the technique, which may have posed obstacles to definitive answers, consists of the requirement of large amounts of protein in order to perform the experiment optimally. In trying to recover the protein after each experiment, significant losses occurred because of the small differences in molecular weight between the trisaccharide and protein (≈4,000 Da). This limited the amounts of protein that could be recovered for additional experiments. Moreover, it was necessary to perform several experiments at different temperatures to obtain ΔCp values. However, since the ΔH values calculated at different temperatures revealed inconsistencies, reliable ΔCp values could not be calculated in any of the binding experiments.

Ideal experimental conditions require protein concentrations that, when multiplied by the association constant, yield values (C) between 10 and 100 ($C = K_a \times [prot]_{tot}$) (C ranges between 1 and 1000 are generally acceptable). This product C describes the binding isotherm, which is sigmoidal in shape. Extremely flat ($C<1$) or extremely steep curves ($C\rightarrow\infty$), do not yield useful information. In the experiments carried out here, C varied between 0.5 and 4.6 (Table 8). To improve the binding isotherm, higher

protein concentrations would be necessary. However, due to solubility considerations this was not feasible with the recombinant domains. Interestingly, Bains et al (1992) performed their ITC experiments using protein concentrations which yielded C values between 0.15 and 0.7. However, in their calculations 4 binding sites per molecule have to be considered.

The decision to use $(GlcNAc)_3$ as the specific saccharide to carry out quantitative binding studies of each recombinant domain, was based on previous studies with WGA which showed that the affinity of $(GlcNAc)_3$ was optimal (Privat et al., 1974a; Bains et al., 1992). Increasing affinities were reported as a function of (GlcNac)-oligomer length up to $(GlcNAc)_3$. This suggested an extended binding site for WGA. The K_a was essentially constant for longer oligomers.

There are three parameters which can be varied (n, K_a, ΔH) in equation 1 (see above) when analyzing the data. Wiseman et al (1989) recommended holding one of these variables constant, especially when working at low C values. Thus, prior knowledge of the stoichiometry can be applied to fix the number of binding sites (n). In the data analysis for BS and BA, n was held constant at 1 for the monomeric proteins, which possess only one saccharide binding site (Table 8). Alternatively, one could perform a single injection experiment from which ΔH can be determined by

graphical analysis. This involves plotting $dQ/d[ligand]_{tot}$ as a function of $[ligand]_{tot}/[protein]_{tot}$ to obtain the binding isotherm. A value for ΔH can be obtained from the intersection of the binding isotherm with the y-axis. This value can subsequently be held constant for the analysis of the titration data. However, the C value has to be sufficiently high so that the binding isotherm intersects the Y-axis at the true ΔH. Very weak binding (as in the case of the recombinant domain B proteins) produces a binding isotherm that generally intersects the Y-axis below the true ΔH. Thus, we concluded that it was reasonable to hold n constant as otherwise an accurate ΔH could not be obtained.

In the data analysis of WGA and nHev, all three variable parameters were allowed to float. This yielded n values of 4.4 and 4.9, for (GlcNAc)3 and NeuLac binding to WGA (Table 8), and agrees with earlier predictions of 4 sites/WGA dimer (Privat et al., 1974b; Kronis and Carver, 1982, Bains et al., 1992). If n was held constant, the only value that would yield a satisfactory fit of the data points was a value of 8. Considering the prior knowledge of 4 sites/dimer, this value seemed unreasonable. Limitations of this type of data analysis can be encountered with weak binding, as is the case with WGA and the recombinant domains.

The reliability of the experimental data quoted in Table 8, can be judged by the magnitudes of the standard deviations

(SD). These values reflect the goodness of fit of the theoretical binding isotherm to the experimental data. As apparent from Table 8 the values obtained were small (<4 for the recombinant domains and <12 for WGA) and comparable to the values quoted by Bains et al (1992) (SD = 8 to 32).

Comparison of the Saccharide Affinities of BS versus BA

BS and BA behaved similarly in all ITC experiments with respect to both binding constants and thermodynamic parameters (Table 8). The $(GlcNAc)_{3\&4}$ oligosaccharides bound with similar binding constants to both recombinant proteins (K_d ≈2.1 to $5.3x10^{-4}M$). We therefore concluded that the mutation at position 28 had no effect on GlcNAc binding, as expected. Binding was not detected with $(GlcNAc)_2$ or NeuLac under the experimental conditions used. However, it cannot be ruled out that the disaccharide may bind weakly and that detection is beyond the limits of the experimental conditions used.

Analysis of the energetics accompanying binding of $(GlcNac)_3$ also revealed comparable values for BS and BA. Positive values were obtained for $T\Delta S$ (3 to 4 kcal/mol) and small negative values were obtained for ΔH (-0.8 to -1.9 kcal/mol) (Table 8). The magnitude and sign of these values indicate that the binding event is entropically driven with only small enthalpic contributions. This trend is characteristic of changes in the electrostatic or hydrophobic

environments and can also be indicative of conformational changes (Ream et al, 1992; Eftink and Biltonen, 1980).

Comparable values were obtained for the binding of both $(GlcNAc)_{3\&4}$ to BA (Table 8). However, $(GLcNAc)_4$ binding to BS, carried out at higher protein concentration (≈3mg/ml), displayed some variations in the thermodynamic parameters. It is possible that these values may not be reliable, if the higher protein concentration caused aggregation.

As postulated above, the binding affinities of BS and BA could only be distinguished using the trisaccharide NeuLac, if these mutants formed true 2-fold related dimers in solution reminiscent of the mode of domain/domain association in WGA (Wright, 1987). Since NeuLac binding to WGA was shown to be weak (K_d ≈$3.88x10^{-4}$M), our inability to detect NeuLac binding with the BS and BA proteins was not surprising. Judging from the difference in the known binding constants for NeuLac versus $(GlcNAc)_3$ binding to WGA (K_d's of $3.88x10^{-4}$ and $0.54x10^{-4}$ M, respectively), an approximately seven fold weaker affinity was expected for NeuLac binding in case of a putative BS dimer. A further decrease in the binding affinity would have been estimated for the putative BA dimer. Our finding that the recombinant domains are monomeric, further explains the absence of binding as judged by this technique. It is doubtful, that even if much higher protein concentrations were used, binding could be detectable with this trisaccharide.

The binding interactions from the aromatic pocket alone are identical in WGA for these two types of saccharides (GlcNAc and NeuNAc), and involve two H-bonds (see Figure 3B) (Wright, 1990). In the WGA dimer NeuLac is stabilized by two H-bonds across the domain/domain interface (Figure 4). One corresponds to the interaction between the amido NH of the saccharide and Glu29, which is also present when GlcNAc is bound. The other H-bond is unique to NeuLac binding and is between the carboxylate group of the saccharide and the OH of Ser28.

Comparison of Saccharide Binding of the Recombinant Domains versus WGA

For comparison purposes, binding studies were also performed on WGA using identical saccharides. As apparent from Table 8, much weaker binding affinities and different trends for ΔH and ΔS were obtained for BS and BA, in comparison to those for WGA. Binding of $(GlcNAc)_3$ to BS and BA is eight fold weaker compared with its binding to WGA (K_d $\approx 4 \times 10^{-4}$M and 0.5×10^{-4}M, respectively)(Table 8). This may be attributed to the monomeric nature of the recombinant domains. The auxiliary H-bond present between Glu29 (from domain C) and the amido NH group of GlcNAc in the WGA B-site is absent in the saccharide complexes of the monomeric single domain proteins. This H-bond and numerous van der Waals contacts

between the saccharide and domain C could account for the slightly better ΔG and the higher association constants in WGA (Table 8). Moreover, rearrangement and displacement of solvent, expected to differ substantially in the single domain versus the multi-domain situation of WGA, have to be taken into account when interpreting the ΔH and ΔS contributions to the free energy of binding. These values differ in both magnitude and sign. In the case of WGA, binding is accompanied by a negative ΔH and TΔS (-8.2 kcal/mol and -3.2 kcal/mol, respectively) (Table 8). This trend is consistent with an enthalpically driven binding event involving H-bonding and van der Waals contacts in protein/ligand interactions (Eftink and Biltonen, 1980).

In contrast, the thermodynamic parameters obtained with the single domain proteins, although low in magnitude, suggest an entropically driven event (ΔH≈-1.3 kcal/mol; TΔS≈3.3 kcal/mol) (Table 8), in which hydrophobic or electrostatic interactions are expected to dominate (Eftink and Biltonen, 1980). This result is difficult to rationalize in terms of the binding interactions observed in the WGA-$(GlcNAc)_2$ complex (Wright, 1984). In the crystal structure of all complexes examined, three H-bonds are formed with the non-reducing terminal GlcNAc residue. Two of these are made with residues in the aromatic pocket and the third is the auxiliary H-bond from domain C (Figure 4). H-bonds typically contribute ≈2

kcal each to ΔH. While a positive ΔS value suggests that hydrophobic interactions are involved, the hydrophobic contacts are expected to be weak and restricted only to contacts between apolar atoms of the three tyrosine residues and the N-Acetyl methyl group and the apolar parts of the saccharide ring. Nevertheless, hydrophobic interactions may be the logical driving force for saccharide binding in the case of the single domain proteins. The small change in enthalpy may suggest that little heat gain is associated with displacement of solvent by the saccharide in comparison to WGA. Conceivably, the gain in entropy may reflect a lesser degree of conformational restrictions of the protein in addition to apolar inter-atomic contacts and effects caused by the release of ordered water around the aromatic residues. A positive ΔS and negligible ΔH can also indicate the involvement of either electrostatic forces or conformational changes. There is no evidence for either from the crystal structure.

The dimeric multi-domain situation in WGA represents a much more complex system, as both global inter-subunit adjustments and rearrangements of larger networks of ordered solvent may contribute substantially to the overall ligand-binding energetics. For instance, a tightening of the dimer structure, suggesting a loss of rotational and translational freedoms, was observed in the NeuLac-WGA crystal structure

when comparing atomic temperature factors relative to the unliganded structure (Wright, 1990).

Comparison of (GlcNAc)4 Binding Between nHevein and the Recombinant Domains

Comparing the B domain binding constants to that of hevein suggests that hevein has a binding site with ≈20 fold higher affinity for $(GlcNAc)_4$. This large difference may be accounted for by the different character of the aromatic pocket in hevein, which contains two tryptophan residues. The equal enthalpy and entropy components suggest that both are driving the binding interactions in hevein ($T\Delta S \approx 3.09$ kcal/mol; $\Delta H \approx -3.1$ kcal/mol). The larger enthalpy change observed for hevein, can be interpreted in terms of stronger van der Waals interactions present in the hevein-$(GlcNAc)_4$ complex. In contrast to the tyrosine residues in domain B, tryptophan residues allow more extensive van der Waals stacking interactions with the pyranoside rings of the two adjacent GlcNAc residues at the non-reducing end of the tetrasaccharide.

Evaluation of the WGA ITC Binding Results

Table 9 summarizes the binding constants for GlcNAc oligomers and NeuLac measured by different investigators using various techniques. Except for the values based on

equilibrium dialysis (Nagata and Burger, 19??), all the K_d values compare within a factor of two of the data presented here. This variation is not large and probably reflects differences in the exact technique employed. For example, only half of the potential binding sites in WGA contain a Trp residue. Thus monitoring the intrinsic tryptophan fluorescence as a function of GlcNAc binding probably reflects binding at those sites only.

The results presented here on $(GlcNAc)_3$ binding compare well with those obtained by Bains et al (1992) using ITC (Table 9). The binding constants we obtained agree to within a factor of two ($K_d \approx 0.54 \times 10^{-4}$M versus $\approx 0.9 \times 10^{-4}$M) (Table 9). Although the magnitudes of ΔG were also in good agreement ($\Delta G \approx$ -5.0 kcal/mol versus -5.5 kcal/mol, Bains et al), the enthalpy and entropy parameters showed significant discrepancies ($\Delta H \approx$ -8.23 kcal/mol versus -16.0 kcal/mol, and for $T\Delta S \approx$-3.21 kcal/mol versus -10.4 kcal/mol). It is possible that this difference is related to the different buffer conditions used. The titrations of Bain et al (1992) were performed at higher ionic strength and pH than used in these studies. Such variations could have caused significant variation in ΔH and ΔS (Eftink and Biltonen, 1980).

The results obtained for Neulac binding to WGA reflect extremely weak binding ($K_d \approx 3.9 \times 10^{-4}$M) and may not be reliable. However, the K_d value was comparable with those quoted by

Kronis and Carver (1982) based on NMR titration data ($K_d \approx 8.7 \times 10^{-4} M$). Analysis of our data in terms of the energetics of the binding of this trisaccharide, suggests that the reaction is entropically driven with a significant positive $T\Delta S$ value (4.3 kcal/mol) and a negligable enthalpic contribution ΔH (-0.39 kcal/mol). These trends are similar to those observed for the recombinant domains. However, they are in contrast with Kronis and Carver (1985), who reported both a larger negative value for ΔH (-13.3 kcal/mol) and $T\Delta S$ (-9.5 kcal/mol). Although the ΔG values are in close agreement (Rice, ≈-4.7 kcal/mol; Kronis and Carver, ≈-3.8 kcal/mol), we have no explanation for the discrepancies in ΔH and ΔS. It is possible that they are related to our failure to perform the ITC titration at sufficiently high protein concentrations (C=0.5), or to differences in experimental conditions, such as the choice of buffers and ionic strength (Rice used 50mM sodium phosphate, pH 6.8; Kronis and Carver used 100mM sodium phosphate, 150mM NaCl, pD 6.1). The trends in the enthalpy and entropy changes reported by Kronis and Carver (1985) compare, however, very closely with those quoted by Bains et al (1992) for $(GlcNAc)_2$ and $(GlcNAc)_3$ binding, and with values reported for other lectin-saccharide studies (Munske et al., 1984; Schwarz et al., 1991; Mandal et al., 1994). These trends suggest that H-bonding and van der waals forces dominate lectin-saccharide binding interactions.

CONCLUSION

In conclusion, we have developed an efficient protocol for generating functional recombinant WGA domains, which are capable of binding GlcNAc-oligosaccharides. It has been established from the intrinsic fluorescence data and preliminary NMR studies, that the domain binding site is identical to that of WGA consisting of three aromatic residues. As in WGA, the binding site is extended in terms of allowing maximal interaction with a GlcNAc oligomer which is at least a trimer in length. However, the binding strength of all saccharides tested was found to be substantially lower. This was explained based molecular weight determination, which suggested that the recombinant B-domain proteins are monomeric. We concluded, therefore, that interactions from the aromatic pocket alone contribute to the observed binding parameters and are sufficient for GlcNAc binding, and that the entire WGA binding site (B-C domain contact) could not be reproduced. Nevertheless, the expression system described here now opens up the possibility to study the folding of the disulfide stabilized domain by site directed mutagenesis.

LITERATURE CITED

LITERATURE CITED

Allen, A.K., A. Neuberger and N. Sharon, "The Purification, Composition and Specificity of Wheat-Germ Agglutinin", Biochemical Journal, 131:155-162, (1973).

Amrein, Kurt E., Baerbel Panholzer, Nicholas A. Flint, Willi Bannwarth, and Paul Burn, "The Src homology 2 domain of the protein-tyrosine kinase $p56^{lck}$ mediates both intermolecular and intramolecular interactions" Procedings from the National Academy of Sciences, 90:10285-10289, (1993).

Andersen, Niels H., Bolong Cao, Adela Rodrigues-Romero, and Barbarin Arreguin, "Hevein: NMR Assignment and Assessment of Solution-State Folding for the Agglutinin-Toxin Motif", Biochemistry, 32:1407-1422, (1993).

Aub, Joseph C., Carol Tieslau, and Ann Lankester, "Reaction of Normal and Tumor Cell Surfaces to Enzymes, I. Wheat-Germ Lipase and Associated Mucopolysaccharides", Procedings of the National Academy of Sciences, 50:613-619, (1963).

Bains, Gabrielle, Reiko T. Lee, Yuan C. Lee, and Ernesto Freire, "Microcalorimetric Study of Wheat Germ Agglutinin Binding to N-Acetylglucosamine and Its Oligomers", Biochemistry, 31:12624-12628, (1992).

Berger, Shelby L. and Alan R. Kimmel (eds), Methods in Enzymology Guide to Molecular Cloning Techniques, Volume 152 Academic Press, Inc., San Diego, (1987).

Burger, Max M., "A Difference in the Architecture of the Surface Membrane of Normal and Virally Transformed Cells", Procedings of the National Academy of Sciences, 62:994-1001, (1969).

Burger, Max M. and Allan R. Goldberg, "Identification of a Tumor-Specific Determinant on Neoplastic Cell Surfaces", Procedings of the National Academy of Sciences, 57:359-366, (1967).

Burnette, W. Neal, ""Western Blotting": Electrophoretic Transfer of Proteins from Sodium Dodecyl Sulfate-

Polyacrylamide Gels to Unmodified Nitrocellulose and Radiographic Detection with Antibody and Radioiodinated Protein A", Analytical Biochemistry, 112:195-203, (1981).

Chrispeels, Maarten J. and Natasha V. Raikhel, "Lectins, lectin genes and their role in plant defense", Plant Cell, 3(1):1-9, (1991).

Davis, J.M., H.R.G. Clarke, H.D. Bradshaw, Jr. and M.P. Gordon, :Populus chitinase genes: structure organization, and similarity of translated sequences to herbaceous plant chitinases", Plant Molecular Biology, 17:631-639, (1991).

Drickamer, Kurt, "Minireview: Clearing Up Glycoprotein Hormones", Cell, 67:1029-1032, (1991).

Eftink, Maurice and Rodney Biltonen "Thermodynamics of Interacting Biological Systems", Biological Microcalorimetry, Ed. A.E. Beezer, Academic Press, London, 343-412, (1980).

Etzler, Marilynn E., "Plant Lectins: Molecular and Biological Aspects", Annual Review of Plant Physiology, 36:209-234, (1985).

Farmer, Terry B. and Richard M. Caprioli, "Assessing the Multimeric States of Proteins: Studies Using Laser Desorption Mass Spectrometry", Biological Mass Spectrometry, 20:796-800, (1991).

Farrell, F.X., C.M. Sax and Z.E. Zehner,. "A Negative Element Involved in Vimentin Gene Expression", Molecular and Cellular Biology, May 10(5):2349-2358, (1990).

Freire, Ernesto, Obdulio L. Mayorga, and Martin Straume, "Isothermal Titration Calorimetry", Analytical Chemistry, 62:950A-959A, (1990).

Hemsley, Anne, Norman Arnheim, Michael D. Toney, Gino Cortopassi and David J. Galas, "A simple method for site-directed mutagenesis using the polymerase chain reaction", Nucleic Acids Research, 17(16):6545-6551, (1989).

Huang, J.K., L. Wen, M. Swegle, H.C. Tran, T.H. Thin, H.M. Naylor, S. Muthukrishnan and G.R. Reeck, "Nucleotide sequence of a rice genomic clone that encodes a class I endochitinase", Plant Molecular Biology, 16:479-480, (1991).

Imberty, A., Y. Bourne, C. Cauibillau, P. Rouge, and S. Peres, "Protein/Carbohydrate Complexes", Advances in Biophysical Chemistry, Vol.3, 106-117, (1993).

Jenkins, W. Terry, "Hydroxylapatite For Protein Chromatography", Biochemical Preparations, 9:83-85, (1962).

Kobata, Akira, "Review: Structures and functions of the sugar chains of glycoproteins", European Journal of Biochemistry, 209:483-501 (1992).

Kronis, K. Anne and Jeremy P. Carver, "Specificity of Isolectins of Wheat Germ Agglutinin for Sialyloligosaccharides: A 360-MHz Proton Nuclear Magnetic Resonance Binding Study" Biochemistry, 21:3050-3057, (1982).

Kronis, K. Anne and Jeremy P. Carver, "Thermodynamics of Wheat Germ Agglutinin-Sialyloligosaccharide Interactions by Proton Nuclear Magnetic Resonance", Biochemistry, 24:834-840, (1985).

Laemmli, U.K., "Cleavage of Structural Proteins during the Assembly of the head of Bacteriophage T4", Nature, 227:680-685, (1970).

Lee, Y.C. "Biochemistry of carbohydrate-protein interaction", The FASEB Journal, 6:3193-3200 (1992).

Lotan, Reuben and Nathan Sharon, "The Fluorescence of Wheat Germ Agglutinin and of its Complexes with Saccharides", Biochemical and Biophysical Research Communications, 55:1340-1346, (1973).

Mandal, Dipak K., Nand Kishore, and C. Fred Brewer, "Thermodynamics of Lectin-Carbohydrate Interactions. Titration Microcalorimetry Measurements of the Binding of N-Linked Carbohydrates and Ovalbumin to Concanavalin A", Biochemistry, 33:1149-1156, (1994)

Maniatis, T., E.F. Fritsch and J. Sambrook, Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, (1982).

Mansfield, Michael A., Willy J. Peumans and Natasha V. Raikhel, "Wheat-germ agglutinin is synthesized as a glycosylated precursor", Planta, 173:482-489, (1988).

Mansfield, Michael A. and Natasha V. Raikhel, "Abscisic acid enhances the transcription of wheat-germ agglutinin mRNA without altering its tissue-specific expression", Planta, 180:548-554, (1990).

Midoux, P., J.-Ph. Grivet and M. Monsigny, "Lectin-Sugar Interactions: the Binding of 1-O-Methyl-di-N-Trifluoroacetyl-B-Chitobioside to Wheat Germ Agglutinin", FEBS Letters,

120:29-32, (1980).

Mishkind, Michael L., Barry A. Palevitz, Natasha V. Raikhel and Kenneth Keegstra, "Localization of Wheat Germ Agglutinin-Like Lectins in Various Species of the Gramineae", Science, 220:1290-1292, (1983).

Monsigny, Michel, Annie-Claude Roche, Claude Sene, Regine Maget-Dana, and Francis Delmotte, "Sugar-Lectin Interactions: How Does Wheat-Germ Agglutinin Bind Sialoglycoconjugates?", European Journal of Biochemistry, 104:147-153, (1980).

Munske, Gehard, R., Henry Krakauer, and James A. Magnuson, "Calorimetric Study of Carbohydrate Binding to Concanavalin A", Archives of Biochemistry and Biophysics, 233(2):582-587, (1984).

Nagata, Yoshiho and Max M. Burger, "Wheat Germ Agglutinin", The Journal of Biological Chemistry, 249:3116-3122, (1974).

Oda, Yuko, Jorg Herrmann, Michael A. Gitt, Christoph W. Turck, Alma L. Burlingame, Samuel H. Barondes, and Hakon Leffler, "Soluble Lactose-binding Lectin from Rat Intestine with Two Different Carbohydrate-binding Domains in the Same Peptide Chain", The Journal of Biological Chemistry, 268:5929-5939, (1993).

Parsons, T.G., H.D. Bradshaw, Jr. and M.P. Gordon, "Systemic accumulation of specific mRNA's in response to wounding poplar trees", Proceedings from the National Academy of Sciences, 86:7895-7899, (1989).

Peters, Barry P., Shigeyuki Ebisu, Irwin J. Goldstein, and Michael Flashner, "Interaction of Wheat Germ Agglutinin with Sialic Acid", Biochemistry, 18:5505-5511, (1979).

Privat, Jean-Paul, Francis Delmotte, Gerard Mialonier, Philippe Bouchard, and Michel Monsigny, "Fluorescence Studies of Saccharide Binding to Wheat-Germ Agglutinin (Lectin), European Journal of Biochemistry, 47:5-14, (1974a).

Privat, Jean-Paul, Francis Delmotte and Michel Monsigny, "Protein-Sugar Interactions, Association of Wheat Germ Agglutinin (Lectin) and O-(-4-Methyl-Umbelliferyl)-Glycosides", FEBS Letters, 46:229-232, (1974b).

Privat, Jean-Paul, Reuben Lotan, Philippe Bouchard, Nathan Sharon, and Michel Monsigny, "Chemical Modification of the Tryptophan Residues of Wheat-Germ Agglutinin", European Journal of Biochemistry, 68:563-572, (1976).

Protocols for the MBP Vectors, New England Biolabs, Inc., Boston, MA, 1-20, (1989).

Rademacher, T.W., R.B. Parekh, and R.A. Dwek, "Glycobiology", Annual Review of Biochemistry, 57:785-838 (1988).

Raikhel, Natasha V., Sebastian Y. Bednarek, and Thea A. Wilkins, "Cell-type-specific expression of a wheat-germ agglutinin gene in embryos and young seedlings of *Triticum aestivum*", Planta, 176:406-414, (1988).

Raikhel, N.V., H.-I. Lee and W.F. Broekaert, "Structure and Function of Chitin-Binding Proteins", Annual Review of Plant Physiology and Plant Molecular Biology, 44:591-615, (1993).

Ransone, Lynn J., Jane Visvader, Paolo Sassone-Corsi, and Inder M. Verma, "Fos-Jun interaction: mutational analysis of the leucine zipper domain of both proteins", Genes and Development, 3:770-781, (1989).

Ream, Joel E., Henry K. Yuen, Ronald B. Frazier, and James A. Sikorski, "EPSP Synthase: Binding Studies Using Isothermal Titration Microcalorimetry and Equilibrium Dialysis and Theri Implications for Ligand Recognition and Kinetic Mechanism", Biochemistry, 31:5528-5534, (1992).

Rice, Robert H. and Marilynn E. Etzler, "Subunit Structure of Wheat Germ Agglutinin", Biochemical and Biophysical Research Communications, 59:414-419, (1974).

Rice, Robert H. and Marilynn E. Etzler, "Chemical Modification and Hybridization of Wheat Germ Agglutinins", Biochemistry, 14:4093-4099, (1975).

Rodriguez-Romero, A., K.G. Ravichandran and M. Soriano-Garcia, "Crystal structure of hevein at 2.8 Å resolution", FEBS Letters, 291:307-309, (1991).

Sairam, M.R., "Role of carbohydrates in glycoprotein hormone signal transduction", The FASEB Journal, 3:1915-1926 (1989).

Sambrook, J., E.F. Fritsch, T. Maniatis, Molecular Cloning: A Laboratory Manual, 2ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, (1989).

Schroeder, Martin R. and Natasha V. Raikhel, "Isolation and Characterization of Pro-Barley Lectin Expressed in *Escherichia coli*", Protein Expression and Purification, 3:508-511, (1992).

Schwarz, Frederick P., Kameldeep Puri, and Avadhesha Surolia,

"Thermodynamics of the Binding of Galactopyranoside Derivatives to the Basic Lectin from Winged Bean (*Psophocarpus tetrogonolobus*), Journal of Biological Chemistry, 266(36):24344-24350, (1991).

Sharon, Nathan and Halina Lis, "Lectins as Cell Recognition Molecules", Science, 246:227-234 (1989).

Shinshi, H., J.M. Neuhaus, J. Ryals and F. Meins, Jr., "Structure of a tobacco endochitinase gene: evidence that different chitinase genes can arise by transposition of sequences encoding a cysteine-rich domain", Plant Molecular Biology, 14:357-368, (1990).

Stubbs, Marlene E., Jeremy P. Carver, and Robert J. Dunn, "Production of Pea Lectin in *Escherichia coli*", The Journal of Biological Chemistry, 261:6141-6144, (1986).

Triplett, Barbara A. and Ralph S. Quatrano, "Timing, Localization, and Control of Wheat Germ Agglutinin Synthesis in Developing Wheat Embryos", Developmental Biology, 91:491-496, (1982).

Tsuji, Takashi, Ryusuke Nakagawa, Nobuyuki Sugimoto, and Ken-ichi Fukuhara, "Characterization of Disulfide Bonds in Recombinant Proteins: Reduced Human Interleukin 2 Inclusion Bodies and Its Oxidative Refolding", Biochemistry, 26:3129-3134, (1987).

Vallette, Francois, Emmanuelle Mege, Allison Reiss and Milton Adesnik, "Construction of mutant and chimeric genes using the polymerase chain reaction", Nucleic Acids Research, 17(2):723-733, (1989).

Van Buuren, M., J.M. Neuhaus, H. Shinshi, J. Ryals and F. Meins, Jr., " The structure and regulation of homeologous tobacco endochitinases genes of *Nicotiana sylvestris* and *N. tometosiformis* origin", Molecular Genes and Genetics, 232:460-469, (1992).

Van Landschoot, Anita, Frank G. Loontiens, Robert M. Clegg, Nathan Sharon, and Clemens K. De Bruyne, "Binding of 4-Methylumbelliferyl N-Acetyl-Chitooligosaccharides to Wheat-Germ Agglutinin, A Reinvestigation of Equilibrium Studies", European Journal of Biochemistry, 79:275-283, (1977).

Van Parijs, Jan, Willem F. Broekaert, Irwin J. Goldstein, and Willy J. Peumans, "Hevein: an antifungal protein from rubber-tree (*Hevea brasiliensis*) latex", Planta, 183:258-264, (1991).

Wales, Richard, Peter T. Richardson, Lynne M. Roberts, Hugh R. Woodland, and J. Michael Lord, "Mutational Analysis of the Galactose Binding Ability of Recombinant Ricin B Chain", The Journal of Biological Chemistry, 266:19172-19179, (1991).

Wales, Richard, Peter T. Richardson, Lynne M. Roberts, and J. Michael Lord, "Recombinant Ricin B Chain Fragments Containing a Single Galactose Binding Site Retain Lectin Activity", Archives of Biochemistry and Biophysics, 294:291-296, (1992).

Wiseman, Thomas, Samuel Williston, John F. Brandts, and Lung-Nan Lin, "Rapid Measurement of Binding Constants and Heats of Binding Using a New Titration Calorimeter", Analytical Biochemistry, 179, 131-137, (1989).

Wright, Christine S., "The Crystal Structure of Wheat Germ Agglutinin at 2.2 A Resolution", Journal of Molecular Biology, 111:439-457, (1977).

Wright, Christine S., "Crystallographic Elucidation of the Saccharide Binding Mode in Wheat Germ Agglutinin and its Biological Significance", Journal of Molecular Biology, 141: 267-291, (1980).

Wright, Christine S., "Structural Comparison of the Two Distinct Sugar Binding Sites in Wheat Germ Agglutinin Isolectin II", Journal of Molecular Biology, 178:91-104, (1984).

Wright, Christine S., Francisco Gavilanes, and Darrell L. Peterson, "Primary Structure of Wheat Germ Aglgutinin Isolectin 2. Peptide Order Deduced from X-ray Structure", Biochemistry, 23:280-287, (1984).

Wright, Christine S. and Itzhak Kahane, "Preliminary X-ray Diffraction Results on CO-Crystals of Wheat Germ Agglutinin with a Sialoglycopeptide from the Red Cell Receptor Glycophorin A", Journal of Molecular Biology, 194:353-355, (1987).

Wright, Christine S. and Natasha Raikhel, "Sequence Variability in Three Wheat Germ Agglutinin Isolectins: Products of Multiple Genes in Polyploid Wheat", Journal of Molecular Evolution, 28:327-336, (1989).

Wright, Christine S., "2.2Å Resolution Structure Analysis of Two Refined N-acetylneuraminyl-lactose-Wheat Germ Agglutinin Isolectin Complexes", Journal of Molecular Biology, 215:635-651, (1990).

Wright, Christine S., "Crystal Structure of a Wheat Germ Agglutinin/Glycophorin-Sialoglycopeptide Receptor Complex", The Journal of Biological Chemistry, 267:14345-14352, (1992).

Wright, Christine S., Martin R. Schroeder and Natasha V. Raikhel, "Crystallization and Preliminary X-ray Diffraction Studies of Recombinant Barley Lectin and Pro-barley Lectin", Journal of Molecular Biology, 233:322-324, (1993).

Wright, H. Tonie, Gnanakaran Sandrasegaram and Christine S. Wright, "Evolution of a Family of N-Acetylglucosamine Binding Proteins Containing the Disulfide-Rich Domain of Wheat Germ Agglutinin", Journal of Molecular Evolution, 33:283-294 (1991).

Zheng, Jian, Florian Schodel, and Darrell L. Peterson, "The Structure of Hepadnaviral Core Antigens", The Journal of Biological Chemistry, 267:9422-9429. (1992).

VITA