uOttawa

L'Université canadienne
Canada's university

FACULTÉ DES ÉTUDES SUPÉRIEURES ET POSTOCTORALES

uOttawa

L'Université canadienne
Canada's university

FACULTY OF GRADUATE AND POSDOCTORAL STUDIES

Melissa McNulty
AUTEUR DE LA THÈSE / AUTHOR OF THESIS

M.Sc. (Biochemistry)
GRADE / DEGREE

Department of Biochemistry, Microbiology and Immunology
FACULTÉ, ÉCOLE, DÉPARTEMENT / FACULTY, SCHOOL, DEPARTMENT

Preliminary Characterization of Wheat, *Triticum aestivum,* Embryo Globulins

TITRE DE LA THÈSE / TITLE OF THESIS

Illimar Altosaar (absent)
DIRECTEUR (DIRECTRICE) DE LA THÈSE / THESIS SUPERVISOR

CO-DIRECTEUR (CO-DIRECTRICE) DE LA THÈSE / THESIS CO-SUPERVISOR

EXAMINATEURS (EXAMINATRICES) DE LA THÈSE / THESIS EXAMINERS

K. Baetz

F. Scott

Gary W. Slater
Le Doyen de la Faculté des études supérieures et postdoctorales / Dean of the Faculty of Graduate and Postdoctoral Studies

Preliminary characterization of wheat, *Triticum aestivum*, embryo globulins

Melissa S. M^{c}Nulty

Thesis submitted to the
Faculty of Graduate and Postdoctoral Studies
In partial fulfillment of the requirements
For the MSc degree in Biochemistry

Biochemistry, Microbiology and Immunology
Faculty of Medicine
University of Ottawa

Your file Votre référence
ISBN: 978-0-494-25805-7
Our file Notre référence
ISBN: 978-0-494-25805-7

Abstract

The 7S globulins are a subclass of seed storage proteins characterized by their solubility in saline solution. These proteins are major legume storage reserves, and have been studied extensively; in the cereals, they represent a minor seed protein fraction and have been characterized to a lesser extent. Recently, a study associating a wheat 7S globulin, WP5212, with the development of type 1 diabetes in diabetes-prone rats, has renewed interest in this protein class. The present study aimed to better characterize the 7S globulins of wheat. The majority of the wheat embryo globulins were detected by anti-WP5212 polyclonal serum. These proteins varied significantly in their molecular masses and isoelectric points. Six major polypeptides were identified by mass spectrometry and/or N-terminal sequencing as belonging to the globulin1 family. These results combined with prior studies have allowed the construction of a hypothetical model of the post-translational events contributing to the wheat 7S globulin profile in mature kernels.

Acknowledgements

I would like to thank my supervisor, Dr. Illimar Altosaar, for the opportunity to study under his guidance. A big thank you goes out to all my lab mates, particularly, David Blais, Evelin Loit, Charles Melnyk, Yaseen Mottiar, Julianne Staebler and Mohsin Zaidi, each of whom contributed to this project in some way, shape or form.

I am grateful to my TAC committee members Dr. Steve Gleddie (ECORC), Dr. Hans Schernthaner (ECORC) and Dr. Fraser Scott (University of Ottawa and OHRI) for their time and advice.

I also extent thanks to the many graduate students, technicians and researchers who have offered encouragement, advice and equipment, particularly the members of the following labs: Dr. Fraser Scott (University of Ottawa and OHRI), Dr. Zemin Yao (University of Ottawa, BMI), Dr. Steve Gleddie (ECORC) and Dr. Ken Dimock (University of Ottawa, BMI).

Finally, I would like to thank Carol Ann Kelly and Nicole Trudel who make life as graduate student much easier.

Table of Contents

List of Tables

List of Figures

Abbreviations

1D	1-dimensional
2D	2-dimensional
2DE	2-dimensional electrophoresis
3D	3-dimensional
BCA	Bicinchoninic acid
Beg1	Barley embryo globulin1
Beg2	Barley embryo globulin2
BLAST	Basic local alignment search tool
bp	base pair
CAPS	3-(cyclohexylamino)-1-propanesulfonic acid
CBB	Coomassie Brilliant Blue
cDNA	complementary deoxyribonucleic acid
cds	coding sequence
CHAPS	3-([3-cholamidopropyl]dimethylamino)-1-propanesulfonate
d35S	Double 35S promoter (cauliflower mosaic virus)
DMSO	Dimethylsulfoxide
DNA	Deoxyribonucleic acid
dNTP	2'-deoxynucleoside 5'-triphosphates
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic acid
EPP	Endoproteolytic processing
ER	Endoplasmic reticulum
EST	Expressed sequence tag
GG2.0	GrainGenes2.0
Glb1	Globulin1
Glb2	Globulin2
HRP	Horse radish peroxidase
IEF	Isoelectric focussing
IPG	Immobilized pH gradient
Kb	Kilo base (pair)
kDa	kilo Dalton
LB	Left border
LC-MS/MS	Liquid chromatography tandem mass spectrometry
M_r	Relative molecular mass
MVP	Matteuccia struthiopteris (fern) vicilin-like protein
NCBI	National Center for Biotechnology Information
NOSter	Nopaline synthase terminator
NT	Non-transgenic (tobacco plant)
OISB	Ottawa Institute of Systems Biology
ORF	Open reading frame
PCR	Polymerase chain reaction
PDB	Protein Data Bank
pI	Isoelectric point
PSV	Protein storage vacuole
PTM	Post-translational modification
PVDF	Polyvinylidene difluoride
RB	Right border
Reg1	Rice embryo globulin1

Reg2	Rice embryo globulin2
rER	rough endoplasmic reticulum
RNA	Ribonucleic acid
RT	Reverse transcriptase
RT-PCR	Reverse transcriptase-polymerase chain reaction
SBP	Sucrose binding protein
SDS	Sodium dodecyl sulfate
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
T0	Primary transformants
T1	Progeny from T0 transformants
T1D	Types 1 diabetes
TBP	Tributylphosphine
TBST	Tris-buffered saline with 0.5% Tween-20
TCA	Trichloroacetic acid
TIGR	The Institute for Genomic Research
UTMB	University of Texas Medical Branch

Chapter I – Introduction

Seed storage proteins

The process of seed development and maturation is marked by massive protein deposits consisting primarily of storage proteins, proteases and their inhibitors. The storage proteins that accumulate in the seed will become the sources of carbon, nitrogen, sulphur and energy during germination and early embryo development (1). These proteins are of significant agricultural value to the human population. The high protein content of the Leguminoseae remains the main source of protein in many developing and underdeveloped nations (2). While storage protein content is significantly lower in the seeds of the Gramineae, the cereals nonetheless contribute a larger proportion of protein to human and livestock nutrition than the legumes, by the sheer volume of their consumption (3).

The first systematic classification of plant proteins was based on their solubility in various solvents. Four main categories were outlined: the water-soluble albumins, the alcohol-soluble prolamins, the alkali or acid soluble glutelins and the salt-soluble globulins (4). To a large extent, this method of classification has been preserved over time for the purposes of seed storage protein nomenclature. While many of the storage proteins are classified according to their solubility, this can lead to confusion, particularly when a protein is soluble in more than one solvent or when a protein's sequence and/or structure is closely related to that of another solubility class. The α-globulins of wheat, maize and rice are one such example. These 18-25kDa proteins are soluble in saline solutions but share significant homology over approximately half their sequence with a high molecular weight glutenin from wheat (5-9). The rice glutelins are another example. These require NaOH for solubility yet share structure with, and are antigenically-related to, the 11S globulins (10, 11).

The globulins

The globulins are a class of storage proteins, soluble in saline solutions, generally deficient in cysteine and methionine, and widely distributed throughout the plant kingdom (12-14). The majority of seed globulins fall into one of two categories according to their sedimentation coefficients: 7-9S and 11-13S (15). The generalized terms 7S globulins and 11S globulins are used to describe proteins within these groups, respectively.

The 7S and 11S globulin families have been extensively studied in the Leguminoseae, where they are designated vicilins and legumins, respectively. These terms are often used interchangeably to describe the respective groups. For the purposes of this study, the terms vicilin and legumin will be reserved for the specific globulins of pea (*Pisum sativum*). The 7S and 11S globulins share sufficient sequence similarity between the globulins of cereals, legumes and gymnosperms to indicate structure preservation during two major divergence events: angiosperms from gymnosperms and monocots from dicots (14).

The cupin superfamily

The evolutionary history of the 7S and 11S seed storage globulins can be traced back to a small, single domain protein, generically termed "cupin" for its characteristic β-barrel structure (16). The conserved β-barrel structure is an arrangement of two motifs each composed of two β-sheets linked by a variable-length loop region. The compact nature of the cupin domain confers a high degree of thermostability which is a likely explanation for the success of this superfamily. Along with this evolutionary advantage, elaborations and modifications of the basic cupin structure have given rise to proteins of vast functional diversity and extensive distribution. Collectively, these proteins show low primary sequence conservation, yet each retains an adaptation of the characteristic three-dimensional fold (17).

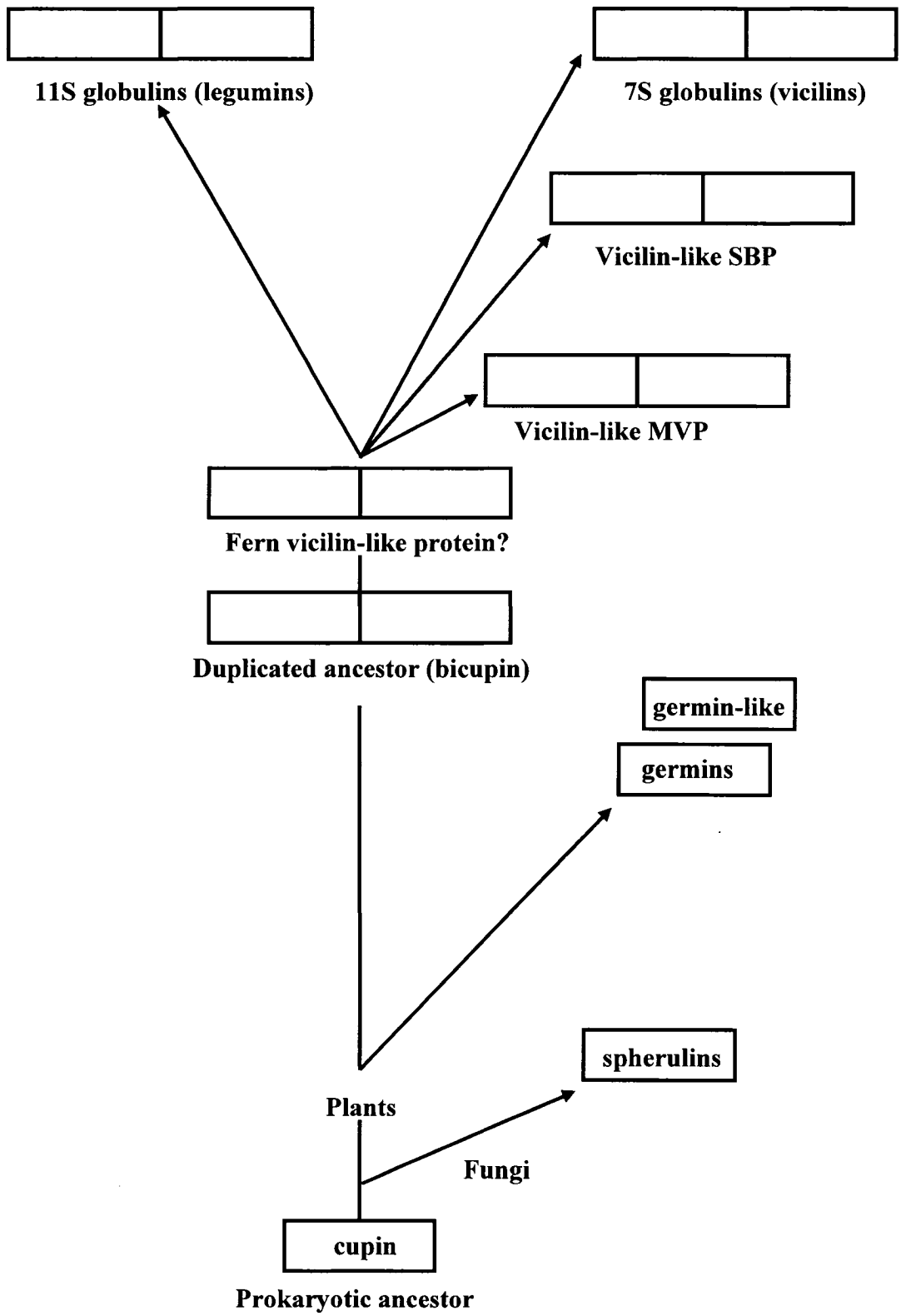
Globulin evolution within the cupin superfamily

Recognition of sequence similarity between the thermostable wheat protein germin and the *Physarum polycephalum* (slime mould) spherulin (18) marked the first alignment in the 7S/11S globulin evolutionary tree. At this point, the 11S and 7S globulins were already established to be related in sequence and structure (19, 20), but these were not added to the cupin tree until later (21). While different evolutionary models have attempted to explain the relationship between the 7S and 11S members of this cupin branch (12, 20, 22), the widely accepted duplication model is corroborated by structural data as well as the numerous related sequences some of which have provided missing evolutionary links (13, 14, 23).

The theorized evolutionary pathway of the globulins (13), shown in Figure 1.1, suggests that the common ancestral cupin belonged to a prokaryote and may have been involved in dessication or hydration processes. This protein evolved into the single domain spherulin proteins of fungi and into a separate branch of germin and germin-like plant proteins which benefit from the enzymatic and molecule-binding abilities afforded by the adaptability of the cupin structure (17). Sometime after this divergence event, an internal duplication resulted in an ancestral bicupin, perhaps best represented by the extant fern spore-specific protein, *Matteuccia struthiopteris* vicilin-like protein (MVP) (22). This protein is thought to have evolved along two separate pathways that lead to the emergence of the compact, desiccation tolerant 7S and 11S globulin proteins. Other divergence events along the 7S globulin pathway established vicilin-like sequences such as the non-storage sucrose binding protein (SBP) of soybean (24). This evolutionary model, based on available structure and sequence information, implies that the 7S and 11S globulin families acquired their storage functions independently, their common ancestor having had a non-storage role (25).

The 7S and 11S globulins can be classified into their separate groups based on sequence homology, their three-dimensional structures and immunological cross-reactivity. For the

Figure 1.1 Theorized evolutionary pathway of the 7S and 11S globulins. The prokaryotic cupin ancestor may have been involved in dessication or hydration processes. From this protein evolved the single domain spherulin proteins (fungi) and the germin and germin-like plant proteins (17). An internal duplication resulted in an ancestral fern vicilin-like bicupin, perhaps best represented today by the fern spore-specific vicilin-like protein (MVP) (22). Divergence after the duplication event, lead to the emergence of the 7S and 11S globulin seed storage proteins. Other divergence events along the 7S globulin pathway established vicilin-like sequences such as the non-storage sucrose binding protein (SBP) of soybean (24). This figure was adapted from Figure 7 (13), with permission from the publisher.


11S globulins (legumins)
7S globulins (vicilins)
Vicilin-like SBP
Vicilin-like MVP
Fern vicilin-like protein?
Duplicated ancestor (bicupin)
germin-like
germins
spherulins
Plants
Fungi
cupin
Prokaryotic ancestor

purposes of this study, the 7S globulins alone will be reviewed, though a brief description of the 11S globulins is provided below for the purpose of completeness.

The 11S globulins

Like their 7S relatives, the 11S globulins are synthesized as precursor proteins that mature along the secretory pathway. These generally undergo post-translational disulfide bond formation followed by cleavage between the N- and C-terminal cupin modules, generating monomers composed of disulfide-linked polypeptides of ~20kDa and ~40kDa. Thus, the 11S polypeptide pattern under reducing conditions is almost always different from that under non-reducing conditions (26, 27). The 11S globulins are stored in protein storage vacuoles (PSVs) where monomers associate to form large hexameric structures with M_r ~300,000-450,000 (26). Hexameric conformation is hypothesized to be the result of an interaction between two trimers that share the same general structure as the 7S globulins (28).

The 7S globulins

In contrast to their larger counterparts, the 7S globulins are homo or heterotrimeric complexes of M_r ~140,000-280,000 whose monomeric subunits consist of a single polypeptide chain (29, 30). Interdomain disulfide bond formation does not occur for 7S monomers, such that their polypeptide profiles under reducing and non-reducing conditions are indistinguishable (26).

While the mature 7S globulins can differ significantly in size within and between species, all are translated as preproproteins belonging to one of two size categories: large (generally 60-80kDa) and small (generally 45-50kDa) (31). These are illustrated in Figure 1.2. All storage globulins mature along the secretory pathway and thus have short hydrophobic N-terminal signal/leader peptides for translocation into the endoplasmic reticulum (ER). These peptides are cleaved off co-translationally within the ER lumen (30, 32). The large size of the 60-80kDa preproproteins is

Figure 1.2 Large versus small 7S globulins. The 7S globulins can be classified as either large or small (31). Both types of globulin are translated as preproproteins with N-terminal signal sequences (white) allowing for co-translational insertion into the endoplasmic reticulum (30, 32), and a core sequence (black), known as the vicilin domain. Structurally, the vicilin domain can be viewed as two modules that share a similar three-dimensional fold. This is further discussed in Figure 1.3. The large and small 7S globulins thus differ in their N-terminal segments (grey). The N-terminal segment in small 7S globulins is absent or very short; in large 7S globulins, it can range from 80 residues in maize (36) to 350 residues in winter squash (37). In some plants, the N-terminal segments are processed considerably such that the final products (monomers), which will associate to form the trimeric complex, are of similar size to the small 7S globulin monomers (39).

Large 7S globulin

Small 7S globulin

attributed to a hypervariable N-terminal segment which occurs between the leader peptide and the remainder of the molecule, which is absent or very short in small 7S globulins. High proportions of charged and acidic residues make up this hydrophilic N-terminal segment (31). Likely the vestige of an ancient transposable element (33-35), this domain can differ significantly in size from 80 residues in maize (36), to 183 residues in soybean (33), to 350 residues in winter squash (37). These large globulin segments share little or no homology between and within plant species, apart from short CXXXC motifs found in most, but not all N-terminal domains (31, 35, 38). The 7S globulins all share a common vicilin domain composed of two related modules (described below and in Figure 1.3).

The mature 7S subunits vary more significantly in size than do their preproproteins, owing to a series of post-translational modifications which often include signal sequence removal, N-terminal sequence cleavage and glycosylation. In some instances, the subunits are further digested into smaller peptides, but these remain associated, likely through non-covalent interactions within the globulin trimer (43).

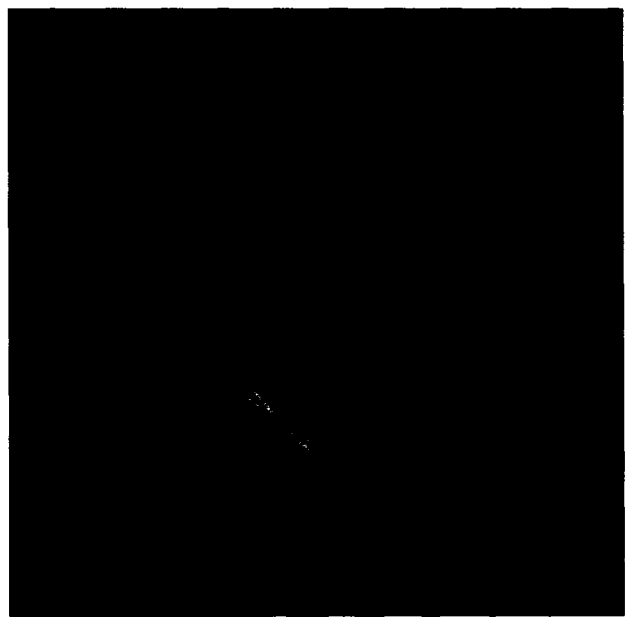
The 7S globulin structure

Determination of the three-dimensional (3D) structures for the model 7S globulins phaseolin (44), canavalin (40, 42), β-conglycinin β subunit (45) and β-conglycinin α' subunit (46) has provided important insight into the evolution of the globulins. The 7S globulin trimer (Figure 1.3) can be composed of identical or mixed subunits. These subunits, or monomers, are single polypeptide chains composed of two structurally-related domains termed the N- and C-terminal modules. The modules arose from a duplication event (Figure 1.1), which is reflected by weak sequence homology (20). Together, these modules related by a pseudo dyad axis, make up the bicupin vicilin domain, shown as the asymmetric unit (Figure 1.3). Each module is composed of a cupin-like β-barrel region, formed by 12-13 antiparallel β-strands, followed by a predominantly

Figure 1.3 The canonical structure of the 7S globulin. Ribbon diagrams of the three-dimensional structure of canavalin, the 7S globulin from jack bean, *Canavalia ensiformis* (40), were obtained from the Protein Data Bank (PDB), http://www.rcsb.org/pdb/Welcome.do, last viewed 13/6/2006, (PDB ID: 2CAV). Canavalin was the first 7S globulin to be crystallized and subjected to X-ray diffraction studies (41). High resolution structural data (40, 42) has shown that the native, or biological, molecule is composed of three monomers (asymmetric units). Trimeric structure is the result of the association between the alpha helical domains of the monomeric subunits. These subunits are related in structure by a pseudo dyad axis, reflecting an ancient sequence duplication. The N- and C-terminal modules that make up the asymmetric unit adopt comparable folds – each is composed of a β-barrel motif and an α-helical region – but share limited sequence identity. All 7S globulins are expected to exhibit similar three-dimensional structures (23).

Biological molecule

Asymmetric unit



alpha helical region (42, 44-46). When the 7S globulin sequences are aligned, and the three-dimensional structures are overlaid, it is apparent that the β-barrel regions are the most highly conserved, while regions outside are subject to more variability. Insertions and deletions relative to other 7S globulins map to variable loop regions within the 3D structure. These are outlined and further discussed by Lawrence, 1999 (23).

7S globulin synthesis and accumulation

The 7S globulins are synthesized on the rough endoplasmic reticulum (rER) (47, 48), with co-translational removal of their N-terminal signal peptide and deposition in the ER (30, 49). The ER is the site of N-linked co-translational glycosylation, partial processing of the oligosaccharides (50-52) and formation of the trimeric complex (43, 53, 54). Glycosylation is dependent on the consensus sequence: Asn-X-Thr/Ser (30). Following trimer formation, the globulins are transported to protein bodies via dense vesicles exiting the Golgi apparatus (32, 55, 56). Further post-translational modifications occur along this transport pathway. In the Golgi apparatus, additional oligosaccharide processing can take place (30, 51, 52, 57, 58), while endoproteolytic events occur in the protein bodies (43, 59, 60). From synthesis on the rER to maturation in the PSV, the 7S globulins can require a few hours or up to 72 hours for completion (43, 61).

The co- and post-translational modifications, including glycosylation and proteolytic processing, observed for some 7S globulins appear to be completely absent in others (Table 1.1). Thus it is concluded that neither of these types of modifications is required for proper processing, transport or storage of the 7S globulins. For those where processing does occur, for example the pea vicilins, proteolytic events are not complete. Pea vicilins are present at molecular masses (M_r) suggesting that intact, partially and fully cleaved polypeptides co-exist in mature 7S globulin trimers, held together by non-covalent forces (43). Differential glycosylation within a single

Table 1.1 7S globulins in selected dicotyledonous and monocotyledonous species. The 7S globulin proteins of selected dicotyledonous species (pea, soybean and common bean) and monocotyledonous species (maize, barley, rice, wheat) are summarized. When information is available, the globulin name, classification as either a large or small 7S globulin, and observed M_r are provided. Glycosylation status and the number of genes encoding these proteins are also indicated. Post-translational cleavage events and glycan modifications are described under the 'processing' heading. N-terminal processing refers to proteolytic modifications of the N-terminal segment of large 7S globulins - this excludes signal sequence removal, which is expected to occur for all 7S globulins. Internal processing indicates proteolytic events within the vicilin domain. N/A: information is not available.

Plant	7S globulin	Large/Small globulin	Observed M_r (kDa)	Glycosylation	Gene number	Processing
Pea	Vicilin - three types (47, 50 and 68kDa)	Small	68, 50, 47, 33, 30, 25, 19, 18, 16, 14, 13, 12 (72-74)	Mixed: unglycosylated & glycosylated at one position (48, 73, 74)	Up to 24 genes (60, 67-71)	Internal (74)
	Convicilin	Large	~70 (29)	No (67)	2 genes (35, 38)	No evidence to suggest processing occurs (67)
Soybean	β-Conglycinin (α and α' subunits)	Large	~72-83 (75-77)	At two positions (78, 79)	~3 genes (80)	Hypothetical alteration to oligosaccharides and N-termini (51, 58)
	β-Conglycinin (β subunit)	Small	~52-54 (75-77)	At one position (78, 79)	~6 genes (80) with other genes resembling β, α and α'	Hypothetical alteration to oligosaccharides (51, 58) no other modifications proven (45)
Common bean	Phaseolin - two types (α and β)	Small	51-53; 47-48, 43-46 (50, 62, 63)	Mixed: unglycosylated & glycosylated at one or two positions (50, 52, 62, 63, 81)	~6-8 genes (82)	Alteration to oligosaccharides, not polypeptides (30, 52, 58)

Plant	7S globulin	Large/Small globulin	Observed M_r (kDa)	Glycosylation	Gene number	Processing
Maize	Glb1	Large	60-70 (61, 83)	Yes (61, 83)	1 (36, 84)	N-terminal (61, 83)
	Glb2	Small	45, 27 (83)	Yes (83)	1 (83, 85)	Possible internal (83, 85)
Barley	Beg1	Large	70,50,40,25 (39, 86)	N/A	1 (39)	N-terminal and internal (39, 86)
	Beg2	N/A	N/A	N/A	N/A	N/A
Rice	Reg1	N/A	49, 35 (87)	N/A	N/A	Internal (87)
	Reg2	Small	46 (87)	N/A	1 (87)	N/A
Wheat	Gbl1	Large	N/A	N/A	6 (88)	N/A

polypeptide species indicates that glycosylation is not a prerequisite for proper folding, transport or accumulation (50, 62-64) . Nonetheless, for storage globulins where glycosylation does occur, it appears that complete lack of glycosylation can affect the extent of protein deposition to protein storage bodies. Thus glycosylation in some globulins may have a stabilising effect on the trimer (65). The various post-translational modifications (PTMs) that can occur for 7S globulins often result in complex expression patterns of multiple iso/glycoforms. These are further complicated when the 7S globulins belong to multigene families.

What do we know about the 7S globulins?

Our current knowledge of the 7S globulin family owes to thorough studies performed on a limited number of legume and crop plants. The most meticulously studied 7S globulins occur in pea, common bean, soybean and maize. The 7S globulins in each of these species will be briefly described below. Table 1.1 is a summary of 7S globulins from a select group of dicotyledonous and monocotyledonous species.

Legume 7S globulins

In the legume family, the globulins are generally the dominant seed storage proteins, accumulating primarily in the cotyledon (66). Generally, one globulin family predominates in the seed: either the 7S or the 11S globulins.

Pea (vicilin and convicilin)

The pea globulin fraction consists of 11S globulins (legumins) and of 7S globulins of the large (convicilin) and small (vicilin) size classes. The ratio of legumin to convicilin/vicilin varies among genotypes from 4 to 0.5 (67). The pea vicilins are encoded by up to 24 genes forming large multigene families (60, 67-71). These proteins are synthesized as preprovicilins that undergo various post-translational modifications including partial glycosylation (48, 73, 74) and

limited endoproteolytic events (72-74) giving rise to a heterogeneous polypeptide population with apparent molecular masses varying from 68kDa to 12kDa. The resultant 30-33kDa proteins are the most abundant (72). These PTMs are outlined in Figure 1.4, panel A. In contrast, the convicilins appear only as ~70kDa proteins (29) and are thought to be encoded by two genes (35, 38). The convicilins appear to avoid the post-translational modifications common to vicilin and thus are likely to be packaged in trimeric form as large globulins (29, 67).

Common bean (phaseolin)

The major reserves of common bean are the phaseolin proteins, small-sized 7S globulins, which accrue at M_r 51-53; 47-48 and 43-46kDa (50, 62, 63). This protein accumulation pattern is attributed to the six to eight genes that encode the phaseolin proteins (82) and to various post-translational glycosylation events. Phaseolin has two potential N-linked glycosylation sites of the pattern Asn-X-Thr/Ser (89) and differential glycosylation is observed for each of these sites (50, 52, 62, 63, 81). In contrast to pea vicilins, no known proteolytic processing occurs after co-translational removal of the signal peptide from nascent phaseolin (30, 52, 58).

Soybean (β-conglycinin)

The soybean β-conglycinin 7S globulins represent ~25% of the seed protein and are usually found in similar (or slightly lesser) proportions to the 11S glycinins. Like in pea, the 7S globulin fraction of soybean is encoded by multigene families of both large and small 7S globulins. At least 15 β-conglycinin transcriptional products have been identified, three of which encode large (α or α') subunits, while at least six encode small 7S globulins (β-subunits) (78, 80). It has been shown in soybean that these different-sized subunits, representing large and small 7S globulins, can associate with one another in the trimeric complex (92-94). The large α/α'-subunits have

Figure 1.4 Post-translational modifications of pea vicilin and maize globulin1. Cleavage sites are indicated by arrowheads; glycosylation, by an asterisk. Black rectangles represent the vicilin domain of the 7S globulins. Diagrams are not drawn to scale.

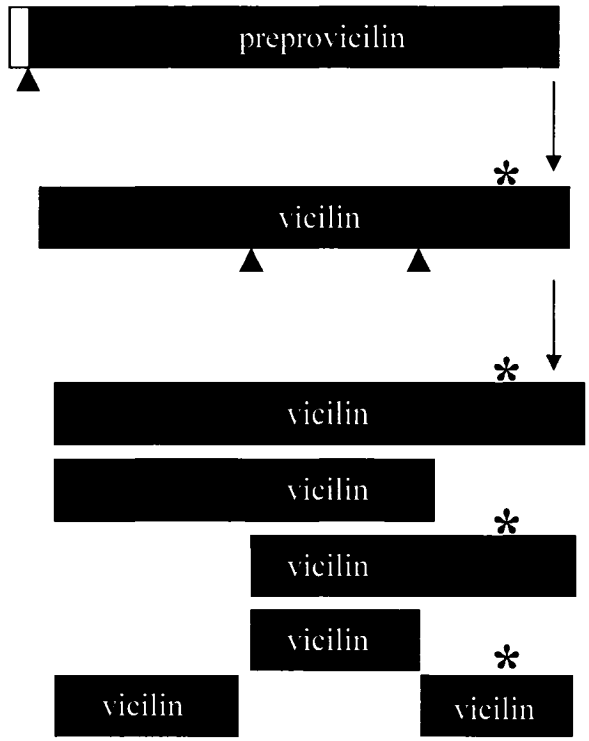
Panel **A** shows the post-translational modifications (PTMs) giving rise to the mature pea vicilin polypeptides. Co-translational removal of the signal peptide (white), deposition in the endoplasmic reticulum (ER) and glycosylation are common events for 7S globulins. Once transferred to the protein bodies, the pea vicilins undergo a series of limited endoproteolytic processing (EPP) events resulting in a number of polypeptides (48, 73, 74).

In panel **B**, maize globulin 1, Glb1, is translated as a preproprotein, preproGlb1', which undergoes a series of PTMs. A leader sequence (white), which signals for insertion into the ER, is co-translationally removed. Loss of the signal peptide results in proGlb1' (61). This immature form of Glb1 can be detected in the ER but not in the protein storage vacuoles (PSVs) of developing embryos suggesting that processing of proGlb1' into Glb1' occurs in the Golgi apparatus (90). ProGlb1' has a larger M_r than that predicted for the primary translation product suggesting that co-translational glycosylation occurs in the ER. The mature proteins Glb1' and Glb1 are glycoproteins, but their nature is undetermined (61, 83). The proGlb1' intermediate is processed to generate Glb1' by removal of the N-terminal propeptide (light grey). The final step in the maturation of Glb1 is a cleavage event, likely occurring in the protein storage bodies, catalyzed by the gene *mep* (modifier of embryo protein), which removes the N-terminal sequence (mid-grey) of Glb1' (83, 84). This reaction is not complete as both Glb1' and Glb1 can be detected in mature maize seeds (61, 83, 91). The remaining dark grey fragment represents the N-terminal segment of globulin1, not removed during the PTMs. Panel B was adapted from Figure 3 (90), with permission from the publisher.

A

Co-translational cleavage of signal peptide, transport into ER and glycosylation

Limited EPP events within protein bodies generate multiple products of varying sizes

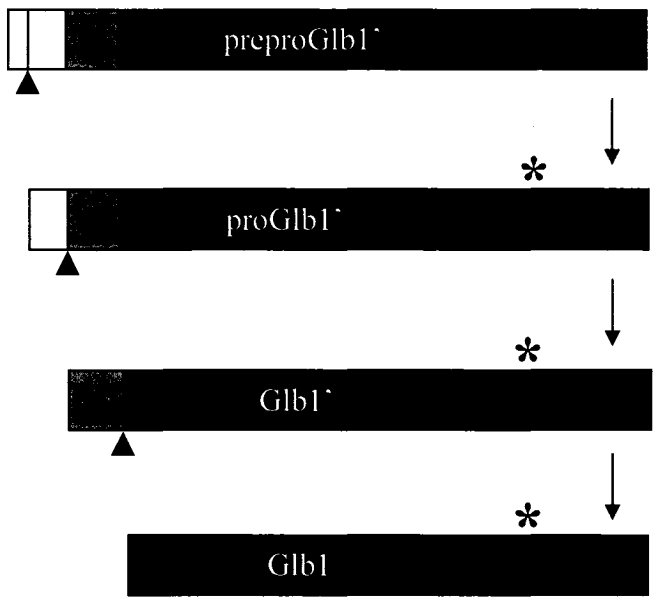


B

Co-translational cleavage of signal peptide, transport into ER and glycosylation

Cleavage of N-terminal propeptide

mep-catalyzed cleavage of N-terminal peptide to yield mature Glb1



mature M_r of 72-83kDa while the β-subunits are ~ 53kDa (75-77). The α/α'-subunits also differ from the β-subunits in their glycosylation pattern: α/α' can be N-glycosylated twice whereas β has one N-linked glycan (78, 79). Like phaseolin and pea vicilins, the β-conglycinin genes contain 5 introns and 6 exons (80, 89, 95).

Cereal 7S globulins

The endosperm storage proteins account for the majority of the cereal seed reserve. In this respect, the monocotyledonous cereals differ from dicotyledonous legumes. The two groups also differ in the predominant types of storage proteins they accumulate. Whereas the legumes store globulins, the cereals generally store prolamins or glutelins (96). Oat is an exception, chiefly amassing 11S globulins (97-103). While the rice storage proteins are considered glutelins (104) on the basis of solubility, some are structurally similar (10) and antigenically-related (105) to the 11S globulins.

Thus, the cereal 7S globulins are considered secondary reserve proteins. They appear to be almost exclusively expressed in the embryo or aleurone layer (39, 87, 91, 106-108) and are regulated at the transcriptional level by phytohormone abscisic acid (ABA) (39, 87, 109-111). The rapid degradation of maize embryo globulins, which begins within 24 hours of imbibition, suggests that these may provide immediate sources of nitrogen to the germinating embryo, prior to mobilization of endosperm proteins (106, 110). Preliminary characterization of this class of proteins has been performed, yet the cereal 7S globulins remain a somewhat obscure group. Much of our current understanding hinges on the data collected from legume studies and from the detailed study of the maize 7S globulins.

Maize 7S globulins

Maize globulin1 (Glb1) is a major embryonal protein and is encoded by a single copy gene for which null alleles have been found (36, 84, 112). For these reasons, Glb1 is the only cereal 7S globulin whose PTM pattern has been unambiguously deciphered (61, 83, 84, 106, 113). Maize Glb1, a large 7S globulin, is the product of a series of PTMs that occur along the secretory pathway before the mature proteins are deposited in PSVs. The maturation of Glb1 is depicted in Figure 1.4, panel B. In contrast to pea vicilins, Glb1 does not undergo endoproteolytic processing (EPP) events within the vicilin domain; rather, these are restricted to the N-terminal segment. The final cleavage of Glb1' to Glb1 in the PSV is not complete, such that both proteins accumulate within the seed (83, 91). A second 7S globulin has also been reported in maize embryos: globulin2 (Glb2). While Glb2 is confirmed to mature along the secretory pathway, its processing is less well understood (83, 85, 91). Based on the size of its N-terminus in relation to the vicilin domain, Glb2 is a small 7S globulin, exhibiting ~28% identity over its length with Glb1 (85, 90).

Barley and Rice

The two 7S globulin gene families in maize, *Glb1* and *Glb2*, have also been identified in barley and rice. The homologues have been named according to their maize counterparts: barley embryo globulins (*Beg1* and *Beg2*) and rice embryo globulins (*Reg1* and *Reg2*). In barley, Beg1 has been cloned and identified at the protein level; Beg2 has only been detected at the transcript level (39, 86, 114). Both Reg1 and Reg2 transcripts and proteins have been confirmed in the rice embryo; only *Reg2* has been cloned (87).

The globulin1 family

Past studies of the Glb1, Beg1 and Reg1 family indicate that each is a single copy gene (36, 39, 84, 87, 112). However, the recent report of two *Beg1* isoforms suggests that *Beg1* may belong to

a small multigene family (114). Like Glb1, Beg1 is a large 7S globulin, possessing an N-terminal segment. Glb1 and Beg1 share ~63% identity, and are thus more closely related than Glb1 and Glb2 (90). Despite being strongly related to Glb1, Beg1 is thought to undergo different processing events along the pathway to maturation. While Glb1 experiences trimming for the removal of a portion of its N-terminal segment, Beg 1 (and probably Reg1) may be subject to additional processing, including almost complete removal of the N-terminal segment and limited proteolytic cleavage events within the vicilin domain (39, 86, 87). The barley embryo globulin profile consists of major bands at 70, 50, 40 and 25kDa. Each of these has been proposed to be a differential processing product of the 72kDa Beg1 preproprotein (39) although to date, it has not been shown that the 50, 40 and 25kDa proteins derive from a 70kDa holoprotein.

The globulin2 family

This cereal globulin family is not nearly as well characterized as the Glb1 group. The available cDNA sequences suggest that these are small 7S globulins, lacking a significant N-terminal segment (85, 87). The Glb2 homologues may undergo processing similar to the pea vicilins (Figure 1.4, panel A) but this has not been investigated (83, 87). To date, protein products of this family have only been identified in maize and rice.

The other cereals: oat, rye, wheat

The identification of the Glb1 and Glb2 families in three separate cereal species suggests that other cereals have similar globulin profiles. Immunologically, oat, rye, wheat and barley all possess major protein bands of ~50-60kDa that react with antisera raised against the oat 7S globulins and the pea vicilins. These proteins differ slightly in size between cereals and are accompanied by larger and smaller immunoreactive species. The relationship between these polypeptides has not been deciphered (101, 105, 108).

In wheat, a partial cDNA with homology to *Glb1* has been identified (109, 115), and a genomic sequence (*Gbl1*) having strong identity with *Glb1* has been deposited to Genbank (accession: M81719). However, the origin of the *Gbl1* sequence is in doubt because its protein coding sequence is identical to the *Beg1* cDNA and does not correspond to the partial cDNA (90). *Gbl1* is expected to belong to a small multigene family in wheat (88). Mass spectrometry and N-terminal sequencing have identified Glb1/Beg1-like proteins in wheat (116-118). Glb2-like sequences have not been reported.

The globulins: storage and beyond

While the 7S globulins have been studied for almost a century (119), prior investigations focussed on the biosynthesis of the seed storage proteins, their structure, their evolution and the multigene families that gave rise to their protein profiles. Until the last decade, the only role attributed to the globulins was that of a carbon and nitrogen source during seedling development. As research into the role of these storage proteins advances, new functions are being proposed.

Of particular interest is the assignment of function to peptides derived from the N-terminal segment of large 7S globulins. As previously discussed, some large 7S globulins are processed minimally such that the mature globulin has retained a significant portion of its N-terminal segment. It is generally accepted that this domain is retained in the trimeric complex, although it is not understood how the polar, highly variable region is accommodated in the tertiary and quaternary folds (46). In other globulins, the N-terminal segment is partially or completely proteolytically processed. Some of these processing events may take place prior to trimer assembly; others may occur after, being retained in the trimeric complex. The fate and purpose of the resultant polypeptides is uncertain, although current studies are providing new insights. Vacuolar processing of a large 7S globulin-like protein from winter squash results in peptides that exhibit *in vitro* trypsin inhibition, perhaps as means of defence against animals (37). Peptides

derived from the N-terminal segments of macadamia nut and cotton 7S globulins also show *in vitro* evidence of protective abilities in antimicrobial and antifungal trials, respectively. The N-terminal sequences may provide additional protection to the seed during the critical period of germination (120, 121).

The compact structure of the bicupin fold is thought to have been a driving force in the acquisition of storage function among the seed 7S and 11S globulins. The tertiary and quaternary structures adopted by the globulins allow the proteins to tolerate elevated temperatures and desiccation, two important properties among seed storage proteins. These configurations also help protect the proteins from premature and uncontrolled proteolysis (25). The structure of some 7S globulins may have also evolved as a defence mechanism. Some 7S globulins for example, have the ability to bind chitin matrices. In this capacity, they are shown to interfere with the larval development of bruchids and spore development of certain fungi, conferring insect and fungal resistance (122-124). Moreover, the Glb1 and Glb2 proteins of maize kernels are associated with constitutive aflatoxin resistance, through a yet undetermined mechanism (125).

The structure adopted by the globulins may also be having adverse effects on human health. Heat tolerance and proteolytic resistance allow some of these proteins to evade the degradative effects of food treatments and digestive processes, reaching the gut intact or partially folded (126-132). As a result, antigenic epitopes are preserved and cross the epithelial barrier lining the intestine where they may elicit adverse immune responses. The induction of food allergies and of autoimmune diseases, such as celiac sprue by wheat gliadins, is proposed to occur through this mechanism (133). The 7S globulins of a number of seeds, including soybean (β-conglycinin and Gly m Bd 28K) (134, 135), peanut (Ara h 1) (136, 137), walnut (Jug r 2) (138), lentil (Len c 1) (139), cashew (Ana o 1) (140), sesame (Ses i 3) (141) and pea vicilin (142) have been found to harbour IgE-binding epitopes.

Rationale and objectives

The wheat globulins are a poorly characterized group, a number of factors having lead to their neglect. Firstly, they make up a proportionally small percentage of the grain's protein content; their only identified role at present is that of secondary reserve proteins. In contrast, the thoroughly studied gliadins and glutenins are the major storage proteins in wheat (96). These proteins make up the gluten fraction, bestowing upon wheat its bread-making quality, a unique and industrially valuable property (143). The implication of a number of gliadins and glutenins as culprits in the development of celiac disease has also been a factor contributing to their detailed study (133). Under the shadow of the gluten proteins, the globulins have received little attention. Thus, our current understanding of the wheat 7S globulins, as previously discussed, is limited to a handful of studies, an unresolved genomic sequence and inferences from other cereal globulin research.

Recently, a wheat cDNA clone, *WP5212*, was isolated from a developing wheat seed expression library and was found to be associated with the development of type 1 diabetes (T1D) in diabetes-prone rats. Sequencing of the *WP5212* clone, revealed a protein coding sequence with 90% identity to *Gbl1/Beg1* (144, 145). This investigation has brought renewed interest to the study of the 7S globulins of wheat.

The current study was a preliminary look at the wheat WP5212 antigenically-related proteins using a proteomic approach. The purpose of this investigation was to assess the complexity of the WP5212-related fraction, to identify major components of this fraction, and to determine their relatedness.

The specific goals of this thesis were:

1) The separation of the 7S globulins from the wheat endosperm and embryo-enriched fractions by two-dimensional gel electrophoresis (2DE) and identification of WP5212 antigenically-related proteins.
2) Identification of abundant WP5212-related polypeptides by mass spectrometric analysis and N-terminal sequencing.
3) Determination of the relatedness of the identified globulin components, by sequence analysis and review of the current literature.

Effort was also invested in expressing the WP5212 sequence in transgenic tobacco. The production of WP5212 in a transgenic plant could facilitate the obtainment of sufficient quantities of protein for feeding trials. This work is shown in Appendix B.

For the purposes of this thesis, "salt-soluble fraction" and "globulins" are two terms that will be used interchangeably to describe the proteins from the wheat seed that are soluble in 1M NaCl. Glb1-like and Glb2-like will be used to describe wheat proteins with strong identity to maize Glb1 and Glb2, respectively. WP5212 is a Glb1-like protein.

Chapter II – Materials and Methods

2.1 Materials

Plant material

Wheat (*Triticum aestivum* cultivar AC Barrie) seeds were provided by Agriculture and Agri-Food Canada, Indian Head Research Farm and Seed Increase Unit (Indian Head, SK).

Reagents

Pre-immune serum and polyclonal rabbit anti-WP5212 serum, raised against two synthetic peptides corresponding to predicted B-cell epitopes from the WP5212 translated sequence (145) were donated by Dr. F. Scott (OHRI, Ottawa). Biotin-SP-conjugated AffiniPure goat anti-rabbit IgG (H+L) and biotin-SP-AffiniPure donkey anti-mouse IgG (H+L) secondary antibodies were purchased from Jackson ImmunoResearch Laboratories, Inc (West Grove, PA). The affinity purified goat anti-biotin-Horse Radish Peroxidase (HRP)-conjugated antibody was obtained from Cell Signaling Technology, Inc (Danvers, MA). The Prestained Benchmark (protein ladder) was purchased from Invitrogen Life Technologies (Burlington, ON), while the Biotinylated Protein Ladder Detection Pack was sold by Cell Signaling Technology, Inc (Danvers, MA). The Precision Plus protein standard is a product of Bio-Rad (Mississauga, ON).

2.2 Methods

Wheat seed protein extraction and sample preparation

For each extraction, 4 g of AC Barrie wheat seeds or 50-100 embryos or endosperm were employed. Seeds were initially ground in a coffee grinder. The crushed seeds were milled by hand using a mortar and pestle under liquid nitrogen until a powder was obtained. Wheat powder was mixed with 10 volumes of fresh ice-cold acid denaturing solution (10% (w/v) trichloroacetic acid (TCA) and 0.05% (w/v) dithiothreitol (DTT) in acetone). Samples were stirred at 4°C for 1

h then left at -20°C overnight. The resulting suspension was centrifuged for 30 min at 35,000 x *g* at 4°C (Beckman Avanti J-25, rotor JA 25.50). The supernatant was decanted and the pellet resuspended in ice-cold acetone containing 0.05% (w/v) DTT. The mixture was extracted by incubation for 1 h at -20°C. The suspension was centrifuged for 20 min at 35,000 x *g* at 4°C; the supernatant decanted and the pellet dried on ice. Dried powder was extracted with 15 ml of 1 M NaCl solution (1.0 M NaCl; 0.05 M Tris; pH 8.0). The mixture was stirred for 1 h at room temperature and centrifuged (27,000 x *g* for 30 min at 22°C). Supernatants were collected and the extraction with 1 M NaCl solution was repeated. The supernatants were pooled and dialysed against 5 changes of water for 24 h at 4°C. Precipitates were collected by centrifugation at 35,000 x *g* at 4°C for 45 min. Pellets were drained and extracted for either one-dimensional (1D) or two-dimensional (2D) protein separation.

Sample preparation for 1D separation

Pellets were resuspended in 1 M NaCl solution. Samples were centrifuged at 20,000 x *g* for 1 h at 22°C and quantified by the bicinchoninic acid (BCA) assay (146). Supernatants were aliquoted and stored at -80°C until use.

Sample preparation for 2D separation

Proteins from pellets were extracted with 1.5 – 3 ml rehydration buffer (8M/2M deionized urea/thiourea, 2% 3-([3-cholamidopropyl]dimethylamino)-1-propanesulfonate (CHAPS), 50 mM DTT, 0.0005% bromophenol blue) containing pH 3-10 ampholytes (Bio-Rad). Following sonication, samples were centrifuged at 200,000 x *g* for 1 h at 22°C in the Beckman TL-100 Ultracentrifuge. Supernatants were collected, quantified by Bradford method (147), aliquoted and stored at -80°C until use. Throughout the 2D separation process, care was taken not to heat the urea/thiourea-containing solutions above 30°C to avoid carbamylation of amino groups which can lead to artifactual spot heterogeneity (148).

One-dimensional sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) and western analysis

SDS-PAGE

Protein extracts were fractionated under SDS-PAGE reducing conditions. Protein samples were combined in a 1:1 (v/v) ratio with 2X sample buffer (4% SDS, 20% glycerol, 0.12 M Tris (pH 6.8), 10% (v/v) β-mercaptoethanol, 0.01% bromophenol blue). Samples were boiled for 5-10 min then centrifuged at 20,000 x *g* for 10 min. All samples were loaded on discontinuous (5% stacking, 10%-12% resolving) 1.5 mm SDS polyacrylamide gels (149). The SDS running buffer, adjusted to pH 8.3 with NaOH, consisted of 25 mM Tris base, 0.19 M glycine and 0.1% SDS. Electrophoresis was performed in the Mini-PROTEAN 3 System (Bio-Rad) at 100-150 V (400 mA) until dye front reached the bottom of the gel. Gels were either stained with Coomassie Brilliant Blue (CBB) R-250 or were used for western blot analysis.

Western blot analysis

Proteins from SDS-PAGE were transferred under semi-dry conditions by means of the Trans-Blot SD Semi-Dry Transfer Cell (Bio-Rad). Gels were rinsed 5-10 min in semi-dry transfer buffer (24 mM Tris, 192 mM glycine and 15% methanol) and transferred to nitrocellulose membranes (Bio-Rad) also soaked in transfer buffer. Electroblotting proceeded for 1-1.5 h at 11 V (400 mA). Once complete, the transfer was verified by staining membranes with Ponceau S (0.2% (w/v) Ponceau S in 1% glacial acetic acid) followed by several washes of ddH_2O. Prior to immunoblotting, membranes were destained with TBST buffer (10 mM Tris-HCl (pH 7.3), 0.1 M NaCl and 0.5% Tween-20) and were incubated 30 min with shaking, in 5% skim milk powder in TBST buffer at room temperature. Membranes were incubated with the primary antibody (diluted 1:10,000 in 5% skim milk powder in TBST and 0.05% NaN_3) and left overnight at 4°C with gentle rocking. Next day, membranes were washed 4 x 5 min in TBST, and incubated 1 h in the secondary, biotinylated, antibody (1:100,000). This incubation was followed by another set

of washes in TBST and a final, tertiary (anti-biotin-HRP-linked, diluted 1:2,000) antibody incubation for 1 h. Following 4 x 5 min washes, membranes were treated with ECL™ Western Blotting Detection Reagents (Amersham Biosciences, Piscataway, NJ) for 1 min then exposed to Kodak BioMax Light Film (Fisher Scientific, Nepean, ON). Film was developed by the Kodak X-OMAT 2000A Processor.

Two-dimensional gel electrophoresis

Proteomic analysis of the AC Barrie salt-soluble proteins was conducted by 2DE. The proteins were separated in the first dimension by isoelectric focussing (IEF) according to their isoelectric points (pI). Passive rehydration was used to load extracts on linear IPG (immobilized pH gradient) 7 cm strips with pH ranges 3-10 and 7-10 (Bio-Rad). A protein load of 100-150 μg was applied to strips for staining and immunoblot applications; 250 μg for mass spectrometry and N-terminal sequencing applications. Strips were covered in mineral oil and rehydrated overnight in a reswelling tray in 125-150 μl volumes of sample diluted in rehydration buffer. For strips in the pH ranges 3-10, DTT was used as a reducing agent; tributylphosphine (TBP) was used for strips in the pH range of 7-10. Rehydrated strips were focussed in the PROTEAN IEF Cell (Bio-Rad) according to the following program: step1, 250 V (30 min), linear ramp; step2, 4000 V (2 h), linear ramp; step3, 4000 V (10,000 Vh), rapid ramp. Immediately following isoelectric focussing, the strips were either stored at -80°C for future use or equilibrated prior to running the second dimension. Strips were thawed, if necessary, and reduced for 15 min in 2 ml of equilibration buffer (6 M urea, 2% SDS, 20% glycerol, 0.375 M Tris-HCl, pH 8.8 containing 2% (wt/vol) DTT or 5 mM TBP). A second alkylation step was performed for 15 min in the same equilibration buffer with the exception that 2.5% (w/v) iodoacetamide was substituted for reducing agents DTT or TBP. Overlay agarose (Bio-Rad) was used to seal the equilibrated strips at the top of 1 mm vertical resolving gels (10% or 12% polyacrylamide). Second dimension SDS-PAGE was performed using the Bio-Rad MiniProteanIII at 100 V (400 mA); gels were

terminated when the bromophenol dye front had reached the end of the gel. After running the second dimension, gels were used for multiple purposes: (1) staining with CBB R-250, (2) semi-dry transfer for western blot analysis (as previously described), (3) mass spectrometry analysis, (4) N-terminal sequencing on polyvinylidene difluoride (PVDF) membranes.

Liquid chromatography tandem mass spectrometry (LC-MS/MS)

Gels were carefully manipulated under a laminar flow hood at all times to avoid keratin contamination. Following separation in the second dimension, gels were stained with Bio-Safe Coomassie (Bio-Rad) according to manufacturer's instructions and destained with multiple changes of ddH_2O. Selected spots were excised from the gel by means of a clean scalpel blade and sent to the Ottawa Institute of Systems Biology (OISB) for mass spectrometric analysis. Proteins were digested in-gel with chymotrypsin as previously described (150). Tryptic digest was avoided because of the high frequency of arginine and lysine residues in the one reported wheat globulin sequence (90). Peptides were separated by liquid chromatography on an Agilent 1100 Series HPLC System (Agilent Technologies, Palo Alto, CA) and applied by electrospray to a QSTAR Pulsar quadrupole-TOF mass spectrometer (ABI/MDS Sciex, Concord, ON) as described (151). Resulting peptide masses were used to interrogate the NCBI (National Center for Biotechnology Information) database using the Mascot search engine (152). Fixed modifications were set for carbamidomethyl (C) and variable modifications for oxidation (M). Three missed cleavages were allowed. Peptide and MS/MS mass tolerances permitted were ±100 ppm and 0.2 Da, respectively.

Protein sequences and alignments

The wheat data set available to Mascot through NCBI is limited. Therefore, globulin sequences derived from wheat expressed sequence tag (EST) libraries, were also interrogated. The *WP5212* coding sequence (cds) (145) was used as the query sequence in Basic Local Alignment Search

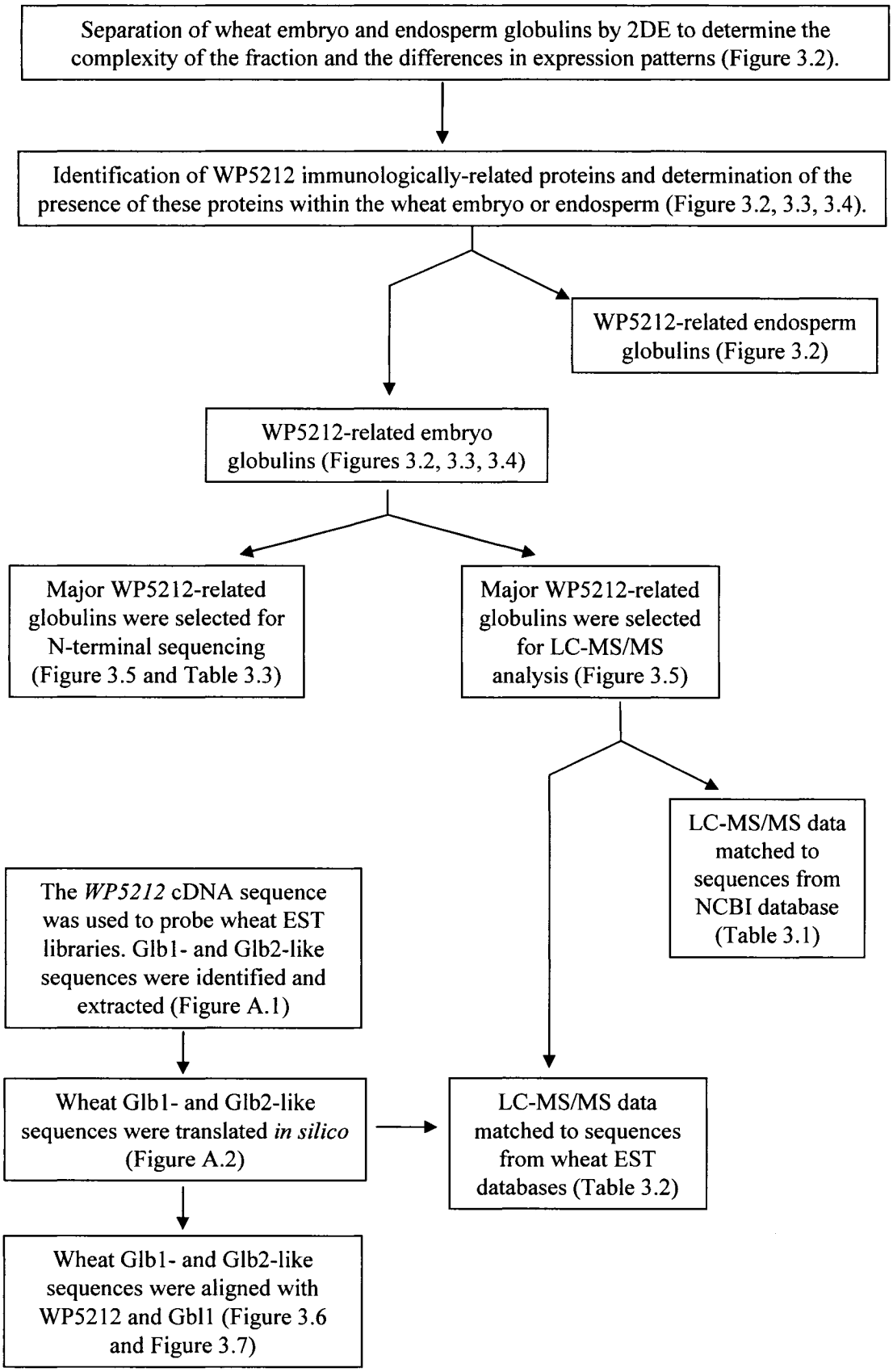
Tool (BLAST) searches, performed under default conditions. Two wheat genome databases were searched for *WP5212*-related sequences: TIGR (The Institute for Genomic Research, http://www.tigr.org/tdb/e2k1/tae1/, Wheat Gene Index, last viewed 31/5/06) and GG2.0 (GrainGenes: A Database for Triticeae and Avena, http://wheat.pw.usda.gov/GG2/blast.shtml, last viewed 31/5/06). Sequences homologous to the other known cereal 7S globulin, Glb2 (85) were also extracted from the TIGR database. Contigs obtained from these searches are reported in Figure A.1. Glb1- and Glb2-related sequences were analyzed using the NCBI open reading frame (ORF) finder (http://www.ncbi.nih.gov/gorf/gorf.html, last viewed 31/5/06) to determine protein sequence (Figure A.2). To assess the relationship of the globulin sequences, these were aligned using ClustalX (153) and manually adjusted using Genedoc (154). All amino acids are reported according to the standard one-letter code (155).

N-terminal sequencing

Gels used for N-terminal sequencing were treated for wet transfer to PVDF membranes. Briefly, the gels were soaked in CAPS (3-(cyclohexylamino)-1-propanesulfonic acid) electroblotting buffer (10 mM CAPS/NaOH pH 11 and 10% methanol) for 5 min. Sequiblot PVDF (Bio-Rad) membranes were wet in 100% methanol then soaked in CAPS electroblotting buffer along with Whatman sheets and transblot sponges. The blotting sandwich was assembled and run in the Mini Trans-Blot Cell (Bio-Rad) at 50 V (170 mA) for 45 min. After disassembly, the membranes were rinsed thoroughly with ddH_2O, saturated in 100% methanol and stained for a minimum of 1 min in amido black stain (0.1% amido black, 1% acetic acid and 40% methanol). Membranes were rinsed in multiple changes of ddH_2O and air dried before excision of spots. Spots of interest were sent to the University of Texas Medical Branch (UTMB) Biomolecular Resource Facility Core for protein N-terminal sequencing by Edman degradation, using the 494 cLC PROCISE Sequencing System (Applied Biosystems; Foster City, CA) for high sensitivity (femtomole quantities) protein sequencing.

Figure 2.1 is a summary of the techniques and steps outlined in chapter 2 and the results presented in chapter 3.

Figure 2.1 Summary of the techniques used to identify and characterize the major wheat 7S globulins. A flow chart format summarizes the techniques and procedures followed throughout this study, which are described in Chapter 2 (Materials and Methods) and Chapter 3 (Results). Reference is made to the figures representing each stage of the study.


Separation of wheat embryo and endosperm globulins by 2DE to determine the complexity of the fraction and the differences in expression patterns (Figure 3.2).
Identification of WP5212 immunologically-related proteins and determination of the presence of these proteins within the wheat embryo or endosperm (Figure 3.2, 3.3, 3.4).
WP5212-related endosperm globulins (Figure 3.2)
WP5212-related embryo globulins (Figures 3.2, 3.3, 3.4)
Major WP5212-related globulins were selected for N-terminal sequencing (Figure 3.5 and Table 3.3)
Major WP5212-related globulins were selected for LC-MS/MS analysis (Figure 3.5)
LC-MS/MS data matched to sequences from NCBI database (Table 3.1)
The *WP5212* cDNA sequence was used to probe wheat EST libraries. Glb1- and Glb2-like sequences were identified and extracted (Figure A.1)
Wheat Glb1- and Glb2-like sequences were translated *in silico* (Figure A.2)
LC-MS/MS data matched to sequences from wheat EST databases (Table 3.2)
Wheat Glb1- and Glb2-like sequences were aligned with WP5212 and Gbl1 (Figure 3.6 and Figure 3.7)

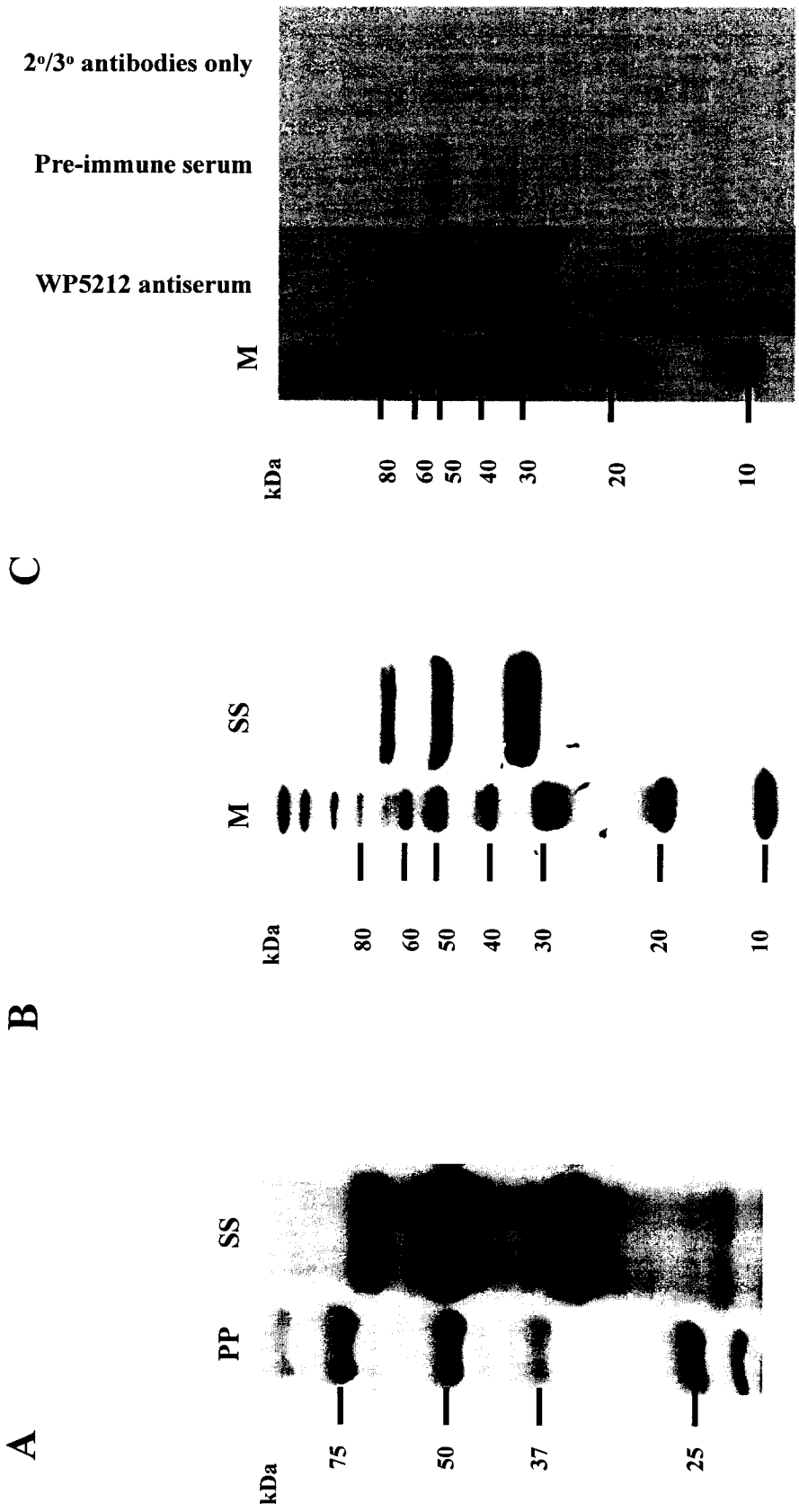
Chapter III – Results

Wheat globulin cDNA *WP5212* (145) is homologous to the 7S globulins, a group of salt-soluble storage proteins, which accumulate in the seed during development. WP5212 has been found to have 80% identity with the only registered 7S globulin from wheat, Gbl1 (Genbank accession AAA34269.1) (144). Few studies have sought to characterize the 7S globulins of wheat because they are minor storage proteins and do not contribute to the bread-making properties of wheat. Mass spectrometry and N-terminal sequencing have shown that peptides matching internal sequences of Gbl1 and WP5212 occur in wheat seeds (116-118). The present thesis research further characterized the 7S globulins of wheat.

WP5212-related proteins: major components of the wheat globulin fraction.

To characterize the WP5212 antigenically-related proteins, 1D and 2D protein separations were carried out. Blots were probed with antibodies raised against two predicted antigenic epitopes from the WP5212 sequence (145). Initially, the globulins were extracted by the classical method in 1 M NaCl, precipitated by dialysis against water and resuspended in 1 M NaCl (106, 156). Proteins were separated by SDS-PAGE and their reactivity against the polyclonal rabbit anti-WP5212 serum was verified by immunoblot (Figure 3.1, panels A and B). Samples prepared for two-dimensional separation were resuspended in rehydration buffer rather than 1 M NaCl. To verify whether the globulins were solubilized during this extraction method, this fraction was also separated by SDS-PAGE, transferred and probed with polyclonal WP5212 antiserum. Pre-immune serum and secondary/tertiary antibody controls were also performed (Figure 3.1, panel C). The WP5212 antigenically-related proteins appear to be major components of the salt-soluble fraction from AC Barrie seeds. The two, less abundant, proteins appear at M_r of approximately 66-68 and 64-65 kDa. Two, more intense regions occur at molecular mass ranges of ~47-53kDa and ~33-37kDa.

Figure 3.1 SDS-PAGE and western blot analysis of *Triticum aestivum* var. AC Barrie salt-soluble proteins. The salt-soluble (**SS**) fraction from AC Barrie seeds was separated under reducing conditions by SDS-PAGE (12% polyacrylamide). In panel **A**, the proteins are stained with CBB R-250. Standard lane (**PP**) is Precision Plus Protein (Bio-Rad). In panel **B**, proteins (**SS**) were immunoblotted with polyclonal anti-WP5212 serum at a 1:10,000 dilution. In panel **C**, the salt-soluble fraction dissolved in rehydration buffer, was separated by SDS-PAGE and immunoblotted with the anti-WP5212 serum (1:10,000), with pre-immune serum (1:10,000), or with secondary and tertiary antibodies alone. Marker lanes (**M**) are Biotinylated Protein Ladder (Cell Signaling Technology, Inc.).


C
2°/3° antibodies only
Pre-immune serum
WP5212 antiserum
M
kDa
80
60
50
40
30
20
10
B
SS
M
kDa
80
60
50
40
30
20
10
A
SS
PP
kDa
75
50
37
25

The WP5212-related proteins are primarily located in the embryo

The 7S cereal globulins are known to occur primarily in the embryo, with endosperm expression attributable to the aleurone layer (39, 87, 91, 106, 108). It was decided to compare the AC Barrie endosperm and embryo-enriched fractions by 2D separation according to pI and M_r, using the anti-WP5212 antibodies. In Figure 3.2, there exists a significant difference in the globulin expression patterns of wheat AC Barrie endosperm (panel A) and embryo (panel B). On the basis of the lack of immunogenicity of the endosperm fraction compared with the embryo-enriched fraction, the latter was selected for further globulin analysis. The immunoreactive spots in the embryo-enriched globulin blot correspond to the same molecular masses as the globulins from the one-dimensional study in Figure 3.1

A number of WP5212-related polypeptides make up the embryo globulin fraction

To determine whether any of the immunoreactive spots from the embryo-enriched fraction owed their signal to pre-immune antibodies, the fraction was separated by 2DE and probed with either pre-immune serum or with the polyclonal anti-WP5212 serum (Figure 3.3, panels A and B, respectively). A single spot, indicated by an arrowhead, corresponding to a protein of pI 6.2 and M_r of 66-67kDa was identified when probed with the pre-immune serum (panel A) and anti-WP5212 serum (panel B). No other significant reactivity was observed for the pre-immune serum.

In contrast, a number of spots were found to react with the anti-WP5212 serum suggesting that the majority of the AC Barrie embryo salt-soluble proteins are immunologically-related. There is a significant variation in the pI of the immunoreactive proteins ranging from a pH of 3 to ~10, the limits of the IPG strips. Two long trains of spots are observed with pI ranges of approximately 7.2-9, and ~7.4-10. Figure 3.4 shows a narrower pH range (7-10) in which these protein trains can be better resolved. The first, fainter train (pH 7.2-9.0) has a M_r ranging from approximately 65-70kDa; the second (pH 7.4-10), is in the Mr range of 47-53kDa. The 7-10 pH range also resolves a number of

Figure 3.2 Comparison of the globulin composition of the AC Barrie endosperm and embryo-enriched fractions. Globulins were extracted from wheat seed endosperm **(A)** and embryo-enriched **(B)** fractions in 1M NaCl and dialyzed against water. Precipitates were dissolved in rehydration buffer, centrifuged and stored at -80°C until use. IPG strips (pH 3-10) were passively rehydrated with 120 µg of globulin extract and focussed by IEF in the first dimension. SDS-PAGE was used to separate the proteins in the second dimension according to molecular mass. Proteins were stained with CBB R-250 (upper panel) or transferred to nitrocellulose and probed with polyclonal rabbit anti-WP5212 serum (lower panel). Marker lanes (**M**) are Prestained Benchmark (Invitrogen). Molecular masses shown on western blots are approximations.

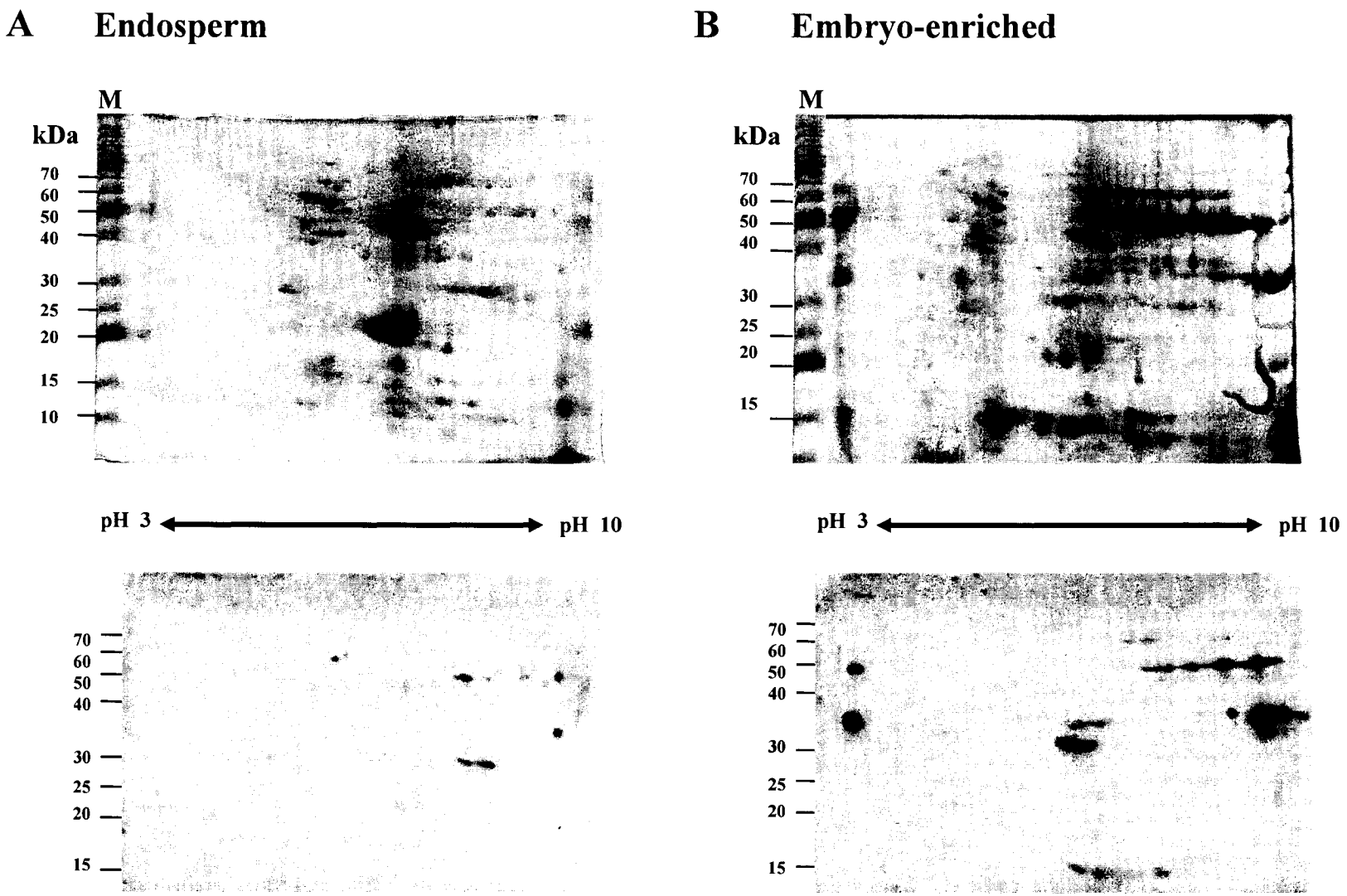
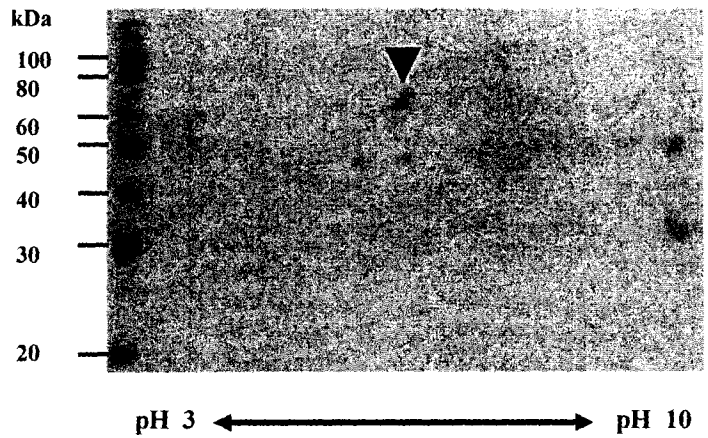

A Endosperm
B Embryo-enriched
M
kDa
70
60
50
40
30
25
20
15
10
pH 3
pH 10

Figure 3.3 Verification of polyclonal antiserum's specific reactivity against AC Barrie globulin components by immunoblot comparison of polyclonal antiserum and pre-immune serum. IPG strips (pH 3-10) were each passively rehydrated with 150 μg of AC Barrie embryo-enriched globulin extract and focused by IEF in the first dimension. Second dimension separation was carried out by SDS-PAGE prior to transfer to nitrocellulose. Membranes were probed with pre-immune serum **(A)** or with polyclonal anti-WP5212 serum **(B)**. The approximate location of a common immunoreactive spot, determined by superimposition of the blots, is indicated by the black arrowhead.

A Pre-immune serum



B Anti-WP5212 serum

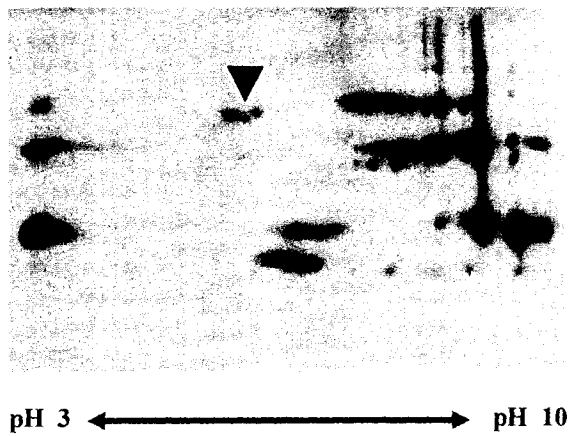
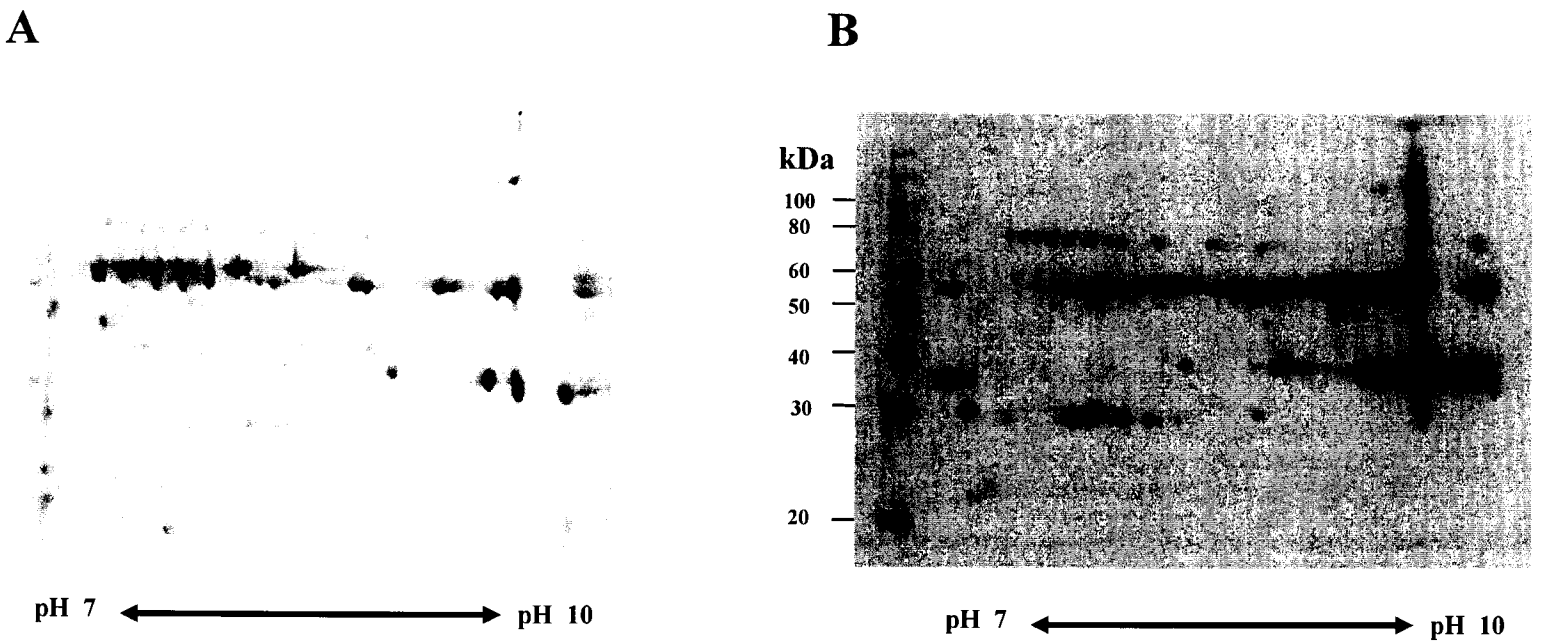


Figure 3.4 Two dimensional separation of salt-soluble proteins from the AC Barrie embryo-enriched fraction on a pH 7-10 gradient. IPG strips (pH 7-10) were passively rehydrated with 100 μg of globulin extract and focussed by IEF in the first dimension. SDS-PAGE was used to separate the proteins in the second dimension according to molecular mass. Proteins were stained with CBB R-250 **(A)** or transferred to nitrocellulose, blocked and probed with polyclonal anti-WP5212 serum **(B)**.


A
pH 7
pH 10
B
kDa
100
80
60
50
40
30
20
pH 7
pH 10

spots in the 28-30kDa and ~35kDa regions. The ~35kDa and ~50kDa basic proteins with pI greater than 8.5 appear to be the most abundant in the embryo fraction. In Figure 3.3, panel B, two smaller trains are also observed at pH~6.5-7 (Mr ~30kDa) and pI 6.2-6.8 (Mr ~28kDa). Some of these spots appear at lower molecular masses than were observed by 1D western analyses. This size difference may be a technical matter, possibly arising from the lack of a stacking gel in the second dimension.

Identification of selected WP5212-related polypeptides

A preliminary investigation, by mass spectrometry, of six immunoreactive spots was undertaken. The spots selected reacted with the anti-WP5212 polyclonal antibodies but fell outside the predicted M_r and pI values of the WP5212 protein: 66.6kDa and 8.5, respectively. These spots represent major polypeptides identified by 2DE and cover a range of M_r and isoelectric points. Figure 3.5, panel A, shows an overloaded IEF/SDS-PAGE gel of the wheat embryo-enriched globulin fraction. Each of the labelled spots was excised, as shown in panel B, and analyzed by LC-MS/MS. Mass spectrometry results are summarized in Table 3.1. Five of the six spots submitted were identified, by interrogating the NCBI database, as Glb1 or Beg1, maize and barley globulin1 homologues, respectively. One spot could not be identified: spot 3. The five identified spots could only be matched to two globulins in the NCBI database.

Glb1 and Glb2-like sequences and alignments

The only full-length wheat 7S globulin sequence deposited to the NCBI database is Glb1 (AAA34269). The coding sequence from the *Glb1* genomic clone (M81719) is identical to that of barley embryo globulin, *Beg1*. Given that prior literature suggests sequence divergence between the wheat and barley coding sequences (31, 109), it has been suggested that the wheat genomic clone is actually a barley sequence (90). The group that deposited the sequence was contacted but could not comment to the effect (personal communication, Ralph Quatrano). The WP5212 cDNA, isolated from wheat cv. AC Barrie, shares 90% identity with a 1387 nucleotide region of the Glb1/Beg1

Figure 3.5 Two-dimensional separation of the AC Barrie embryo-enriched globulin fraction by SDS-PAGE for mass spectrometry. IPG strip (pH 3-10) was passively rehydrated with 250 μg of globulin extract and focussed by IEF in the first dimension. The strip was overloaded to ensure sufficient quantities of each protein would be present. SDS-PAGE was used to separate the proteins in the second dimension according to molecular mass. Proteins were stained with Coomassie BioSafe (Bio-Rad) **(A)**; spots of interest were excised for mass spectrometry **(B)**. Spots chosen for LC-MS/MS reacted with the anti-WP5212 polyclonal antibodies and represented a sampling of the major observed molecular masses (~30kDa and ~50kDa) and isoelectric points in the acidic (pH 3), neutral (pH 6-7) and basic (pH 9-10) regions. Spots were randomly labeled 1 – 6.

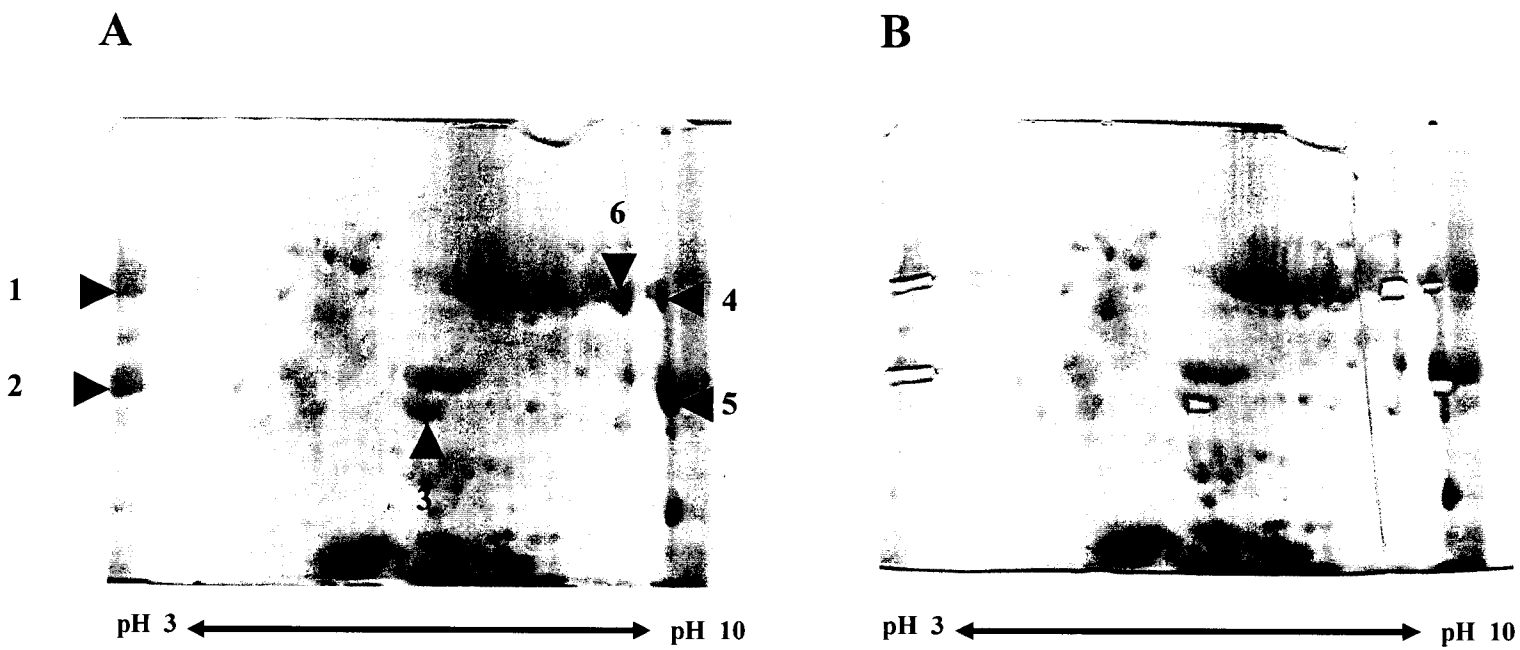

A
1
2
3
4
5
6
pH 3
pH 10
B
pH 3
pH 10

Table 3.1 Mass spectrometry search results of selected spots from the AC Barrie embryo-enriched salt-soluble fraction. Protein spots shown in Figure 3.5 were excised, digested with chymotrypsin and subjected to LC-MS/MS. The search engine Mascot (152) was used to match peptide masses to proteins in the NCBI database. Observed M_r/pI are the approximate molecular masses and isoelectric points of the corresponding proteins observed in Figure 3.5. Theoretical M_r/pI correspond to calculated values for the identified sequence. Sequence coverage is the ratio of the total number of residues in matched peptides to the number of residues in the identified protein. Non-duplicated peptide sequences matching the identified protein are reported.
*oxid. (M): oxidation of methionine

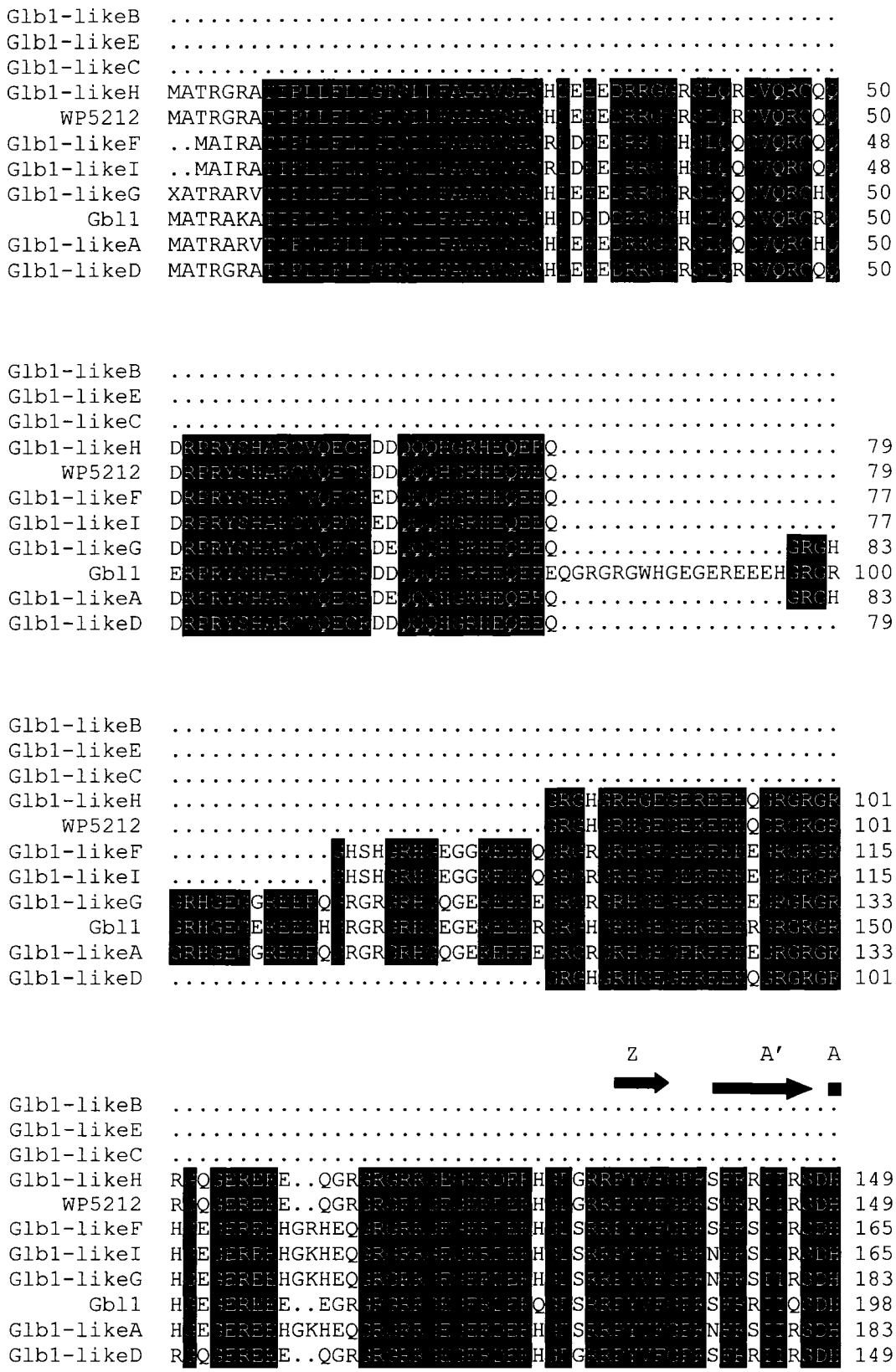
Spot ID	Approx. observed M_r/pI	Database match (species, protein and accession)	Theoretical M_r/pI	Sequence coverage (%)	Non-repetitive matching peptide sequences
1	48-50kDa/ 3	*Zea mays subs parviglumis*, globulin-1, gi|3414837	27kDa/ 8.6	8.9	DEVSRLL; HTISVPGKF; HTISVPGRF; RPFDEVSRLL
	48-50kDa/ 3	Barley, globulin Beg1 precursor, gi|421978	72kDa/ 7.3	6.3	RVAIMEVNPRAF + oxid. (M)*; VAQGEGVLTVIENGEKRSY
2	32-33kDa/ 3	Barley, globulin Beg1 precursor, gi|421978	72kDa/ 7.3	5.5	LASL; RVAIMEVNPRAF + oxid. (M)*; VAQGEGVLTVIENGEKRSY
3	28-30kDa/ 6.5	no significant match			N/A
4	49-50kDa/ 9.75	*Zea mays subs parviglumis*, globulin-1, gi|3414837	27kDa/ 8.6	2.5	DEVSRLL; HTISVPGKF; HTISVPGRF
5	30-33kDa/ 9.75	*Zea mays subs parviglumis*, globulin-1, gi|3414837	27kDa/ 8.6	8.1	DEVSRLL; HTISVPGMF+ oxid. (M)*; HTISVPGRF; RPFDEVSRLL
	30-33kDa/ 9.75	Barley, globulin Beg1 precursor, gi|421978	72kDa/ 7.3	6.3	RVAIMEVNPRAF; RVAIMEVNPRAF + oxid. (M); VAQGEGVLTVIENGEKRSY
6	48-49kDa/ 9.5	*Zea mays subs parviglumis*, globulin-1, gi|3414837	27kDa/ 8.6	8.1	DEVSRLL; HTISVPGKF; HTISVPGRF; RPFDEVSRLL
	48-49kDa/ 9.5	Barley, globulin Beg1 precursor, gi|421978	72kDa/ 7.3	3.3	HTISVPGKF; RVAIMEVNPRAF + oxid. (M)*

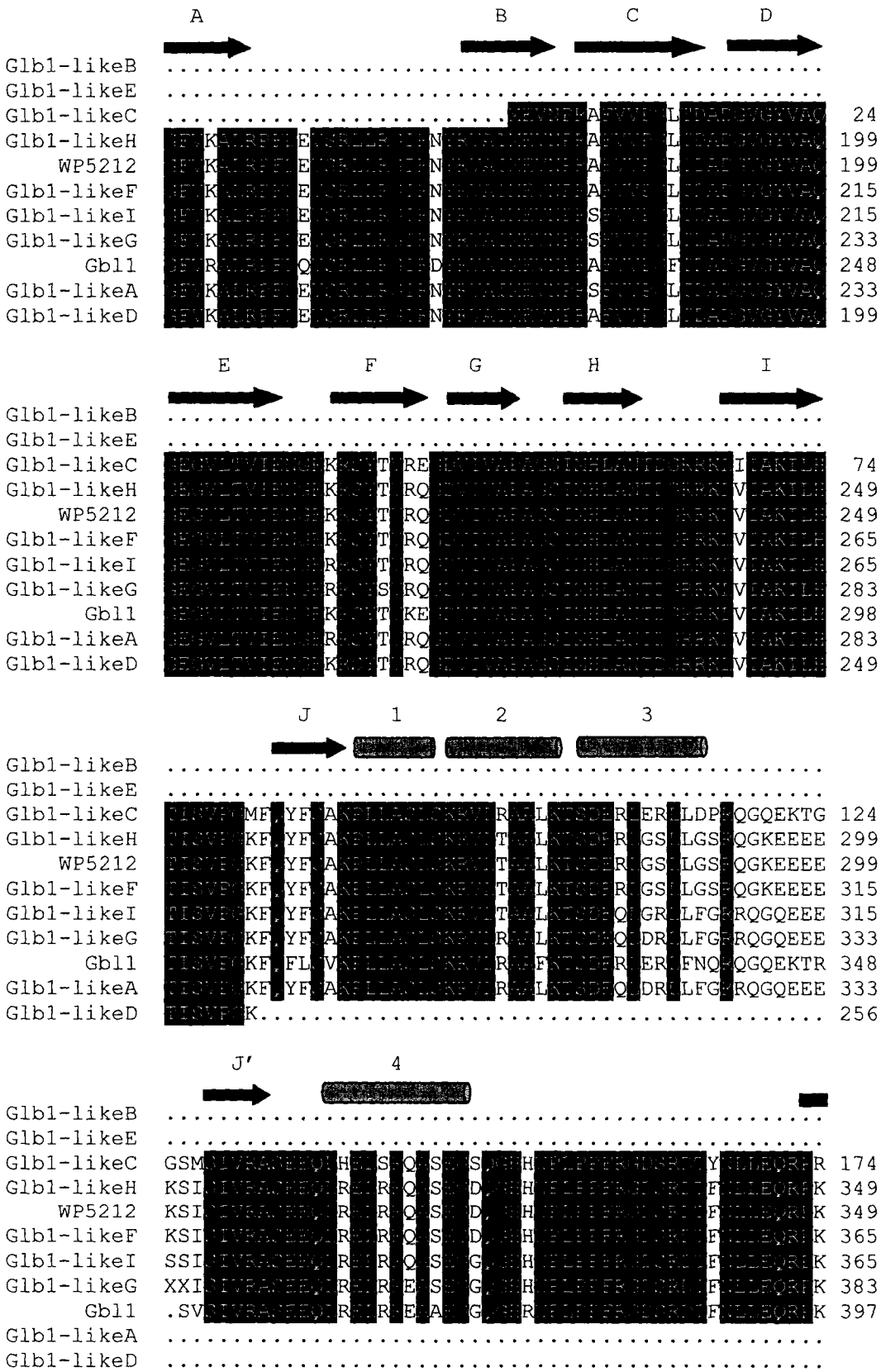
coding sequence. Since there is evidence that the wheat 7S globulins are encoded by a multigene family (88), the Gbl1 sequence was retained for this study.

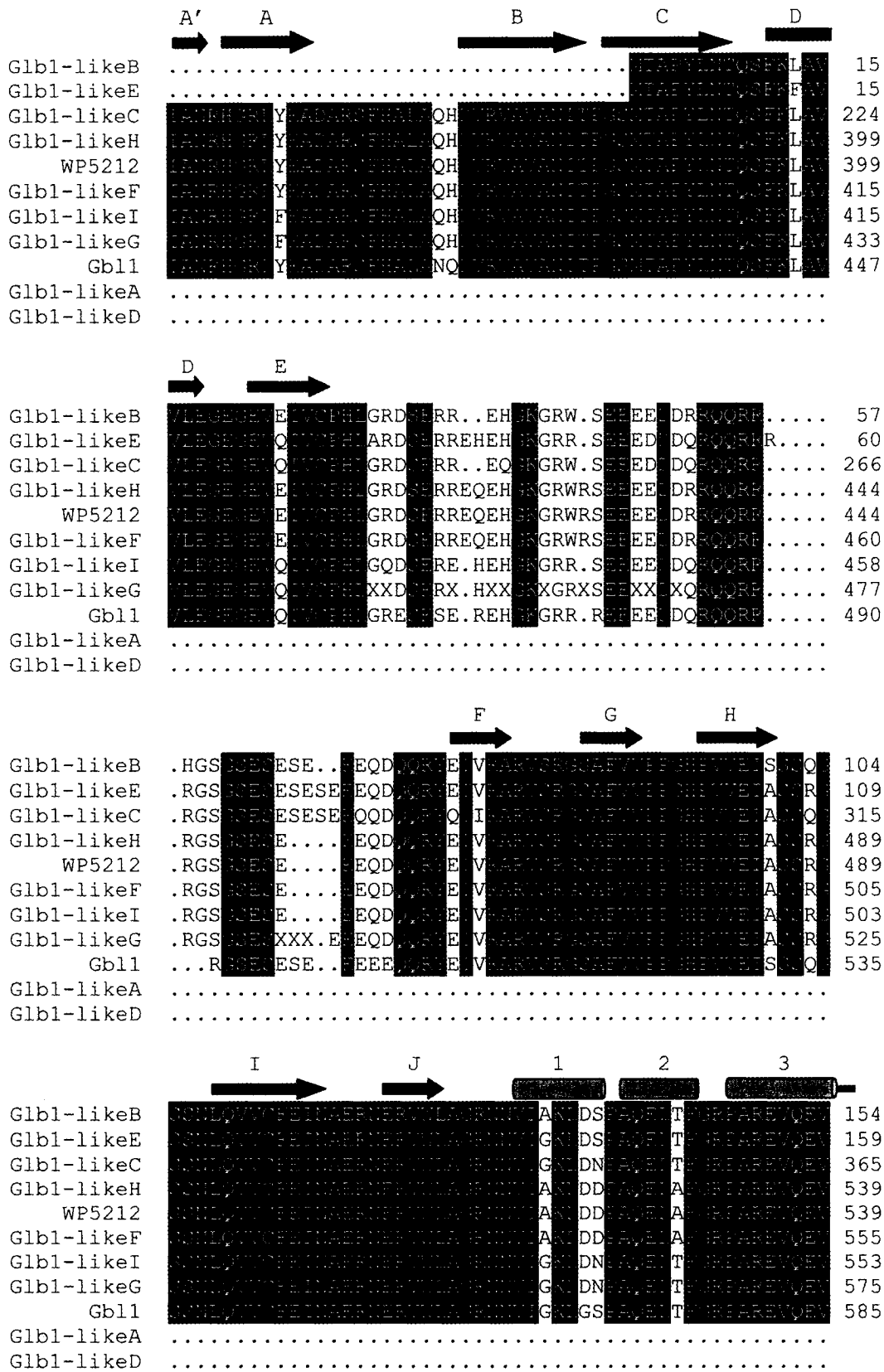
Since only one wheat/barley globulin sequence is currently available to the Mascot search engine via NCBI, two EST databases were explored. The TIGR wheat genome database is a centralized resource for all available wheat sequence data (http://www.tigr.org/, last viewed 31/5/06). The GrainGene2.0 database also offers wheat sequence data (http://wheat.pw.usda.gov/GG2/index.shtml, last viewed 31/5/06). A collection of wheat EST contigs available through these databases were analyzed for homology with WP5212 using BLAST. Sequences identified were classified according to their homology with either the cereal globulin1 (Glb1-like) or globulin2 (Glb2-like) families. The contig sequences were analyzed using the NCBI ORF finder (http://www.ncbi.nih.gov/gorf/gorf.html, last viewed 31/5/06) to determine protein sequence. The resultant proteins are presented in Figure A.2. The conceptually translated proteins were aligned using ClustalX (153), with minor adjustments made manually in Genedoc (154).

Figure 3.6 is an alignment of the Glb1-related proteins, where identity is highlighted in black. A number of identified sequences did not cover the entire length of the *WP5212* cds. These sequences generated conceptual proteins truncated at either the N- or C-terminus, relative to the WP5212 sequence. A subset of these sequences were retained for analysis, but due to the large number of these shorter EST contigs, only a few could be used in this study (Glb1-like A, B, C, D, E, I). Contigs corresponding to the full-length of the WP5212 cds were also identified: Glb1-like F, Glb1-like G, Glb1-like H. One of the resultant proteins, Glb1-like H, is identical to WP5212 over its entire length. Another protein sequence, Glb1-like F, is nearly identical except at its N-terminus. Figure 3.7 shows the similarity between full-length Glb2 proteins (Glb2-like A and Glb2-like B) and Glb1-related proteins, Gbl1 and WP5212. The sequences exhibit 31% identity over the full-length of the Glb2 proteins.

Figure 3.6 Multiple sequence alignment of the Glb1-like sequences of wheat. The protein sequences from Figure A.2, corresponding to Glb1-like proteins, were aligned in ClustalX (153); minor adjustments were made manually using Genedoc (154). Sequences Glb1-like A and Glb1-like D were truncated at their C-termini by 41 and 14 residues, respectively, to remove regions that did not fit the alignment. A variable length N-terminal region characteristic of large 7S globulins (31), is observed in all sequences except Glb1-like B, Glb1-like C and Glb1-like E. Glb1-like I is incomplete at its C-terminus as its coding sequence does not code for a stop codon. Some sequences contain X when the amino acid identity is uncertain. Black highlights show residue identity. To the right of each alignment block, the position of the last residue for a given sequence is reported. The secondary structure elements, occurring along the 7S globulin polypeptide chain, are overlaid on the alignment and labeled according to convention (23, 44). β-strands are represented by arrows; α helices, by cylinders. Helix 4 is part of the linker region that connects the structurally similar N- and C-terminal modules.


Glb1-likeB
Glb1-likeE
Glb1-likeC
Glb1-likeH
WP5212
Glb1-likeF
Glb1-likeI
Glb1-likeG
Gbl1
Glb1-likeA
Glb1-likeD
Z
A′
A


A
B
C
D
Glb1-likeB
Glb1-likeE
Glb1-likeC
Glb1-likeH
WP5212
Glb1-likeF
Glb1-likeI
Glb1-likeG
Glb11
Glb1-likeA
Glb1-likeD
E
F
G
H
I
J
1
2
3
J'
4


A′ A B C D
Glb1-likeB 15
Glb1-likeE 15
Glb1-likeC 224
Glb1-likeH 399
WP5212 399
Glb1-likeF 415
Glb1-likeI 415
Glb1-likeG 433
Gbl1 447
Glb1-likeA
Glb1-likeD
D E
Glb1-likeB 57
Glb1-likeE 60
Glb1-likeC 266
Glb1-likeH 444
WP5212 444
Glb1-likeF 460
Glb1-likeI 458
Glb1-likeG 477
Gbl1 490
Glb1-likeA
Glb1-likeD
F G H
Glb1-likeB 104
Glb1-likeE 109
Glb1-likeC 315
Glb1-likeH 489
WP5212 489
Glb1-likeF 505
Glb1-likeI 503
Glb1-likeG 525
Gbl1 535
Glb1-likeA
Glb1-likeD
I J 1 2 3
Glb1-likeB 154
Glb1-likeE 159
Glb1-likeC 365
Glb1-likeH 539
WP5212 539
Glb1-likeF 555
Glb1-likeI 553
Glb1-likeG 575
Gbl1 585
Glb1-likeA
Glb1-likeD

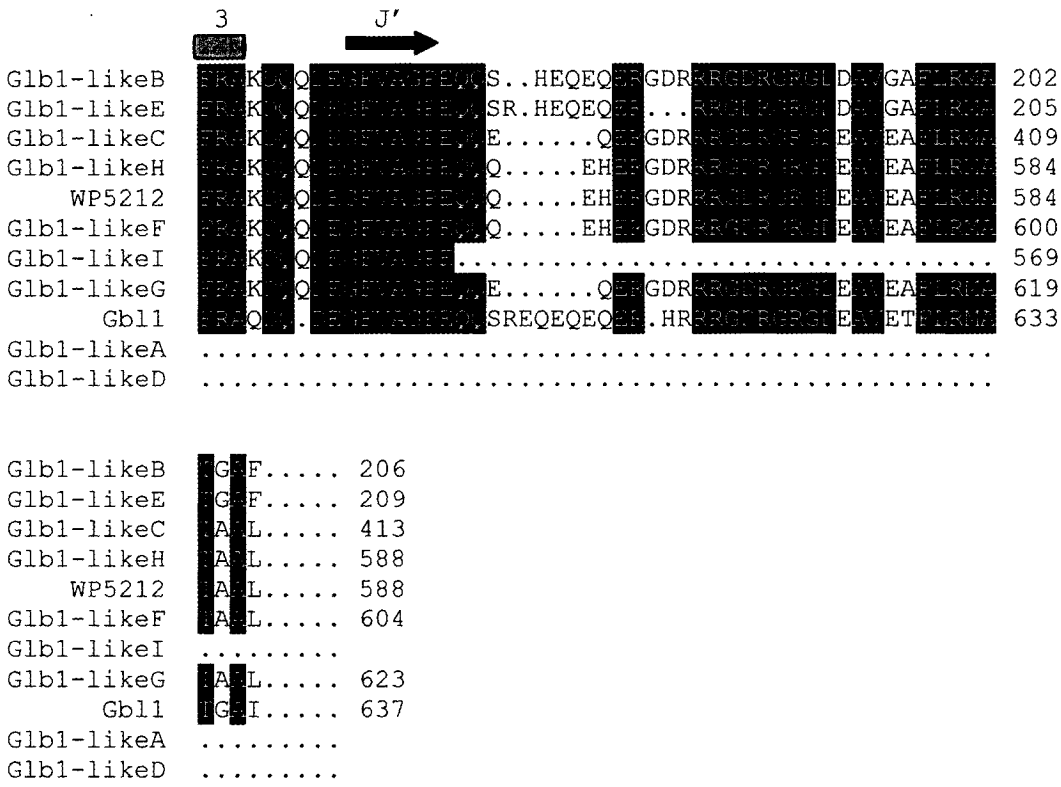
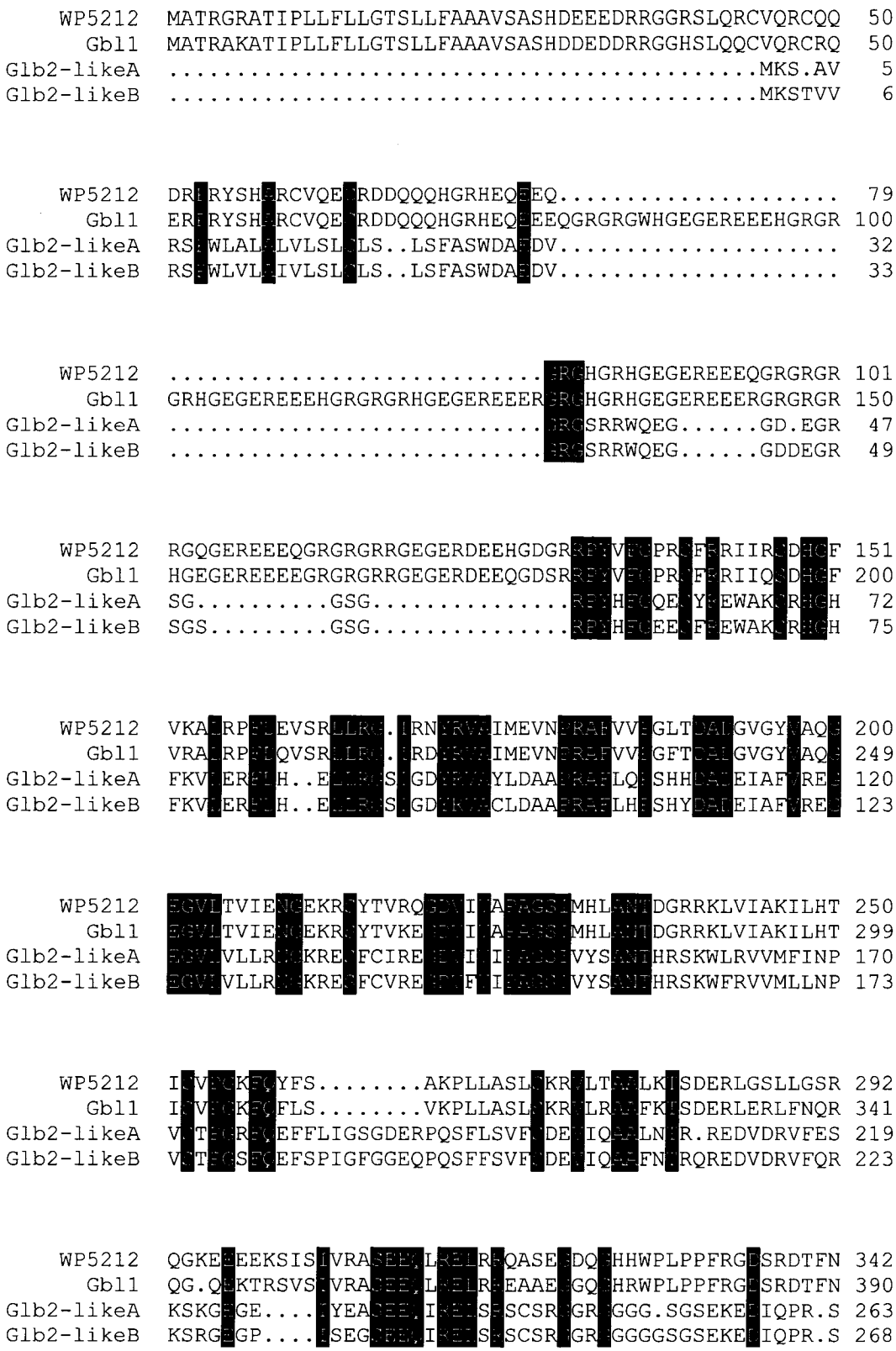
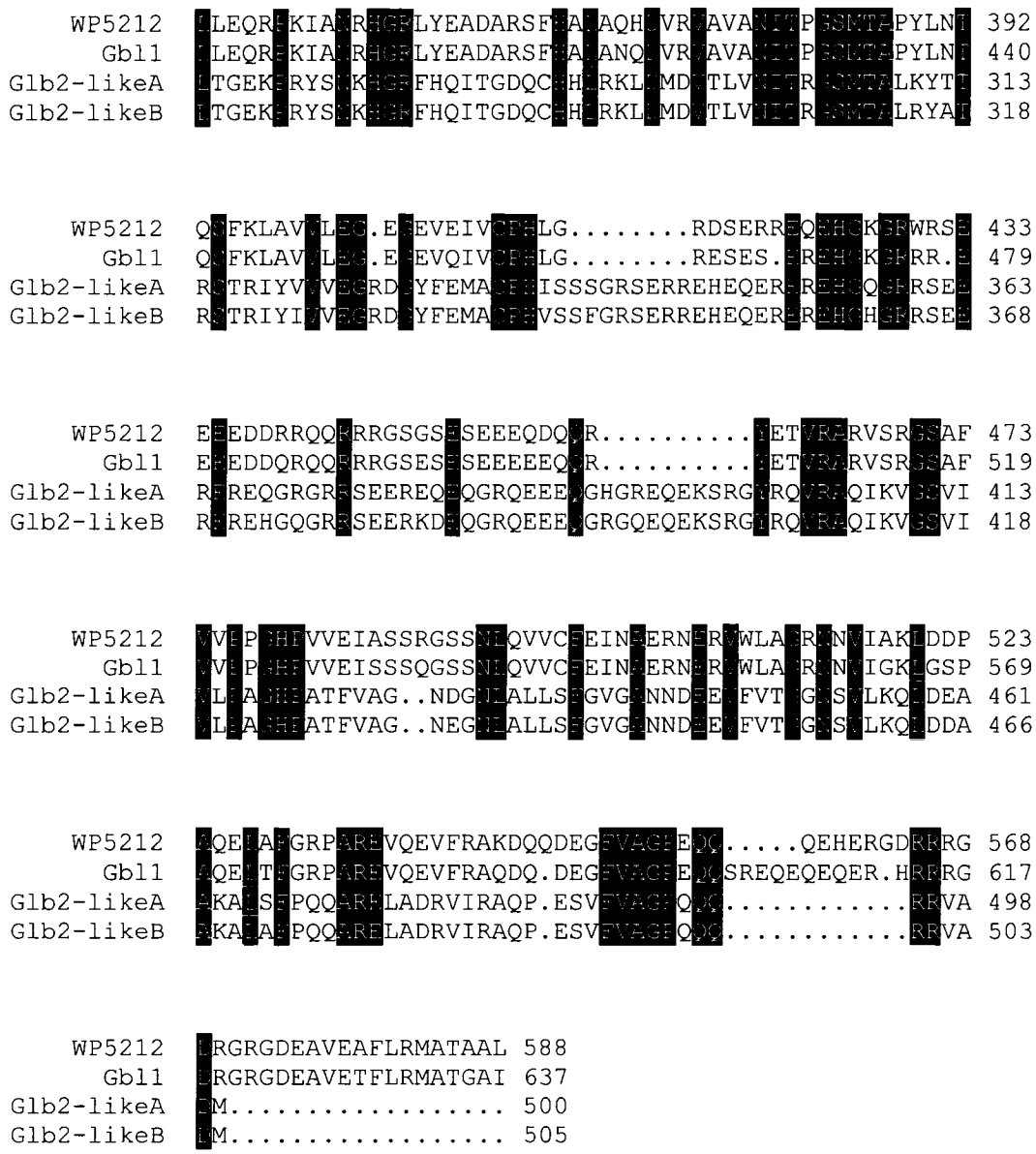

3
J'
Glb1-likeB
Glb1-likeE
Glb1-likeC
Glb1-likeH
WP5212
Glb1-likeF
Glb1-likeI
Glb1-likeG
Gbl1
Glb1-likeA
Glb1-likeD

Figure 3.7 Multiple sequence alignment of wheat Glb2-like sequences to WP5212 and Gbl1. The protein sequences from Figure A.2, corresponding to full-length Glb2-like proteins (85), were aligned in ClustalX (153) to WP5212 and Gbl1 (full-length Glb1-like proteins). Minor adjustments to the alignment were made manually using Genedoc (154). Conserved residues are highlighted with a black background. Within the alignment, 155 residues were conserved. Thus, the Glb1 and Glb2-like proteins share ~ 31% identity over the length of the Glb2-like sequences. To the right of each alignment block, the position of the last residue for a given sequence is reported.


WP5212 MATRGRATIPLLFLLGTSLLFAAAVSASHDEEEDRRGGRSLQRCVQRCQQ 50
Gbl1 MATRAKATIPLLFLLGTSLLFAAAVSASHDDEDDRRGGHSLQQCVQRCRQ 50
Glb2-likeA .....MKS.AV 5
Glb2-likeB .....MKSTVV 6
WP5212 DR RYSH RCVQE RDDQQQHGRHEQ EQ..... 79
Gbl1 ER RYSH RCVQE RDDQQQHGRHEQ EEQGRGRGWHGEGEREEEHGRGR 100
Glb2-likeA RS WLAL LVLSL LS..LSFASWDA DV..... 32
Glb2-likeB RS WLVL IVLSL LS..LSFASWDA DV..... 33
WP5212 ..... RGHGRHGEGEREEEQGRGRGR 101
Gbl1 GRHGEGEREEEHGRGRGRHGEGEREEER RGHGRHGEGEREEERGRGRGR 150
Glb2-likeA ..... RGSRRWQEG.....GD.EGR 47
Glb2-likeB ..... RGSRRWQEG.....GDDEGR 49
WP5212 RGQGEREEEQGRGRGRRGEGERDEEHGDGR PV PR FR IIR DHGF 151
Gbl1 HGEGEREEEEGRGRGRRGEGERDEEQGDSR PV PR FR IIQ DHGF 200
Glb2-likeA SG.....GSG..... HQ QE Y EWAK RHGH 72
Glb2-likeB SGS.....GSG..... HQ EE F EWAK RHGH 75
WP5212 VKA RP EVSR LRG. RN RVA IMEVN RAF VV GLT DAL GVGY AQG 200
Gbl1 VRA RP QVSR LRG. RD RVA IMEVN RAF VV GFT DAL GVGY AQG 249
Glb2-likeA FKV ER H..E S GD YLDAA RAF LQ SHH DAL EIAF RE 120
Glb2-likeB FKV ER H..E S GD CLDAA RAF LH SHY DAL EIAF RE 123
WP5212 EGVL TVIE GEKR YTVRQ DV I A AGS MHL NT DGRRKLVIAKILHT 250
Gbl1 EGVL TVIE GEKR YTVKE DV I A AGS MHL NT DGRRKLVIAKILHT 299
Glb2-likeA EGVL VLLR KRE FCIRE EV I I AGS VYS NT HRSKWLRVVMFINP 170
Glb2-likeB EGVL VLLR KRE FCVRE DV F I AGS VYS NT HRSKWFRVVMLLNP 173
WP5212 I V KF YFS.....AKPLLASL KR LT LK SDERLGSLLGSR 292
Gbl1 I V KF FLS.....VKPLLASL KR LR FK SDERLERLFNQR 341
Glb2-likeA V T RF EFFLIGSGDERPQSFLSVF DE IQ LN R.REDVDRVFES 219
Glb2-likeB V T SF EFSPIGFGGEQPQSFFSVF DE IQ FN RQREDVDRVFQR 223
WP5212 QGKE EEKSIS VRA EE L RE R QASE DQ HHWPLPPFRG SRDTFN 342
Gbl1 QG.Q KTRSVS VRA EE L RE R EAAE GQ HRWPLPPFRG SRDTFN 390
Glb2-likeA KSKG GE....YEA EE I RE S SCSR GR GGG.SGSEKE IQPR.S 263
Glb2-likeB KSRG GP....SEG EE I RE S SCSR GR GGGGSGSEKE IQPR.S 268


WP5212
Gbl1
Glb2-likeA
Glb2-likeB

Further characterization of the selected WP5212-related polypeptides

The protein sequences from Figure A.2 were digested *in silico* and the predicted peptide masses were compared with LC-MS/MS data using Mascot. The results from this search are presented in Table 3.2. Armed with more sequence data, the program was able to match all six spots to wheat Glb1-like sequences. Spot 1 also matched to a peptide of Glb2-like sequences (Glb2-like B and C). Occasionally, more than one protein contained matching peptide masses. In these cases, the highest match is shown first, followed by the next matches in descending order. Though the match for spot 3 was statistically significant ($p<0.05$), only one peptide mass matched known globulin sequences.

To obtain more information about the WP5212 immunologically-related proteins, it was decided to proceed with N-terminal sequencing of four of the spots chosen for mass spectrometry in Figure 3.5. The results are presented in Table 3.3. Two of the sequences (spots 1 and 4) may be N-terminally blocked because no information could be obtained, despite protein visualization after amido black staining. Spot 5 was difficult to visualize following protein transfer and staining; results from N-terminal sequencing suggest little protein was present. The first two residues could not be determined (X) and the last residue was reported as either arginine (R) or glutamic acid (E). The result from this spot must be used with caution as low protein quantity results in reduced accuracy (personal communication with John Smith, UTBM). Nonetheless, six residues from the result XXHGDSRR, matched with Glb1-related sequences: Glb1-like A, F, G and I. Finally, Spot 3, which could not initially be identified by mass spectrometry, was identified by N-terminal sequencing as having an N-terminus that matches Gbl1, WP5212 and Glb1-like F, G, H, I sequences. This finding corroborates the data for spot 3 from Table 3.2.

Table 3.2 Mass spectrometry search results matching known globulin-like sequences from wheat. Protein spots shown in Figure 3.5 were excised, digested with chymotrypsin and subjected to LC-MS/MS. The search engine Mascot (152) was used to match peptide masses to the theoretical wheat globulin-like sequences from Figure A.2. Observed M_r/pI are the approximate molecular masses and isoelectric points of the corresponding proteins observed in Figure 3.5. Theoretical M_r/pI correspond to calculated values for the identified sequence. Sequence coverage (Seq. cover) is the ratio of the total number of residues in matched peptides to the number of residues in the identified protein. Non-duplicated peptide sequences matching the identified protein are reported.

Spot ID	Approx. observed M_r/pI	Globulin name	Theoretical M_r/pI	Seq. cover (%)	Non-repetitive matching peptide sequences	Other proteins matching same set of peptides
1	48-50kDa/ 3	WP5212	67kDa/8.48	13	RPFDEVSRLL; RVAIMEVNPRAF + oxid.(M); VAQGEGVLTVIENGEKRSY; HTISVPGKF; VVPPGHPVVEIASSRGSSNL; GRPAREVQEVF	Glb1-like F & Glb1-like H
	48-50kDa/3	Glb2-like C	37kDa/9.53	2.5	AFPQQARE	Glb2-like B
2	32-33kDa/ 3	WP5212	67kDa/8.48	7	RPFDEVSRLL; RVAIMEVNPRAF + oxid. (M); VAQGEGVLTVIENGEKRSY	Glb1-like D, Glb1-like F and Glb1-like H
3	28-30kDa/ 6.5	WP5212	67kDa/8.48	1.2	VVPPGHP	Gbl1, Glb1-like B, C, E, F, G, H, I
4	49-50kDa/ 9.75	WP5212	42kDa/10.17	2.7	DEVSRLL; HTISVPGKF	Glb1-like A, F, G, H, I
5	30-33kDa/ 9.75	WP5212	67kDa/8.48	13	RPFDEVSRLL; RVAIMEVNPRAF + oxid. (M); VAQGEGVLTVIENGEKRSY; SAKPLLASL TVRQGDVIVAPAGSIMHL; VVPGLTDADGVGY	Glb1-like F & Glb1-like H
	30-33kDa/ 9.75	Glb1-like A	42kDa/10.17	21	RPFDEVSRLL; VAQGEGVLTVIENGERRSY; VVPGLTDADGVGY; TVRQGDVIVAPAGSIMHL; SAKPLLASL; KTSDEQLDRLLF	—
	30-33kDa/ 9.75	Glb1-like C	46kDa/6.13	6.8	SAKPLLASL; HTISVPGMF; VVPGLTDADGVGY; VAQGEGVLTVIENGEKRSY	—
6	48-49kDa/ 9.5	WP5212	67kDa/8.48	10	RPFDEVSRLL; RVAIMEVNPRAF + oxid. (M); HTISVPGKF; KTSDERLGSLL; VVPPGHPVVEIASSRGSSNL	Glb1-like F & Glb1-like H
	48-49kDa/ 9.5	Glb1-like G	70kDa/7.42	5	RPFDEVSRLL; HTISVPGKF; KTSDEQLDRLLF	Glb1-like A

Table 3.3 N-terminal sequencing results of selected spots from AC Barrie embryo-enriched salt-soluble fraction. Selected protein spots from Figure 3.5 were sent for N-terminal sequencing by Edman degradation to complement the sequence information obtained by LC-MS/MS. Observed M_r/pI are the approximate molecular masses and isoelectric points of the corresponding proteins observed in Figure 3.5. Theoretical M_r/pI were calculated by http://www.expasy.ch/tools/pi_tool.html, last viewed 3/6/06). The N-terminal sequences were matched to the putative wheat globulin proteins shown in Figure A.2.

*protein was either blocked or quantity was too low. Given that the protein could be easily visualized by amido black staining, it is possible that the N-terminus was blocked.
**protein quantity was very low, therefore result may not be accurate – the last amino acid could not be determined with certainty - it was identified as either E or R.
***no match could be found among globulin proteins, though other proteins (not shown) do carry this sequence.
X: amino acid identity could not be determined.

Spot ID	Approximate observed M_r/pI	N-terminal sequencing performed?	Sequence obtained	Sequences matched	Theoretical M_r/pI
1	48-50kDa/ 3	Yes	No*	—	—
2	—	No	—	—	—
3	28-30kDa/ 6.5	Yes	SRDTFNLL	Gbl1 WP5212 Glb1-like F Glb1-like G Glb1-like H Glb1-like I	72kDa/6.80 67kDa /8.48 68 kDa/6.69 70kDa/7.42 67kDa/8.48 64kDa/6.69
4	49-50kDa/ 9.75	Yes	No*	—	—
5	30-33kDa/ 9.75	Yes	XXHGDSRR**	Glb1-like A Glb1-like F Glb1-like G Glb1-like I	42kDa/10.17 68kDa/6.69 70kDa/7.42 64kDa/6.69
	30-33kDa/ 9.75		XXHGDSRE**	No match ***	—
6	—	No	—	—	—

Chapter IV – Discussion

The wheat embryo globulins

When the AC Barrie globulin fraction is resolved by SDS-PAGE under reducing conditions, three major protein size ranges are observed at approximately: 64-70kDa; 47-53kDa and 33-37kDa (Figure 3.1). For the purposes of this discussion the size groups will be referred to as 65kDa, 50kDa and 35kDa, respectively. These three major wheat globulin groups were observed in previous studies and were identified as 7S globulins (105, 108). These groups were found to be enriched in the embryo and aleurone layer of wheat (15, 108, 109) and other cereals (39, 86, 87, 91, 106). In previous studies, the wheat globulins were found to occur at slightly higher molecular masses, differing by approximately 2-3kDa. This discrepancy may be a function of the different markers used to determine M_r. Alternatively, allelic variation between wheat varieties used in these studies could be the basis of the observed difference. Consistent with these previous findings, probing with anti-WP5212 polyclonal serum revealed that the wheat endosperm expresses only small amounts of detectable protein (Figure 3.2) in its globulin fraction. Preliminary immunolocalization studies performed in our lab suggest that this expression is attributable to the aleurone layer (personal communication with Charles Melnyk; see frontispiece). In the embryo-enriched fraction (Figures 3.2 to 3.4), anti-WP5212 reactivity is observed for each of the major size groups. The 65kDa protein group appears to make up a minor fraction of the immunoreactive proteins, while the 50kDa and 35kDa groups are major components. Similar observations were made when the wheat embryo globulins were probed with antiserum raised against the 7S globulins of oat (105, 108, 109). A component of wheat M_r 50kDa globulins was found to react with antibodies against Glb1 from maize and vicilin from pea (36, 109).

Composition of the wheat embryo salt-soluble fraction

The 65kDa protein group consists of at least two distinguishable bands on one-dimensional blots (Figure 3.1). In maize, the Glb1 proteins consist of two major polypeptides that differ in size by approximately 2kDa. These have been shown to be products of the same gene, that undergo different post-translational modifications. Glb1 is a cleavage product of Glb1' (as described in Figure 1.4). A similar series of PTMs may generate the wheat globulins observed at ~ 64-65 and ~66-68kDa. When separated in two dimensions (Figures 3.3 and 3.4), this size group appears as a train of spots ranging in pI from 7.2-9.0. As will be discussed below, this train could be the result of PTMs including N-terminal sequence removal and glycosylation. Futhermore, if the wheat globulins are encoded by a family of homoeologous genes, as has been suggested (88), the resulting proteins could differ slightly in M_r and pI to generate the observed pattern. The immunoreactivity observed in Figures 3.1-3.4 suggests that the major components of the wheat embryo salt-soluble proteins, 50kDa and 35kDa, are related to WP5212. The polypeptides demonstrate a range of isoelectric points concentrated in the basic range when separated by 2DE (Figure 3.3).

For each of the globulin components (65, 50 and 35kDa), an immunoreactive protein could be found at the acidic end of the pH spectrum, exhibiting a pI of approximately 3 (Figure 3.3). The wheat globulin 2DE pattern is similar to that published by Robert, *et al*, 1985, in the basic range, although the present study identifies more proteins in this range, as well as proteins not previously identified in the acidic range. Some of the differences may be due to the fractionation procedure used. The previous study used a pH 3.5-10 gradient in the first dimension which could have resulted in the loss of the acidic polypeptides.

Chemical modifications, such as deamidation and carbamylation, of the basic proteins during protein sample preparation could account for the acidic polypeptides observed. Alternatively, these may

arise from naturally-occurring PTMs (157). Another possibility is that the acidic proteins, while immunologically-related to the more basic proteins, are encoded by different genes, resulting in different 2DE patterns.

Wheat 7S globulin sequences

Full- and partial-length wheat EST contigs with homology to either globulin1 or globulin2 were extracted from the TIGR and GG2.0 databases. These two types of 7S globulin coding sequences are known to occur in other cereals and their protein products have been identified in maize and rice (39, 83, 85-87). The Glb1-related sequences among cereals are more closely related in sequence than the Glb1 and Glb2 sequences within a given species. In maize, the Glb1 and Glb2 sequences have diverged over evolutionary time, yet they exhibit approximately 28% identity over the length of the Glb2 protein. In comparision, the Glb1 and Beg1 sequences are 63% identical at the amino acid level (90).

The first alignment performed, Figure 3.6, is an alignment of the Glb1-related sequences. Gbl1, WP5212, Glb1-like F, G and H are full-length sequences. Glb1-like A and D are intact at the N-terminus but are largely incomplete at the C-terminus; Glb1-like I is missing a short sequence at its C-terminus. Finally, Glb1-like B, C and E are truncated at their N-termini. These shorter sequences most likely represent incomplete globulin contigs. The WP5212 sequence is identical to the GG2.0 sequence Glb1-like H and these are nearly identical to Glb1-like F except in their N-terminal segment.

The N-terminal segment, is characterized by 16-amino acid hydrophilic repeats and a CXXXC motif observed in 7S globulins of other plants (31). There appear to be at least three differences in the length of the N-terminus in this region among the wheat Glb1-related sequences. Gbl1 has the longest segment characterized by seven repeating ~16-amino acid tracts beginning at position 55

(39). The other globulin1 sequences differ by either three tracts (WP5212, Glb1-like D and H), by two tracts (Glb1-like F and I) or by one tract (Glb1-like A and G). Each tract translates to a size of approximately 2kDa which may explain some of the heterogeneity observed in the globulin 2DE pattern. These differences in tandem repeats may be attributable to the amplification and insertion of a transposable element, or to the duplication of a pre-existing sequence (158). Repeated motifs are a common phenomenon in Triticeae prolamin storage protein genes (159).

The wheat globulin1 sequences are otherwise highly homologous with occasional amino acid polymorphisms, insertions and/or deletions in specific regions. The overlaid secondary structure elements, derived from the crystal structures of related globulins, phaseolin and β-conglycinin suggest that the insertions and deletions map to loop regions that will not affect the tertiary and quaternary structure of the proteins (23, 44, 45). In regions of the alignment where identity is not conserved, Gbl1 and Glb1-like C appear to diverge to a greater extent than the other sequences. As more full-length sequences become available for the globulin1-like genes and proteins, it will become possible to establish their evolutionary relationships.

The alignment in Figure 3.7 compares the sequences of Glb2-like A and B to the known Glb1-related sequences: WP5212 and Gbl1. The N-terminal segment described for the Glb1-like sequences is significantly reduced in the Glb2-like proteins. This segment separates the large from the small 7S globulins and thus according to this convention Glb2-like A and B can be considered small 7S globulins (23, 31). The Glb2-like sequences share approximately 31% identity with the Glb1-like sequences from wheat, similar to the identity between the maize Glb1 and Glb2 proteins (90).

Relationships of major 7S globulins of wheat

Based on immunoblot analysis, the WP5212-related proteins were expected to make up the majority of the wheat embryo-enriched globulin fraction. Six spots ranging in M_r from approximately 28kDa

to 50kDa and ranging in pI from 3 to 9.75 were chosen for analysis (Figure 3.5). Whether these polypeptides represented different gene products or whether they were derived from a common, larger precursor was investigated. Mass spectrometry and N-terminal sequencing proceeded successfully generating useful proteomic data.

Results from the interrogation of the NCBI database are presented in Table 3.1. The searches consistently identified globulin1-related proteins Beg1 (identical to Gbl1) and Glb1 from maize, except for spot 3 which could not be identified. It should be noted that the M_r and pI of the maize Glb1 gene is not an accurate representation of the protein. The genomic sequence deposited only represents a 1.2Kb portion of the locus. The ORF thus represents less than half the length of the primary translation product (160). The poor sequence coverage observed in Table 3.1 is likely a combination of effects. Firstly, the Glb1 gene is from another organism, maize, a distantly-related member of the Gramineae (161). Therefore, many of the maize Glb1 (and potentially barley Beg1) *in silico* digestion products may differ from the observed wheat peptide masses. Secondly, the spots chosen for sequencing were of significantly lower M_r than the full-length Beg1 sequence matched, reducing percent coverage. Related proteins derived from multigene families and allelic variation that occurs between species might also affect the results. Finally, PTMs such as glycosylation and acetylation, chemical modifications during sample preparation and enzymatic digestion can affect the peptide masses obtained by LC-MS/MS, also reducing sequence coverage (157).

Given that all spots submitted for LC-MS/MS were identified as globulin1-related, the 7S globulin sequences in Figure A.2 were used to construct a 7S globulin database that was also interrogated with the LC-MS/MS ions search data: results are presented in Table 3.2. Slightly higher protein coverage was obtained for these results, although this was often a function of sequence length. Nonetheless, new peptides were matched that had not been previously identified. The location of these peptides

relative to the interrogated sequences is shown in Figure A.2. This database search also provided two new results:

1) Spot 3 was found to have a single peptide matching globulin1-related sequences. The identification of this single peptide was not significant when the NCBI database search was performed, as the individual ions score was below the cutoff at $p<0.05$. By creating a wheat 7S globulin database, the number of peptide masses searched was reduced and an acceptable ions score was obtained for this match (http://www.matrixscience.com/help/interpretation_help.html, last viewed 7/6/06). Spot 3 could be tentatively identified as a 7S globulin.

2) Spot 1 was found to have a peptide matching a Glb2-like sequence (AFPQQARE) from the wheat EST database, as well as multiple peptides matching Glb1-like sequences. Spot 1 can be said with confidence ($p<0.05$) to contain a Glb1-related protein. However, the peptide match corresponding to a Glb2-like sequence, AFPQQARE, suggests that another, possibly less abundant protein may also be present. It is possible, however, that this match is random and that the peptide actually belongs to the Glb1-related protein. Post-translational and chemical modifications can cause changes in peptide masses leading to false identification (157). It is also possible that the identified peptide could correspond to a yet unidentified member of the globulin1-related proteins, sharing the identified sequence with Glb2-like proteins. However, the corresponding peptides from known Glb1-like sequences, TFGRPARE and AFGRPARE, differ sufficiently from AFPQQARE to suggest otherwise. The context in which the match was made should also be considered. The Glb2 and Reg2 proteins are known to occur in the salt-soluble fraction in maize and rice seed embryos, respectively (83, 85, 87); and transcripts homologous to Glb2/Reg2 are found in both barley and wheat seed tissues ((39) and present study). Since the origin of the sample is a salt-soluble fraction from wheat embryos, it is not unreasonable for a Glb2-like match to have been made.

To gain more information about the nature of the Glb1-related proteins, N-terminal sequencing was performed. For economical reasons, four of the six spots from Figure 3.5 and Tables 3.1 and 3.2 were chosen for N-terminal sequencing. Spot 3, with a M_r of approximately 30kDa was chosen because its LC-MS/MS identification as a Glb1-related protein was tentative. The other spots were chosen as a sampling of the different M_r and pI of the globulin fraction. No sequence data were obtained from the 50kDa proteins. The ~30kDa proteins, however, provided useful information (Table 3.3). Firstly, spot 3 was confirmed to be a Glb1-related protein with an internal sequence as its N-terminus (Figure A.2). Secondly, spot 5 which has a similar M_r but a more basic pI, was found to have a different N-terminus which also corresponded to an internal peptide from the Glb1-like sequences (Figure A.2). These findings suggest that the ~30kDa globulin-like polypeptides could arise from different processing events of the same precursor, as was observed for barley Beg1 (39, 86).

Wheat 7S globulins: putting the pieces together

The N-terminal sequence data obtained from this study along with those from two other wheat proteomic studies were consolidated (Table 4.1). Combined, the information provides insight into the nature of the various proteins that make up the Glb1-related protein cluster in wheat. Figure 4.1 is a summary of the potential modifications to Glb1-related sequences. WP5212 was used as the model because it is the only sequence corroborated by independent data (GG2.0: accession 8718.1), representing a full-length globulin1 homologue. Once endoproteololytic processing (EPP) sites could be established, the nature of the sequences in Table 4.1 could be hypothesized. The model suggests that the globulin1-related proteins of varying molecular masses and isoelectric points are not derived from separate genes. Rather, they originate from large 7S globulin precursors possessing a signal peptide, an N-terminal segment and a vicilin domain (Figure 4.1). These precursors may belong to a multigene family whose members differ slightly in sequence, with varying length N-terminal segments. Figure 4.1 can be summarised as follows. The globulin1 protein, prior to

Table 4.1 Pooled N-terminal sequence data for Glb1-related proteins and potential processing events in the maturation of Glb1-related polypeptides. N-terminal sequence data for Glb1-related polypeptides were collected from three separate studies. When available, the polypeptide's observed molecular mass (M_r) and isoelectric point (pI) are presented. Based on these N-terminal sequences and using the WP5212 sequence as a representative protein, a map of potential cleavage events for wheat Glb1-related proteins was constructed (Figure 4.1). For each polypeptide, a potential set of cleavage events (based on Figure 4.1) is provided that could result in a theoretically similar-sized protein. The predicted M_r and pI of theoretical cleavage products are provided, calculated by the ExPASy Compute pI/ M_r tool (http://www.expasy.ch/tools/pi_tool.html, last viewed 3/6/06). ND: not determined.

Study	Observed M_r/pI	N-terminus	Predicted polypeptide	Theoretical M_r/pI of cleavage products
Singh, *et al.* 2001	20kDa/ ND	DERQERRRGS	Polypeptide C-terminal to cleavage site 4	17kDa/5.84
Singh, *et al.* 2001	54kDa/ ND	XEHGDXRRRPYP	Polypeptide C-terminal to cleavage site 2	52kDa/8.49
Singh, *et al.* 2001	37kDa/ ND	XXHGDDRRPYV	Polypeptide between cleavage sites 2 and 4	35kDa/9.14
Singh, *et al.* 2001	38kDa/ ND	XXHGDSRRPYV	Polypeptide between cleavage sites 2 and 4	35kDa/9.14
Skylas, *et al.* 2000	13-15kDa/ 7	SHDEEEDRRR	Polypeptide between cleavage sites 1 and 2	12kDa/6.59
Present study (spot 3)	28-30kDa/ 6.5	SRDTFNLL	Polypeptide C-terminal to cleavage site 3	29kDa/5.76
Present study (spot 5)	30-33kDa/ 9.75	XXHGDSRR	Polypeptide between cleavage sites 2 and 4	35kDa/9.14

Figure 4.1 Features of the WP5212 sequence and potential processing events. The WP5212 sequence is shown as a representative member of the family of Glb1-related proteins in wheat. Features shown are likely common to all Glb1-related proteins though specific sequences may differ among family members.

Panel **A)** The full-length primary translation product is represented at the amino acid level. A signal sequence directing the protein into the secretory pathway is italicized; the predicted cleavage site is indicated by a hollow arrow (1) with signal sequence prediction by TargetP1.1 (162), (http://www.cbs.dtu.dk/services/TargetP/, last viewed 3/6/06). Repetitive tracts of hydrophilic residues, also found in barley Beg1, are underlined (39). These charged motifs are found in storage globulins of other plants, but the sequences are not conserved. A repetitive motif, CXXXC, common to the N-terminal segments of large 7S globulins, is boxed (31). Peptides representing the mature N-termini of Glb1-like polypeptides, identified in this study and previous studies (39, 116, 117) are bolded. Potential cleavage sites, based on the N-terminal sequences, are represented by black arrows, labeled 2-4. The holoprotein is the sequence immediately following cleavage site 1. The polypeptide contained between cleavage sites 1 and 2 represents the N-terminal segment (39). The region C-terminal to cleavage site 2 represents the vicilin domain (23, 31). The only potential site for N-linked glycosylation, NIT, is double-underlined (39).

Panel **B)** A diagram of full-length WP5212 is shown with approximate locations of potential cleavage sites, labeled 1-4, corresponding to those described in A). The protein domains are represented as follows: signal peptide (white); N-terminal segment (grey); vicilin domain (black). The putative N-linked glycosylation site (52, 163, 164) is represented by an asterisk. Arrows beneath the diagram give the predicted fragments resulting from potential cleavage events. The molecular mass (kDa) and isoelectric point (pI) of the resultant polypeptides is indicated on the right. M_r and pI prediction was performed using the ExPASy Compute pI/ M_r tool (http://www.expasy.ch/tools/pi_tool.html, last viewed 3/6/06): the putative glycosylation is not taken into account. The check marks represent hypothetical endoproteolytic products that could result in the observed polypeptides of Table 4.1, based on their observed N-terminal sequence, M_r and pI. The diagram is not drawn to scale.

A

*MATRGRATIPLLFLLGTSLLFAAAVSA***SHDEEEDRRG**GRSLQRCVQRCQ

QDRPRYSHARCVQECRDDQQQHGRHEQEEQGRGHGRHGEGEREEEQGRG

RGRRGQGEREEEQGRGRGRRGEGERD**EEHGDGRRPY**VFGPRSFRRIIRS

DHGFVKALRPFDEVSRLLRGIRNYRVAIMEVNPRAFVVPGLTDADGVGY

VAQGEGVLTVIENGEKRSYTVRQGDVIVAPAGSIMHLANTDGRRKLVIA

KILHTISVPGKFQYFSAKPLLASLSKRVLTAALKTSDERLGSLLGSRQG

KEEEEKSISIVRASEEQLRELRRQASEGDQGHHWPLPPFRGD**SRDTFNL**

LEQRPKIANRHGRLYEADARSFHALAQHDVRVAVANITPGSMTAPYLNT

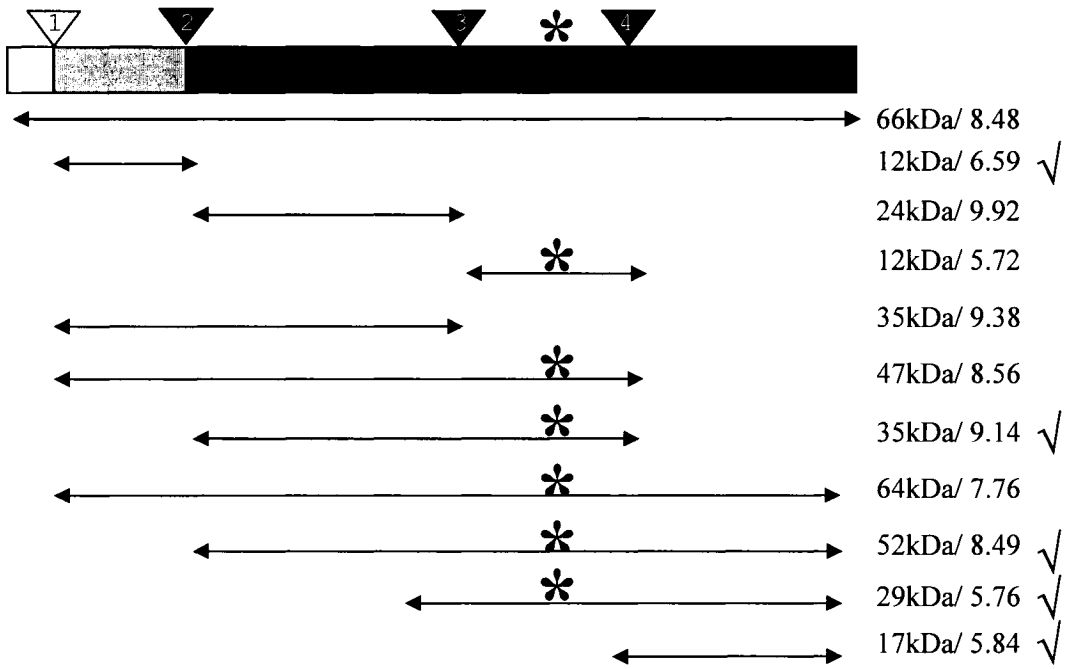
QSFKLAVVLEGEGEVEIVCPHLGRDSERREQEHGKGRWRSEEEED**DRRQ**

QRRRGSGSESEEEQDQQRYETVRARVSRGSAFVVPPGHPVVEIASSRGS

SNLQVVCFEINAERNERVWLAGRNNVIAKLDDPAQELAFGRPAREVQEV

FRAKDQQDEGFVAGPEQQQEHERGDRRRGDRGRGDEAVEAFLRMATAAL

B



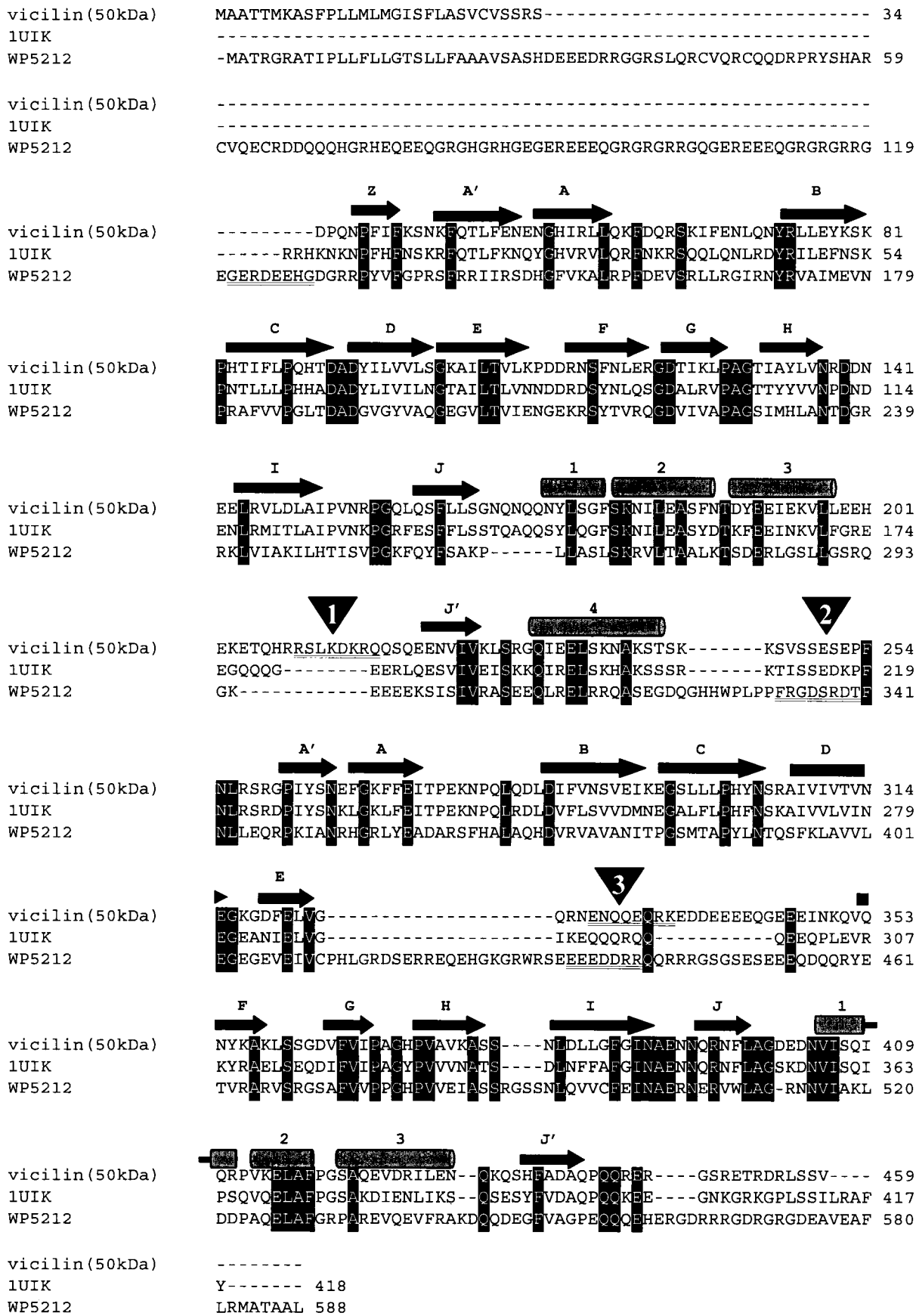
co-translational removal of the signal sequence, is a ~66kDa monomer. After signal sequence removal (cleavage site 1), the protein becomes a ~64kDa holoprotein with a new N-terminus (bolded). Three more internal N-terminal sequences have been identified for globulin1-like sequences in wheat. The homologous sequences in WP5212 are also bolded and the cleavage sites, 2-4, are inferred. In table 4.1, the predicted cleavage events often generated hypothetical polypeptides that had predicted M_r and pI similar to those observed.

Maturation of the 7S globulins within the trimeric fold

The processing sites in Figure 4.1 are approximate, as the nature of the EPP events is undetermined. However, knowledge of similar processing events for pea vicilin suggests that the proteolytic events occur in highly polar regions, after trimeric assembly of the individual subunits (43). Thus, the processing sites must be located in enzyme-accessible locations within the three-dimensional structure. Using the available three-dimensional data from the 7S globulin of soybean (β-conglycinin α'), the potential cleavage events for pea vicilin and WP5212 were modeled to homologous regions of the β-conglycinin structure.

The relative locations of potential cleavage sites of WP5212 (Figure 4.1) and the pea 50kDa vicilin (72, 74) within the trimeric complex, were approximated based on their alignment with the β-conglycinin-α' core domain (Figure 4.2). Since no three-dimensional structure exists for a full-length large 7S globulin, it was not possible to show the N-terminal segment of the WP5212 polypeptide (46) in relation to the β-conglycinin-α' N-terminal segment.

Figure 4.2 Structure-based sequence alignment of the wheat WP5212, pea 50kDa vicilin and soybean β-conglycinin-α' sequences. The WP5212 and pea 50kDa vicilin sequences were aligned to the sequence for the core β-conglycinin-α' subunit, for which a three-dimensional structure is available (46). The core sequence does not include the N-terminal segment of the native β-conglycinin-α' polypeptide. Identical residues are highlighted in black. Approximate locations of potential cleavage sites of WP5212, based on Figure 4.1 (present study) and pea vicilin data (72, 74) are double underlined and labeled with numbered black triangles: 1, 2 and 3. These sites are mapped to the three-dimensional structure of the core β-conglycinin-α' subunit in Figure 4.3. The first proposed cleavage site of WP5212 mapped to the end of the N-terminal segment, a region of the sequence for which 3D structure information is unavailable. This site was not numbered as it could not be shown in Figure 4.3. Signal sequence cleavage events are also omitted. Cleavage site 1 is unique to pea vicilin; cleavage site 2 is unique to WP5212; cleavage site 3 occurs in homologous regions of both sequences. The secondary structure elements, occurring along the β-conglycinin-α' polypeptide chain, are overlaid on the alignment and labeled according to convention (23, 44, 46). β-strands are represented by arrows; α helices, by cylinders. Helix 4 is part of the linker region that connects the structurally similar N- and C-terminal modules. The alignment was performed using ClustalX (153).


vicilin(50kDa) MAATTMKASFPLLMLMGISFLASVCVSSRS---------- 34
1UIK ----------
WP5212 -MATRGRATIPLLFLLGTSLLFAAAVSASHDEEEDRRGGRSLQRCVQRCQQDRPRYSHAR 59
vicilin(50kDa) ----------
1UIK ----------
WP5212 CVQECRDDQQQHGRHEQEEQGRGHGRHGEGEREEEQGRGRGRRGQGEREEEQGRGRGRRG 119
Z A′ A B
vicilin(50kDa) ---------DPQNPFIFKSNKFQTLFENENGHIRLLQKFDQRSKIFENLQNYRLLEYKSK 81
1UIK ------RRHKNKNPFHFNSKRFQTLFKNQYGHVRVLQRFNKRSQQLQNLRDYRILEFNSK 54
WP5212 EGERDEEHGDGRRPYVFGPRSFRRIIRSDHGFVKALRPFDEVSRLLRGIRNYRVAIMEVN 179
C D E F G H
vicilin(50kDa) PHTIFLPQHTDADYILVVLSGKAILTVLKPDDRNSFNLERGDTIKLPAGTIAYLVNRDDN 141
1UIK PNTLLLPHHADADYLIVILNGTAILTLVNNDDRDSYNLQSGDALRVPAGTTYYVVNPDND 114
WP5212 PRAFVVPGLTDADGVGYVAQGEGVLTVIENGEKRSYTVRQGDVIVAPAGSIMHLANTDGR 239
I J 1 2 3
vicilin(50kDa) EELRVLDLAIPVNRPGQLQSFLLSGNQNQQNYLSGFSKNILEASFNTDYEEIEKVLLEEH 201
1UIK ENLRMITLAIPVNKPGRFESFFLSSTQAQQSYLQGFSKNILEASYDTKFEEINKVLFGRE 174
WP5212 RKLVIAKILHTISVPGKFQYFSAKP------LLASLSKRVLTAALKTSDERLGSLLGSRQ 293
1 J′ 4 2
vicilin(50kDa) EKETQHRRSLKDKRQQSQEENVIVKLSRGQIEELSKNAKSTSK-------KSVSSESEPF 254
1UIK EGQQQG--------EERLQESVIVEISKKQIRELSKHAKSSSR-------KTISSEDKPF 219
WP5212 GK------------EEEEKSISIVRASEEQLRELRRQASEGDQGHHWPLPPFRGDSRDTF 341
A′ A B C D
vicilin(50kDa) NLRSRGPIYSNEFGKFFEITPEKNPQLQDLDIFVNSVEIKEGSLLLPHYNSRAIVIVTVN 314
1UIK NLRSRDPIYSNKLGKLFEITPEKNPQLRDLDVFLSVVDMNEGALFLPHFNSKAIVVLVIN 279
WP5212 NLLEQRPKIANRHGRLYEADARSFHALAQHDVRVAVANITPGSMTAPYLNTQSFKLAVVL 401
E 3
vicilin(50kDa) EGKGDFELVG----------QRNENQQEQRKEDDEEEEQGEEEINKQVQ 353
1UIK EGEANIELVG----------IKEQQQRQQ----------QEEQPLEVR 307
WP5212 EGEGEVEIVCPHLGRDSERREQEHGKGRWRSEEEEDDRRQQRRRGSGSESEEEQDQQRYE 461
F G H I J 1
vicilin(50kDa) NYKAKLSSGDVFVIPAGHPVAVKASS----NLDLLGFGINAENNQRNFLAGDEDNVISQI 409
1UIK KYRAELSEQDIFVIPAGYPVVVNATS----DLNFFAFGINAENNQRNFLAGSKDNVISQI 363
WP5212 TVRARVSRGSAFVVPPGHPVVEIASSRGSSNLQVVCFEINAERNERVWLAG-RNNVIAKL 520
2 3 J′
vicilin(50kDa) QRPVKELAFPGSAQEVDRILEN--QKQSHFADAQPQQREE----GSRETRDRLSSV---- 459
1UIK PSQVQELAFPGSAKDIENLIKS--QSESYFVDAQPQQKEE----GNKGRKGPLSSILRAF 417
WP5212 DDPAQELAFGRPAREVQEVFRAKDQQDEGFVAGPEQQQEHERGDRRRGDRGRGDEAVEAF 580
vicilin(50kDa) --------
1UIK Y------- 418
WP5212 LRMATAAL 588

Two of the predicted proteolytic processing sites (2 and 3), shown in Figure 4.3, occur in regions of the β-conglycinin-α' three-dimensional structure that could not be determined because of diffuse main chain electron density (46). These sites are emphasized in panel C by dashed lines. Modeling of the potential cleavage sites lends strength to the data collected for pea vicilins that processing occurs after formation of the 7S trimeric complex and transfer to protein bodies (43, 59). Each of the processing sites occurs on the surface of the folded protein, in regions represented by a loop or by a highly disordered main chain structure. Cleavage sites 1 and 3 are present in regions of the vicilin and WP5212 sequences that have insertions relative to the β-conglycinin-α' core sequence. Increases in the size of these regions have the potential to increase their susceptibility to processing enzymes.

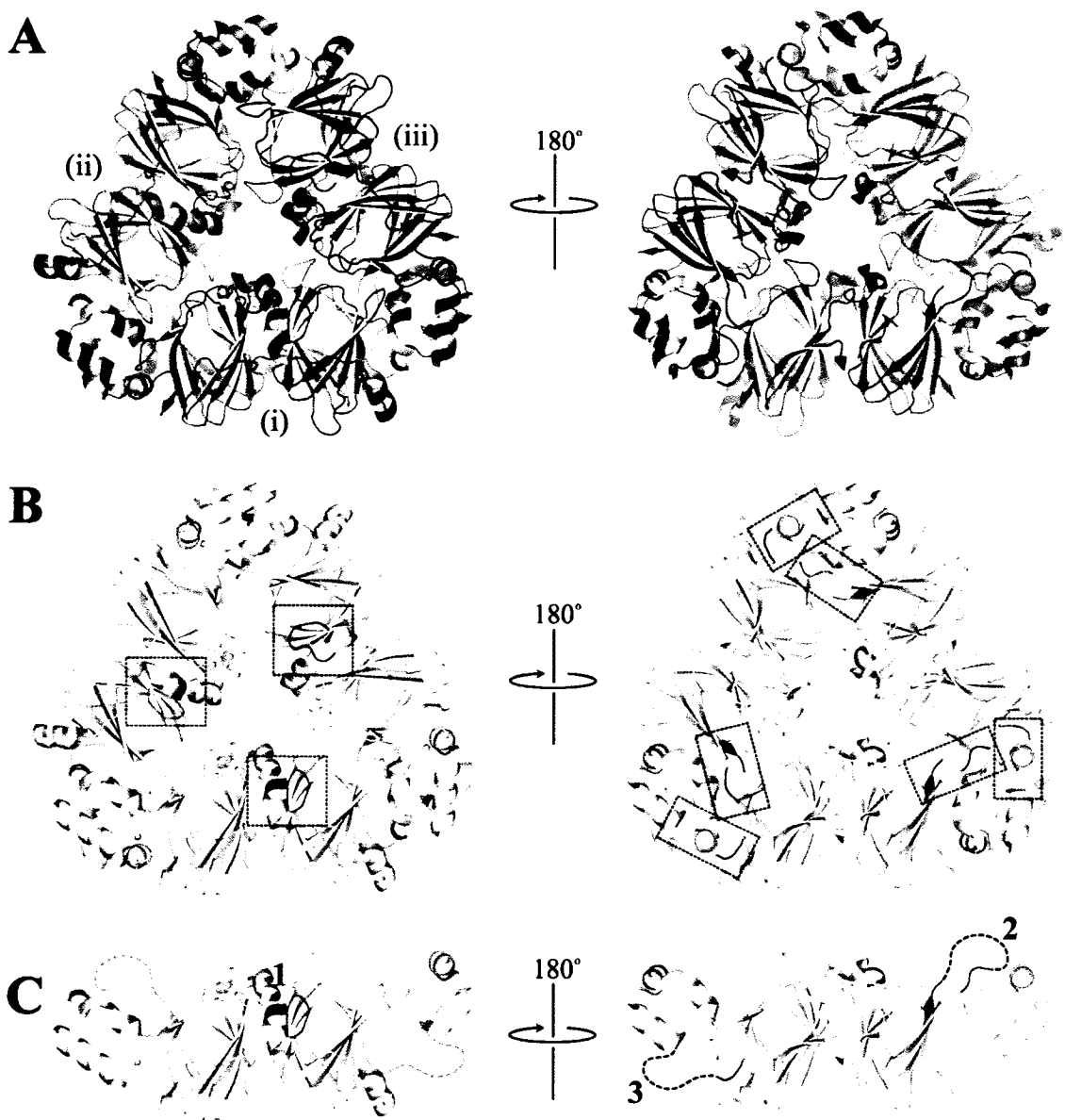
It is not yet known what dictates the proteolytic cleavage of certain 7S globulins into smaller peptides (39, 59, 87) while others remain intact (30, 67, 83). However, processing does appear to occur only in highly polar regions of the polypeptide, where specific sequences are expected to permit, or prevent cleavage (68, 72). Furthermore, proteolytic processing occurs only in a fraction of the susceptible polypeptides within the trimer (39, 43). The heterogeneity of the observed polypeptides has not yet been explained, though it is probable that the chemical environment of the PSV contributes to these processing events to some extent. *In vitro* studies of the 7S globulin phaseolin have shown the protein to oligomerize under acidic conditions (166). If this phenomenon also occurs *in vivo* within the acidic PSV, enzyme accessibility to potential cleavage sites could be reduced. Low enzyme levels in the PSV may also play a role (30). Though the 7S globulin three-dimensional structures provide insight into the nature of the cleavage sites, the exact, complete native conformations of the various trimers, and their accessibility within the protein body, are unknown. The evolutionarily-selected conformation of the 7S globulins and the EPP events involved in their maturation may have favoured easier proteolysis during the early phases of germination, propelling the embryo's growth and success.

Figure 4.3 Structural representation of potential cleavage sites in 7S globulins based on the three-dimensional structure of the β-conglycinin-α' homotrimer core. The right-hand structures are views of the left-hand structures upon 180° rotation. The structure was obtained from the Protein Data Bank (PDB) http://www.rcsb.org/pdb/Welcome.do, last viewed 13/6/06.

Panel **A** is a ribbon diagram of the three-dimensional structure for the β-conglycinin-α' homotrimer core (PDB ID: 1UIK) is shown in panel **A**. Each monomer in the homotrimer has been assigned a different colour: i) red; ii) green; iii) blue. The β-conglycinin-α' monomers shown represent the core domain, lacking the 141 amino acid N-terminal segment of the native protein. The core domain, also referred to as the vicilin domain, consists of two similar modules related by a pseudo-dyad. Each module consists of an elaborated cupin domain associated with an alpha helical loop domain (46).

Panel **B** shows the β-conglycinin-α' homotrimer subunits coloured grey. Based on the alignment in Figure 4.2, potential cleavage sites determined for WP5212 and the pea 50kDa vicilin were mapped to homologous regions on the β-conglycinin-α' core monomer. The cleavage sites are coloured red, green or blue, according to their respective monomer in panel A. Boxes are drawn around the potential cleavage sites to highlight their locations.

Panel **C** is the structural representation of the single β-conglycinin-α' core monomer (i) shown in grey. Dashed lines (red) were drawn to show potential cleavage regions where density was lacking in the three-dimensional structure due to disorder of the polypeptide (46). The dashed lines are not an accurate representation of the position and length of the polypeptide backbone; rather they are used to emphasize the location of the potential cleavage sites. Cleavage sites 1, 2 and 3 represent those indicated in Figure 4.2. Panels B and C try to illustrate the location of these sites on the outer surfaces of the 7S globulin trimer, in disordered regions which appear to protrude from the core structure. Structural visualization was performed using PyMOL (165).


A
(ii)
(iii)
(i)
180°
B
180°
C
1
180°
2
3

Other potential post-translational modifications

The EPP events outlined in Figures 4.1-4.3 are likely part of a series of post-translational events that lead to the maturation of the globulin1-like proteins, similar to those observed for barley, Beg1 (39). Other PTMs could also play a role. Glycosylation is frequently observed in 7S globulins as N-linked complex glycans (52, 163), but is not required for proper folding or export to the PSV (43). Other cleavage events may also occur at the N-terminus (prior to cleavage site 2, Figure 4.1) as observed for maize Glb1 (84), or may include further C-terminal trimming. The mass spectrometry data collected in this study and knowledge of the gene structure of cereal globulin1 homologues does not support alternative splicing as a means of generating different globulin1-related polypeptides (90).

The identification of wheat globulin1 holoproteins (~65-70kDa) by immunoblot analysis suggests that the proteolytic modifications described are partial and may not be prerequisites for proper folding, transport and storage of the globulin1 proteins. The presence of polypeptides of ~50kDa and ~30-38kDa ((117) and present study), with the same N-termini reinforce the idea that not necessarily every hypothesized set of EPP events occurs, and some may occur to different extents.

In a previous study, the persistence of a small polypeptide representing the N-terminus of the wheat globulin1 holoprotein (Table 4.1), was found to occur (116). No smaller M_r proteins were identified in the present study, but this is explained, at least in part, by fact that the antibodies used to probe the globulin fraction were raised against two linear epitopes from the vicilin domain of WP5212: SRDTFNLLEQRPKIAN and RGDEAVEAFLRMATA (145). It is also possible that the different Glb1-like polypeptides derived from holoproteins have slightly different solubilities, dissolving to different extents in 1M NaCl, or being soluble in ddH_20.

Summary

The WP5212 antigenically-homologous proteins make up the major portion of the globulin fraction from the wheat embryo, but are present in low quantities in that of the endosperm. Some of these proteins have been definitively identified as globulin1 homologues by mass spectrometry and N-terminal sequencing. Separation of the proteins by 2DE has allowed the detection of at least one new polypeptide, not previously identified: M_r ~28-30kDa; pI ~6.5; N-terminus is SRDTFNLL.

The data suggest that the Glb1-related proteins of wheat are large 7S globulins, synthesized as preproproteins, that undergo a series of post-translational modifications which include limited endoproteolytic events. This data is corroborated by similar processing events in globulin1 homologues of other cereals (39, 87) and in the vicilins of pea (68, 72, 73). The wheat 7S globulin sequences have putative N-linked glycosylation sites. A previous study has shown that the major size groups in the wheat globulin 7S fraction bind concanavalinA, although the exact nature and extent of glycosylation is unknown (105). These PTMs, among others, may contribute to the microheterogeneity of the observed isoelectric points as well as the wide range of observed molecular masses among the gloublin1-related proteins.

The globulins from wheat embryos are still largely uncharacterized as a group and the lack of sequence data is a limiting factor in their characterization. The expression pattern of the globulin1 antigenically-related proteins is in part the result of post-translational modifications, but may also arise from a small multigene family on the arms of homoeologous chromosomes 4AL, 4BS and 4DS (88).

A Glb2-related protein has been tentatively identified within the wheat globulin fraction by LC-MS/MS. While transcript sequences have been deposited to EST databases, to this author's knowledge, this is the first time a wheat Glb2-like protein has been identified in the wheat proteome.

There is still much that is unknown about the 7S globulins, particularly the large 7S globulins for which no complete 3D structure is available (46). In some of the large 7S globulins studied (36, 37), the majority of the holoproteins experience almost complete removal of their N-terminal segment. However, the persistence of small amounts of full-size holoproteins suggests that the N-terminal segment does not interfere with trimer assembly and storage. Other large 7S globulins (β-conglycinin α' subunits and convicilin) do not experience this N-terminal segment removal and are still folded correctly (26). Why some large 7S globulins experience this processing while others do not is uncertain, but it is probably a function of their specific amino acid sequences.

There is also uncertainty regarding the fate of the propeptides that are cleaved from the 7S globulin N-terminal segments either during seed maturation or during germination. Preliminary studies suggest these may play a role in seed defence. N-terminal segment-derived peptides have been shown to exhibit trypsin inhibitory (37) and antimicrobial properties (120).

Current and Future directions

Characterization of the 7S storage globulins from wheat is at a very early stage. Distinguishing between isoforms derived from different PTMs, members of multigene families, allelic variants and microheterogeneity due to chemical modifications during sample preparation, is currently a difficult task, hampered by the limited sequence data available from wheat. Sequencing of the wheat hexaploid genome is still largely incomplete (161). The task is daunting because of wheat's large genome size (16-17Gb) and its high frequency of repetitive elements (90%) (167). As a result, only a limited number of wheat sequences have been deposited to the publicly available databases, making

mass spectrometry a restricted application in the study of the wheat proteins, especially those that may possess multiple, closely-related family members. Our lab is working on characterizing the globulin1-related sequences at the genomic level. Screening of a hexaploid wheat BAC library identified at least three genes that have open reading frames sharing strong identity with the known *WP5212* sequence (personal communication with Evelin Loit).

Immediate work at the proteomic level will focus on positively identifying the 65-70kDa antigenically-related proteins by mass spectrometry and N-terminal sequencing. Smaller globulin1-related peptides will also be sought, to obtain a more comprehensive globulin1 profile.

WP5212-related sequences have been mapped to the chromosomes 4AL, 4BS and 4DS in hexaploid wheat variety Chinese Spring (88) through the use of nullisomic-tetrasomic lines (168). However, it has not been shown that globulin1-related proteins are derived from the genes on these chromosomes. Separation of the globulins from each aneuploid line by 2DE may simplify the globulin profile and help map the proteins to their respective chromosomes.

Finally, the tentative identification of a second type of 7S globulin in wheat, Glb2-like, deserves more attention. The Glb2 and Reg2 proteins in maize and rice, respectively, are major embryonal globulins (85, 87). It will be interesting to see if globulin2 homologues also accumulate to high levels in wheat embryos, or whether these are minor proteins in comparison to the globulin1-related proteins.

As our understanding of the wheat 7S globulin profile improves, the downstream consequences on flour milling and dietary exposure to putative diabetogens can be pursued with greater accuracy.

Works Cited

1. Shotwell, M.A. and Larkins, B.A. 1989. The biochemistry and molecular biology of seed storage proteins, In *The Biochemistry of Plants,* Marcus, A., Ed.; Academic Press, Inc.: San Diego, Vol.15 pp. 297-344.

2. Sathe, S.K. 2002. Dry bean protein functionality. *Crit. Rev. Biotechnol.* **22**: 175-223.

3. Shewry, P.R. and Halford, N.G. 2002. Cereal seed storage proteins: structures, properties and role in grain utilization. *J. Exp. Bot.* **53**: 947-958.

4. Osborne, T.B. 1924. *The vegetable proteins.* Longmans, Green & Co.: London; pp. 154.

5. Anderson, O.D.; Rausch, C.; Moullet, O.; Lagudah, E.S. 2003. The wheat D-genome HMW-glutenin locus: BAC sequencing, gene distribution, and retrotransposon clusters. *Funct. Integr. Genomics.* **3**: 56-68.

6. Gu, Y.Q.; Coleman-Derr, D.; Kong, X.; Anderson, O.D. 2004. Rapid genome evolution revealed by comparative sequence analysis of orthologous regions from four triticeae genomes. *Plant Physiol.* **135**: 459-470.

7. Kong, X.Y.; Gu, Y.Q.; You, F.M.; Dubcovsky, J.; Anderson, O.D. 2004. Dynamics of the evolution of orthologous and paralogous portions of a complex locus region in two genomes of allopolyploid wheat. *Plant Mol. Biol.* **54**: 55-69.

8. Woo, Y.M.; Hu, D.W.; Larkins, B.A.; Jung, R. 2001. Genomics analysis of genes expressed in maize endosperm identifies novel seed proteins and clarifies patterns of zein gene expression. *Plant Cell.* **13**: 2297-2317.

9. Shorrosh, B.S.; Wen, L.; Zen, K.C.; Huang, J.K.; Pan, J.S.; Hermodson, M.A.; Tanaka, K.; Muthukrishnan, S.; Reeck, G.R. 1992. A novel cereal storage protein: molecular genetics of the 19 kDa globulin of rice. *Plant Mol. Biol.* **18**: 151-154.

10. Zhao, W.M.; Gatehouse, J.A.; Boulter, D. 1983. The purification and partial amino acid sequence of a polypeptide from the glutelin fraction of rice grains - homology to pea legumin. *FEBS Lett.* **162**: 96-102.

11. Robert, L.S.; Nozzolillo, C.; Altosaar, I. 1985. Homology between rice glutelin and oat 12S globulin. *Biochim. Biophys. Acta.* **829**: 19-26.

12. Shutov, A.D.; Kakhovskaya, I.A.; Braun, H.; Baumlein, H.; Muntz, K. 1995. Legumin-like and vicilin-like seed storage proteins: evidence for a common single-domain ancestral gene. *J. Mol. Evol.* **41**: 1057-1069.

13. Shutov, A.D. and Baumlein, H. 1999. Origin and evolution of seed storage proteins, In *Seed proteins.* Shewry, P.R. and Casey, R., Eds.; Kluwer Academic Publishers: Dordrecht; pp. 543-561.

14. Hager, K. and Fischer, H. 1999. Molecular phylogenies and structural diversification of gymnosperm and angiosperm storage globulins, In *Seed Proteins,* Shewry, P.R. and Casey, R., Eds.; Kluwer Academic Publishers: Dordrecht; pp. 499-515.

15. Danielsson, C.E. 1949. Seed globulins of the Gramineae and Leguminoseae. *Biochem J.* **44**: 387-100.

16. Dunwell, J.M. 1998. Cupins: a new superfamily of functionally diverse proteins that include germins and plant storage proteins. *Biotechnol. Genet. Eng. Rev.* **15**: 1-32.

17. Dunwell, J.M.; Culham, A.; Carter, C.E.; Sosa-Aguirre, C.R.; Goodenough, P.W. 2001. Evolution of functional diversity in the cupin superfamily. *Trends Biochem. Sci.* **26**: 740-746.

18. Lane, B.G.; Bernier, F.; Dratewka-Kos, E.; Shafai, R.; Kennedy, T.D.; Pyne, C.; Munro, J.R.; Vaughan, T.; Walters, D.; Altomare, F. 1991. Homologies between members of the germin gene family in hexaploid wheat and similarities between these wheat germins and certain Physarum spherulins. *J. Biol. Chem.* **266**: 10461-10469.

19. Argos, P.; Narayana, S.V.; Nielsen, N.C. 1985. Structural similarity between legumin and vicilin storage proteins from legumes. *EMBO J.* **4**: 1111-1117.

20. Gibbs, P.E.; Strongin, K.B.; McPherson, A. 1989. Evolution of legume seed storage proteins-a domain common to legumins and vicilins is duplicated in vicilins. *Mol. Biol. Evol.* **6**: 614-623.

21. Baumlein, H.; Braun, H.; Kakhovskaya, I.A.; Shutov, A.D. 1995. Seed storage proteins of spermatophytes share a common ancestor with desiccation proteins of fungi. *J. Mol. Evol.* **41**: 1070-1075.

22. Shutov, A.D.; Braun, H.; Chesnokov, Y.V.; Baumlein, H. 1998. A gene encoding a vicilin-like protein is specifically expressed in fern spores. Evolutionary pathway of seed storage globulins. *Eur. J. Biochem.* **252**: 79-89.

23. Lawrence, M.C. 1999. Structural relationships of the 7S and 11S globulins, In *Seed Proteins,* Shewry, P.R. and Casey, R., Eds.; Kluwer Academic Publishers: Dordrecht; pp. 517-541.

24. Grimes, H.D.; Overvoorde, P.J.; Ripp, K.; Franceschi, V.R.; Hitz, W.D. 1992. A 62-kD sucrose binding protein is expressed and localized in tissues actively engaged in sucrose transport. *Plant Cell.* **4**: 1561-1574.

25. Shutov, A.D.; Baumlein, H.; Blattner, F.R.; Muntz, K. 2003. Storage and mobilization as antagonistic functional constraints on seed storage globulin evolution. *J. Exp. Bot.* **54**: 1645-1654.

26. Casey, R. 1999. Distribution and some properties of seed globulins. In *Seed Proteins,* Shewry, P.R. and Casey, R., Eds.; Kluwer Academic Publishers: Dordrecht; pp. 159-169.

27. Yamagata, T.; Kato, H.; Kuroda, S.; Abe, S.; Davies, E. 2003. Uncleaved legumin in developing maize endosperm: identification, accumulation and putative subcellular localization. *J. Exp. Bot.* **54**: 913-922.

28. Adachi, M.; Takenaka, Y.; Gidamis, A.B.; Mikami, B.; Utsumi, S. 2001. Crystal structure of soybean proglycinin A1aB1b homotrimer. *J. Mol. Biol.* **305**: 291-305.

29. Croy, R.R.; Gatehouse, J.A.; Tyler, M.; Boulter, D. 1980. The purification and characterization of a third storage protein (convicilin) from the seeds of pea (*Pisum sativum* L.). *Biochem. J.* **191**: 509-516.

30. Chrispeels, M.J. 1984. Biosynthesis, procesing and transport of storage proteins and lectins in cotyledons of developing legume seeds. *Phil. Trans. R. Soc. Lond. B.* **304**: 309-322.

31. Dure, L. 1990. An unstable domain in the vicilin genes of higher plants. *New Biol.* **2**: 487-493.

32. Chrispeels, M.J. 1991. Sorting of proteins in the secretory system. *Annu. Rev. Plant Phys.* **42**: 21-53.

33. Doyle, J.J.; Schuler, M.A.; Godette, W.D.; Zenger, V.; Beachy, R.N.; Slightom, J.L. 1986. The glycosylated seed storage proteins of *Glycine max* and *Phaseolus vulgaris*. Structural homologies of genes and proteins. *J. Biol. Chem.* **261**: 9228-9238.

34. Borroto, K. and Dure L.III. 1987. The globulin seed storage proteins of flowering plants are derived from two ancestral genes. *Plant Mol. Biol.* **8**: 113-131.

35. Brown, D.; Ellis, T.H.N.; Gatehouse, J.A. 1988. The sequence of a gene encoding convicilin from pea (*Pisum sativum* L.) shows that convicilin differs from vicilin by an insertion near the N-terminus. *Biochem J.* **251**: 717-726.

36. Belanger, F.C. and Kriz, A.L. 1989. Molecular characterization of the major maize embryo globulin encoded by the *Glb1* gene. *Plant Physiol.* **91**: 636-643.

37. Yamada, K.; Shimada, T.; Kondo, M.; Nishimura, M.; Hara-Nishimura, I. 1999. Multiple functional proteins are produced by cleaving Asn-Gln bonds of a single precursor by vacuolar processing enzyme. *J. Biol. Chem.* **274**: 2563-2570.

38. Newbigin, E.J.; deLumen, B.O.; Chandler, P.M.; Gould, A.; Blagrove, R.J.; March, J.F.; Kortt, A.A.; Higgins, T.J. 1990. Pea convicilin: Structure and primary sequence of the protein and expression of a gene in the seeds of transgenic tobacco. *Planta.* **180**: 461-470.

39. Heck, G.R.; Chamberlain, A.K.; Ho, T.H.D. 1993. Barley embryo globulin 1 gene, *Beg1* - Characterization of cDNA, chromosome mapping and regulation of expression. *Molec. Gen. Genet.* **239**: 209-218.

40. Ko, T.P.; Day, J.; McPherson, A. 2000. The refined structure of canavalin from jack bean in two crystal forms at 2.1 and 2.0 A resolution. *Acta Crystallogr. D Biol. Crystallogr.* **56**: 411-420.

41. McPherson, A. Canavalin. 1999. In *Seed Proteins.* Shewry, P.R. and Casey, R., Eds.; Kluwer Academic Publishers: Dordrecht; pp. 241-257.

42. Ko, T.P.; Ng, J.D.; McPherson, A. 1993. The three-dimensional structure of canavalin from jack bean (*Canavalia ensiformis*). *Plant Physiol.* **101**: 729-744.

43. Chrispeels, M.J.; Higgins, T.J.; Spencer, D. 1982. Assembly of storage protein oligomers in the endoplasmic reticulum and processing of the polypeptides in the protein bodies of developing pea cotyledons. *J. Cell Biol.* **93**: 306-313.

44. Lawrence, M.C.; Izard, T.; Beuchat, M.; Blagrove, R.J.; Colman, P.M. 1994. Structure of phaseolin at 2.2 A resolution. Implications for a common vicilin/legumin structure and the genetic engineering of seed storage proteins. *J. Mol. Biol.* **238**: 748-776.

45. Maruyama, N.; Adachi, M.; Takahashi, K.; Yagasaki, K.; Kohno, M.; Takenaka, Y.; Okuda, E.; Nakagawa, S.; Mikami, B.; Utsumi, S. 2001. Crystal structures of recombinant and native soybean beta-conglycinin beta homotrimers. *Eur. J. Biochem.* **268**: 3595-3604.

46. Maruyama, Y.; Maruyama, N.; Mikami, B.; Utsumi, S. 2004. Structure of the core region of the soybean beta-conglycinin alpha' subunit. *Acta Crystallogr. D Biol. Crystallogr.* **60**: 289-297.

47. Bollini, R. and Chrispeels, M.J. 1979. The rough endoplasmic reticulum is the site of reserve protein synthesis in developing *Phaseolus vulgaris* cotyledons. *Planta.* **146**: 487-501.

48. Chrispeels, M.J.; Higgins, T.J.V.; Craig, S.; Spencer, D. 1982. Role of the endoplasmic reticulum in the synthesis of reserve proteins and the kinetics of their transport to protein bodies in developing pea cotyledons. *J. Cell Biol.* **93**: 5-14.

49. Vitale, A.; Ceriotti, A.; Denecke, J. 1993. The role of the endoplasmic reticulum in protein synthesis, modification and intracellular transport. *J. Exp. Bot.* **44**: 1471-1484.

50. Bollini, R.; Vitale, A.; Chrispeels, M.J. 1983. *In vivo* and *in vitro* processing of seed reserve protein in the endoplasmic reticulum: evidence for two glycosylation steps. *J. Cell Biol.* **96**: 999-1007.

51. Sengupta, C.; Deluca, V.; Bailey, D.S.; Verma, D.P.S. 1981. Posttranslational processing of 7S and 11S components of soybean storage proteins. *Plant Mol. Biol.* **1**: 19-34.

52. Sturm, A.; Van-Kuik, J.A.; Vliegenthart, J.F.; Chrispeels, M.J. 1987. Structure, positions and biosynthesis of the high mannose and the complex oligosaccharide side chains of the bean storage protein phaseolin. *J. Biol. Chem.* **262**: 13392-13403.

53. Pedrazzini, E.; Giovinazzo, G.; Bollini, R.; Ceriotti, A.; Vitale, A. 1994. Binding of BiP to an assembly-defective protein protein in plant cells. *Plant J.* **5**: 103-110.

54. Vitale, A.; Bielli, A.; Ceriotti, A. 1995. The binding protein associates with monomeric phaseolin. *Plant Physiol.* **107**: 1411-1418.

55. Hohl, I.; Robinson, D.G.; Chrispeels, M.J.; Hinz, G. 1996. Transport of storage proteins to the vacuole is mediated by vesicles without a clathrin coat. *J. Cell. Sci.* **109**: 2539-2550.

56. Hillmer, S.; Movafeghi, A.; Robinson, D.G.; Hinz, G. 2001. Vacuolar storage proteins are sorted in the cis-cisternae of the pea cotyledon Golgi apparatus. *J. Cell Biol.* **152**: 41-50.

57. Bollini, R.; Van der Wilden, W.; Chrispeels, M.J. 1982. A precursor of the reserve protein phaseolin is transiently associated with the endoplasmic reticulum of developing *Phaseolus vulgaris* cotyledons. *Physiol. Plant.* **55**: 82-92.

58. Sturm, A.; Johnson, K.D.; Szumilo, T.; Elbein, A.D.; Chrispeels, M.J. 1987. Subcellular localization of glycosidases and glycosyl transferases involved in the processing of N-linked oligosaccharides. *Plant Physiol.* **85**: 741-745.

59. Gatehouse, J.A.; Croy, R.R.D.; Morton, H.; Tyler, M.; Boulter, D. 1981. Characterization and subunit strcutures of the vicilin storage proteins of pea (*Pisum sativum* L.). *Eur. J. Biochem.* **118**: 627-633.

60. Casey, R.; Domoney, C.; Ellis, N. 1986. Legume storage proteins and their genes. In *Oxford surveys of plant molecular and cell biology.* Miflin, B.J., Ed.; Oxford University Press: New York; Vol.3 pp. 1-95.

61. Kriz, A.L. and Schwartz, D. 1986. Synthesis of globulins in maize embryos. *Plant Physiol.* **82**: 1069-1075.

62. Lioi, L. and Bollini, R. 1984. Contribution of processing events to the molecular heterogeneity of four banding types of phaseolin, the major storage protein of *Phaseolus vulgaris* L. *Plant Mol. Biol.* **3**: 345-353.

63. Slightom, J.L.; Drong, R.F.; Klassy, R.C.; Hoffman, L.M. 1985. Nucleotide sequences from phaseolin cDNA clones: the major storage proteins from Phaseolus vulgaris are encoded by two unique gene families. *Nucleic Acids Res.* **13**: 6483-6498.

64. Lelievre, J.M.; Dickinson, C.D.; Dickinson, L.A.; Nielsen, N.C. 1992. Synthesis and assembly of soybean beta-conglycinin *in vitro*. *Plant Mol. Biol.* **18**: 259-274.

65. Bustos, M.M.; Kalkan, F.A.; VandenBosch, K.A.; Hall, T.C. 1991. Differential accumulation of four phaseolin glycoforms in transgenic tobacco. *Plant Mol. Biol.* **16**: 381-395.

66. Boulter, D. 1979. Storage and biosynthesis of legume storage proteins. In *Seed protein improvement in cereals and grain legumes;* Int.l Atomic Energy Agency: Vienna; Vol.1 pp. 125-136.

67. Casey, R. and Domoney, C. 1999. Pea Globulins, In *Seed Proteins,* Shewry, P.R. and Casey, R., Eds.; Kluwer Academic Publishers: Dordrecht; pp. 171-208.

68. Gatehouse, J.A.; Lycett, G.W.; Delauney, A.J.; Croy, R.R.; Boulter, D. 1983. Sequence specificity of the post-translational proteolytic cleavage of vicilin, a seed storage protein of pea (*Pisum sativum* L.). *Biochem. J.* **212**: 427-432.

69. Higgins, T.J.V. 1984. Synthesis and regulation of major proteins in seeds. Ann. Rev. Plant Physio. **35**: 191-221.

70. Domoney, C. and Casey, R. 1985. Measurement of gene number for seed storage proteins in Pisum. *Nucleic Acids Res.* **13**: 687-699.

71. Domoney, C. and Casey, R. 1990. Another class of vicilin gene in *Pisum. Planta.* **183**: 39-42.

72. Gatehouse, J.A.; Lycett, G.W.; Croy, R.R.; Boulter, D. 1982. The post-translational proteolysis of the subunits of vicilin from pea (*Pisum sativum* L.). *Biochem. J.* **207**: 629-632.

73. Spencer, D.; Chandler, P.M.; Higgins, T.J.V.; Inglis, A.S.; Rubira, M. 1983. Sequence interrelationships of the subunits of vicilin from pea seeds. *Plant Mol. Biol.* **2**: 259-267.

74. Lycett, G.W.; Delauney, A.J.; Gatehouse, J.A.; Gilroy, J.; Croy, R.R.D.; Boulter, D. 1983. The vicilin gene family of pea (Pisum sativum L.): a complete cDNA coding sequence for preprovicilin. *Nucleic Acids Res.* **11**: 2367-2380.

75. Beachy, R.N.; Jarvis, N.P.; Barton, K.A. 1981. Biosynthesis of subunits of the soybean 7S storage protein. *J. Mol. Appl. Genet.* **1**: 19-27.

76. Coates, J.B.; Medeiros, J.S.; Thanh, V.H.; Nielsen, N.C. 1985. Characterization of the subunits of beta-conglycinin. *Arch. Biochem. Biophys.* **243**: 184-194.

77. Ladin, B.F.; Tierney, M.L.; Meinke, D.W.; Hosangadi, P.; Veith, M.; Beachy, R.N. 1987. Developmental regulation of β-conglycinin in soybean axes and cotyledons. *Plant Physiol.* **84**: 35-41.

78. Thanh, V.H. and Shibasaki, K. 1977. Beta-conglycinin from soybean proteins. Isolation and immunological and physicochemical properties of the monomeric forms. *Biochim. Biophys. Acta.* **490**: 370-384.

79. Yamauchi, F. and Yamagishi, T. 1979. Carbohydrate sequence of a soybean 7S protein. *Agr. Biol. Chem..Tokyo.* **43**: 505-510.

80. Harada, J.J.; Barker, S.J.; Goldberg, R.B. 1989. Soybean β-conglycinin genes are clustered in several DNA regions and are regulated by transcriptional and posttranscriptional processes. *Plant Cell.* **1**: 415-425.

81. Paaren, H.M.; Slightom, J.L.; Hall, T.C.; Inglis, A.S.; Blagrove, R.J. 1987. Purification of a seed glycoprotein: N-terminal and deglycosylation analysis of phaseolin. *Phytochemistry.* **26**: 335-343.

82. Talbot, D.R.; Adang, M.J.; Slightom, J.L.; Hall, T. 1984. Size and organization of a multigene family encoding phaseolin, the major seed storage protein of *Phaseolus vulgaris* L. *Mol. Gen. Genet.* **198**: 42-49.

83. Kriz, A.L. 1989. Characterization of embryo globulins encoded by the maize *Glb* genes. *Biochem Genet.* **27**: 239-251.

84. Schwartz, D. 1979. Analysis of the size alleles of the *Pro* gene in maize - evidence for a mutant protein processor. *Molec. Gen. Genet.* **174**: 233-240.

85. Wallace, N.H. and Kriz, A.L. 1991. Nucleotide sequence of a cDNA clone corresponding to the maize *Globulin-2* gene. *Plant Physiol.* **95**: 973-975.

86. Yupsanis, T.; Burgess, S.R.; Jackson, P.J.; Shewry, P.R. 1990. Characterization of the major protein component from aleurone cells of barley (*Hordeum vulgare L.*). *J. Exp. Bot.* **41**: 385-392.

87. Sun, J.L.; Nakagawa, H.; Karita, S.; Ohmiya, K.; Hattori, T. 1996. Rice embryo globulins: amino-terminal amino acid sequences, cDNA cloning and expression. *Plant Cell Physiol.* **37**: 612-620.

88. Millan, T.; Devos, K.M.; Chinoy, C.N.; Litts, J.L.; Quatrano, R.S.; Gale, M.D. 1992. Chromosomal location and RFLP utility in wheat and barley of a wheat gene with homology to a 7S storage-globulin sequence. *Theor. Appl. Genet.* **85**: 387-388.

89. Slightom, J.L.; Sun, S.M.; Hall, T.C. 1983. Complete nucleotide sequence of a French bean storage protein gene: Phaseolin. *PNAS.* **82**: 3320-3324.

90. Kriz, A.L. 1999. 7S globulins of cereal, In *Seed Proteins,* Shewry, P.R. and Casey, R., Eds.; Kluwer Academic Publishers: Dordrecht; pp. 477-498.

91. Thijssen, M.H.; Spoelstra, P.; Emons, A.M.C. 1996. Immunodetection and immunolocalization of globulin storage proteins during zygotic and somaic embryo development in *Zea mays. Physiol. Plantarum.* **98**: 539-549.

92. Sykes, G.E. and Gaylor, K.R. 1981. Detection and characterization of a new β-conglycinin from soybean seeds. *Arch. Biochem. Biophys.* **210**: 525-530.

93. Yamauchi, F.; Sato, M.; Sato, W.; Kamata, Y.; Shibasaki, K. 1981. Isolation and identification of a new type of β-conglycinin in soybean globulins. *Agr. Biol. Chem. Tokyo.* **45**: 2863-2868.

94. Thanh, V.H. and Shibasaki, K. 1978. Major proteins of soybean seeds. Subunit structure of β-conglycinin. *J. Agr. Food Chem.* **26**: 692-695.

95. Higgins, T.J.V.; Newbigin, E.J.; Spencer, D.; Llewellyn, D.J.; Craig, S. 1988. The sequence of pea vicilin gene and its expression in transgenic tobaccco plants. *Plant Mol. Biol.* **11**: 683-695.

96. Larkins, B.A. 1981. Seed storage proteins: Characterization and Biosynthesis. In *The Biochemistry of Plants.* Marcus, A., Ed.; Academic Press Inc.: New York; vol.6 pp. 449-489.

97. Peterson, D.M. and Smith, D. 1976. Changes in nitrogen and carbohydrate fractions in developing oat groats. *Crop Sci.* **16**: 67-71.

98. Peterson, D.M. 1976. Protein concentration, concentration of protein fractions, and amino acid balance in oats. *Crop Sci.* **16**: 663-666.

99. Peterson, D.M. 1978. Subunit structure and composition of oat seed globulin. *Plant Physiol.* **62**: 506-509.

100. Matlashewski, G.J.; Adeli, K.; Altosaar, I.; Shewry, P.R.; Miflin, B.J. 1982. *In vitro* synthesis of oat globulins. *FEBS Lett.* **145**: 208-212.

101. Burgess, S.R.; Shewry, P.R.; Matlashewski, G.J.; Altosaar, I.; Miflin, B.J. 1983. Characteristics of oat (*Avena sativa L.*) seed globulins. *J. Exp. Bot.* **34**: 1320-1332.

102. Robert, L.S.; Cudjoe, A.; Nozzolillo, C.; Altosaar, I. 1983. Total solubilization of groat proteins in high protein oat (avena sativa L. cv. Hinoat): Evidence that glutelins are a minor component. *Can. Inst. Food Sci. Technol. J.* **16**: 196-200.

103. Burgess, S.R. and Miflin, B.J. 1985. The localization of oat (*Avena sativa* L.) seed globulins in protein bodies. *J. Exp. Bot.* **36**: 945-954.

104. Julianno, B.O. and Boulter, D. 1976. Extraction and composition of rice endosperm glutelin. *Phytochemistry.* **15**: 1601-1606.

105. Robert, L.S.; Adeli, K.; Altosaar, I. 1985. Homology among 3S and 7S globulins from cereals and pea. *Plant Physiol.* **78**: 812-816.

106. Khavkin, E.E.; Misharin, S.I.; Markov, Y.Y.; Peshkova, A.A. 1978. Identification of embryonal antigens of maize: globulins as primary reserve proteins of the embryo. *Planta.* **143**: 11-20.

107. Cross, J.W. and Adams, W.R.J. 1983. Embryo-specific globulins from *Zea mays* L. and their subunit composition. *J. Agric. Food Chem.* **31**: 534-538.

108. Burgess, S.R. and Shewry, P.R. 1986. Identification of homologous globulins from embryos of wheat, barley, rye and oats. *J. Exp. Bot.* **37**: 1863-1871.

109. Quatrano, R.S.; Litts, J.; Colwell, G.; Chakerian, R.; Hopkins, R. 1986. Regulation of gene expression in wheat embryos by ABA: Characterization of cDNA clones for the E_M and putative globulin proteins and localization of the lectin wheat germ agglutinin, In *Molecular Biology of Seed Storage Proteins and Lectins,* Shannon, L.M. and Chrispeels, M.J., Eds.; American Society of Plant Physiologists: Rockville, MD; pp. 127-136.

110. Rivin, C.J. and Grudt, T. 1991. Abscisic acid and the deveopmental regulation of embryo storage proteins in maize. *Plant Physiol.* **95**: 358-365.

111. Paiva, R. and Kriz, A.L. 1994. Effect of abscisic acid on embryo-specific gene expression during normal and precocious germination in normal and viviparous maize (*Zea mays*) embryos. *Planta.* **173**: 110-116.

112. Belanger, F.C. and Kriz, A.L. 1991. Moleular basis for allelic polymorphism of the maize *Globulin-1* gene. *Genetics.* **129**: 863-872.

113. Osterman, J.C. 1988. An allele of the *Prot* locus in maize is a variant for the site of protein processing. *Biochem. Genet.* **23**: 463-474.

114. Zhang, H.; Sreenivasulu, N.; Weschke, W.; Stein, N.; Rudd, S.; Radchuk, V.; Potokina, E.; Scholz, U.; Schweizer, P.; Zierold, U.; Langridge, P.; Varshney, R.K.; Wobus, U.; Graner, A. 2004. Large-scale analysis of the barley transcriptome based on expressed sequence tags. *Plant J.* **40**: 276-290.

115. Williamson, J.D. and Quatrano, R.S. 1988. ABA-regulation of two classes of embryo-specific sequences in mature wheat embryos. *Plant Physiol.* **86**: 208-215.

116. Skylas, D.J.; Mackintosh, J.A.; Cordwell, S.J.; Basseal, D.J.; Walsh, B.J.; Harry, J.; Blumenthal, C.; Copeland, L.; Wrigley, C.W.; Rathmell, W. 2000. Proteome approach to the characterisation of protein composition in the developing and mature wheat-grain endosperm. *J. Cereal Sci.* **32**: 169-188.

117. Singh, J.; Blundell, M.; Tanner, G.; Skerritt, J.H. 2001. Albumin and globulin proteins of wheat flour: Immunological and N-terminal sequence characterisation. *J. Cereal Sci.* **34**: 85-103.

118. Burghardt, K.M. 2002. Molecular characterization of candidate wheat proteins associated with the developement of type 1 diabetes.

119. Osborne, T.B. 1909. *The Vegetable Proteins.* Longmans, Green & Co.: New York.

120. Marcus, J.P.; Green, J.L.; Goulter, K.C.; Manners, J.M. 1999. A family of antimicrobial peptides is produced by processing of a 7S globulin protein in Macadamia integrifolia kernels. *Plant J.* **19**: 699-710.

121. Chung, R.P.T.; Neumann, G.M.; Polya, G.M. 1997. Purification and characterization of basic proteins with *in vitro* antifungal acitivity from seeds of cotton, *Gossypium hirsutum. Plant Sci.* **127**: 1-16.

122. Gomes, V.M.; Mosqueda, M.; Blanco-Labra, A.; Salas, M.P.; Fernandes, K.V.S.; Cordeiro, R.A.; Xavier-Filho, J. 1997. Vicilin storage proteins from *Vigna unguilata* (legume) seeds inhibit fungal growth. *J. Agric. Food Chem.* **45**: 4110-4115.

123. Yunes, A.N.A.; de Andrade, M.T.; Sales, M.P.; Morais, R.A.; Fernandes, K.V.S.; Gomes, V.M.; Xavier-Filho, J. 1998. Legume seed vicilins (7S storage proteins) interfere with the development of cowpea weevil (*Callosobruchus maculatus* (F)). *J. Sci. Food Agric.* **76**: 111-116.

124. Sales, M.P.; Gerhardt, I.R.; Grossi-De-Sa, M.F.; Xavier-Filho, J. 2000. Do legume storage proteins play a role in defending seeds against bruchids? *Plant Physiol.* **124**: 515-522.

125. Chen, Z.; Brown, R.L.; Cleveland, T.E.; Damann, K.E.; Russin, J.S. 2001. Comparison of constitutive and inducible maize kernel proteins of genetypes resistant or suscetible to aflatoxin production. *J. Food Prot.* **64**: 1785-1792.

126. Kamata, Y.; Otsuka, S.; Sato, M.; Shubaki, K. 1982. Limited proteolysis of soybean beta-conglycinin. *Agric. Biol. Chem.* **46**: 2829-2834.

127. Santoro, L.G.; Grant, G.; Pusztai, A. 1988. Differences in the degradation *in vivo* and *in vitro* of phaseolin, the major storage protein of *Phaseolus vulgaris* seeds. *Biochem. Soc. Trans.* **16**: 612-613.

128. Koppelman, S.J.; Bruijnzeel-Koomen, C.A.; Hessing, M.; de Jongh, H.H. 1999. Heat-induced conformational changes of Ara h 1, a major peanut allergen, do not affect its allergenic properties. *J. Biol. Chem.* **274**: 4770-4777.

129. Aalberse, R.C. 2000. Structural biology of allergens. *J. Allergy Clin. Immunol.* **106**: 228-238.

130. Maleki, S.J.; Kopper, R.A.; Shin, D.S.; Park, C.W.; Compadre, C.M.; Sampson, H.; Burks, A.W.; Bannon, G.A. 2000. Structure of the major peanut allergen Ara h 1 may protect IgE-binding epitopes from degradation. *J. Immunol.* **164**: 5844-5849.

131. Mills, E.N.; Huang, L.; Noel, T.R.; Gunning, A.P.; Morris, V.J. 2001. Formation of thermally induced aggregates of the soya globulin beta-conglycinin. *Biochim. Biophys. Acta.* **1547**: 339-350.

132. Teuber, S.S. 2002. Hypothesis: the protein body effect and other aspects of food matrix effects. *Ann. N. Y. Acad. Sci.* **964**: 111-116.

133. Schuppan, D. and Hahn, E.G. Biomedicine. 2002. Gluten and the gut-lessons for immune regulation. *Science.* **297**: 2218-2220.

134. Burks, A.W.Jr.; Brooks, J.R.; Sampson, H.A. 1988. Allergenicity of major component proteins of soybean determined by enzyme-linked immunosorbent assay (ELISA) and immunoblotting in children with atopic dermatitis and positive soy challenges. *J. Allergy Clin. Immunol.* **81**: 1135-1142.

135. Tsuji, H.; Hiemori, M.; Kimoto, M.; Yamashita, H.; Kobatake, R.; Adachi, M.; Fukuda, T.; Bando, N.; Okita, M.; Utsumi, S. 2001. Cloning of cDNA encoding a soybean allergen, Gly m Bd 28K. *Biochim. Biophys. Acta.* **1518**: 178-182.

136. Burks, A.W.; Williams, L.W.; Helm, R.M.; Connaughton, C.; Cockrell, G.; O'Brien, T. 1991. Identification of a major peanut allergen, Ara h I, in patients with atopic dermatitis and positive peanut challenges. *J. Allergy Clin. Immunol.* **88**: 172-179.

137. Burks, A.W.; Cockrell, G.; Stanley, J.S.; Helm, R.M.; Bannon, G.A. 1995. Recombinant peanut allergen Ara h I expression and IgE binding in patients with peanut hypersensitivity. *J. Clin. Invest.* **96**: 1715-1721.

138. Teuber, S.S.; Jarvis, K.C.; Dandekar, A.M.; Peterson, W.R.; Ansari, A.A. 1999. Identification and cloning of a complementary DNA encoding a vicilin-like proprotein, jug r 2, from english walnut kernel (*Juglans regia*), a major food allergen. *J. Allergy Clin. Immunol.* **104**: 1311-1320.

139. Sanchez-Monge, R.; Pascual, C.Y.; Diaz-Perales, A.; Fernandez-Crespo, J.; Martin-Esteban, M.; Salcedo, G. 2000. Isolation and characterization of relevant allergens from boiled lentils. *J. Allergy Clin. Immunol.* **106**: 955-961.

140. Wang, F.; Robotham, J.M.; Teuber, S.S.; Tawde, P.; Sathe, S.K.; Roux, K.H. 2002. Ana o 1, a cashew (*Anacardium occidental*) allergen of the vicilin seed storage protein family. *J. Allergy Clin. Immunol.* **110**: 160-166.

141. Beyer, K.; Bardina, L.; Grishina, G.; Sampson, H.A. 2002. Identification of sesame seed allergens by 2-dimensional proteomics and Edman sequencing: seed storage proteins as common food allergens. *J. Allergy Clin. Immunol.* **110**: 154-159.

142. Wensing, M.; Knulst, A.C.; Piersma, S.; O'Kane, F.; Knol, E.F.; Koppelman, S.J. 2003. Patients with anaphylaxis to pea can have peanut allergy caused by cross-reactive IgE to vicilin (Ara h 1). *J. Allergy Clin. Immunol.* **111**: 420-424.

143. Feldman, M.; Lupton, F.G.H.; Miller, T.E. 1995. Wheats, In *Evolution of crop plants,* Smartt, J. and Simmonds, N.W., Eds.; Longman Scientific: London; pp. 184-192.

144. MacFarlane, A.J.; Burghardt, K.M.; Kelly, J.; Simell, T.; Simell, O.; Altosaar, I.; Scott, F.W. 2003. A type 1 diabetes-related protein from wheat (*Triticum aestivum*) - cDNA clone of a wheat storage globulin, Glb1, linked to islet damage. *J. Biol. Chem.* **278**: 54-63.

145. MacFarlane, A. 2004. Identification and characterization of the Type 1 Diabetes-related wheat stoarge globulin WP5212.

146. Smith, P.K.; Krohn, R.I.; Hermanson, G.T.; Mallia, A.K.; Gartner, F.H.; Provenzano, M.D.; Fujimoto, E.K.; Goeke, N.M.; Olson, B.J.; Klenk, D.C. 1985. Measurement of protein using bicinchoninic acid. *Anal. Biochem.* **150**: 76-85.

147. Bradford, M.M. 1976. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* **72**: 248-254.

148. Shaw, M.M. and Riederer, B.M. 2003. Sample preparation for two-dimensional gel electrophoresis. *Proteomics.* **3**: 1408-1417.

149. Laemmli, U.K. 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature.* **227**: 680-685.

150. Wilm, M.; Shevchenko, A.; Houthaeve, T.; Breit, S.; Schweigerer, L.; Fotsis, T.; Mann, M. 1996. Femtomole sequencing of proteins from polyacrylamide gels by nano-electrospray mass spectrometry. *Nature.* **379**: 466-469.

151. Vasilescu, J.; Smith, J.C.; Ethier, M.; Figeys, D. 2005. Proteomic analysis of ubiquitinated proteins from human MCF-7 breast cancer cells by immunoaffinity purification and mass spectrometry. *J. Proteome Res.* **4**: 2192-2200.

152. Perkins, D.N.; Pappin, D.J.; Creasy, D.M.; Cottrell, J.S. 1999. Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis.* **20**: 3551-3567.

153. Thompson, J.D.; Gibson, T.J.; Plewniak, F.; Jeanmougin, F.; Higgins, D.G. 1997. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* **25**: 4876-4882.

154. Nicholas, K.B.; Nicholas, H.B.J.; and Deerfield, D.W.I. 1997. GeneDoc: analysis and visualization of genetic variation, *EMBNEW. NEWS*, **4**: 14.

155. International Union of Pure and Applied Chemistry and International Union of Biochemistry and Molecular Biology. 1992. *Biochemical Nomenclature and Related Documents.* Portland Press.

156. Robert, L.S. 1985. The expression of seed storage proteins in oat, other cereals and in legumes.

157. Karty, J.A.; Ireland, M.M.; Brun, Y.V.; Reilly, J.P. 2002. Artifacts and unassigned masses encountered in peptide mass mapping. *J. Chromatogr. B. Analyt Technol. Biomed. Life. Sci.* **782**: 363-383.

158. Saenz de Miera, L.E. and Perez de la Vega, M. 2001. Evidence that the N-terminal extension of the vicieae convicilin genes evolved by intragenic duplications and trinucleotide expansions. *Genome.* **44**: 1022-1030.

159. Shewry, P.R. and Tatham, A.S. 1999. The characteristics, structures and evolutionary relationships of the prolamins. In *Seed Proteins,* Shewry, P.R. and Casey, R., Eds.; Kluwer Academic Publishers: Dordrecht; pp. 11-33.

160. Hilton, H. and Gaut, B.S. 1998. Speciation and domestication in maize and its wild relatives: evidence from the globulin-1 gene. *Genetics.* **150**: 863-872.

161. Gill, B.S.; Appels, R.; Botha-Oberholster, A.M.; Buell, C.R.; Bennetzen, J.L.; Chalhoub, B.; Chumley, F.; Dvorak, J.; Iwanaga, M.; Keller, B.; Li, W.; McCombie, W.R.; Ogihara, Y.; Quetier, F.; Sasaki, T. 2004. A workshop report on wheat genome sequencing: International Genome Research on Wheat Consortium. *Genetics.* **168**: 1087-1096.

162. Emanuelsson, O.; Nielsen, H.; Brunak, S.; von Heijne, G. 2000. Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. *J. Mol. Biol.* **300**: 1005-1016.

163. Davey, R.A. and Dudman, W.F. 1979. The carbohydrate of storage glycoproteins from seeds of *Pisum sativum*: Characterization and distribution of component polypeptides. *Aust. J. Plant Physiol.* **6**: 435-447.

164. Duranti, M.; Restani, P.; Poniatowska, M.; Cerletti, P. 1981. The seed globulins of *Lupinus albus*. *Phytochemistry.* **20**: 2071-2075.

165. DeLano, W.L. 2006. The PyMOL Molecular Graphics System (2002).

166. Sun, S.M.; McLeester, R.C.; Bliss, F.A.; Hall, T.C. 1974. Reversible and irreversible dissociation of globulins from Phaseolus vulgaris seed. *J. Biol. Chem.* **249**: 2118-2121.

167. Li, W.; Zhang, P.; Fellers, J.P.; Friebe, B.; Gill, B.S. 2004. Sequence composition, organization, and evolution of the core Triticeae genome. *Plant J.* **40**: 500-511.

168. Sears, E.R. 1966. Nullisomic-tetrasomic combinations in hexaploid wheat. In *Chromosome Manipulation and Plant Genetics.* Riley, R. and Lewis, K.R., Eds.; Oliver & Boyd: Edinburgh; pp. 29-45.

169. Horn, M.E.; Woodard, S.L.; Howard, J.A. 2004. Plant molecular farming: systems and products. *Plant Cell Rep.* **22**: 711-720.

170. Twyman, R.M.; Stoger, E.; Schillberg, S.; Christou, P.; Fischer, R. 2003. Molecular farming in plants: host systems and expression technology. *Trends Biotechnol.* **21**: 570-578.

171. Staebler, J.M. 2006. Clearing the air: expression of nitrous oxide reductase from *Pseudomonas stutzeri* in transgenic tobacco.

172. Kay, R.; Chan, A.; Daly, M.; McPherson, J. 1987. Duplication of CaMV-35S promoter sequences creates a strong enhancer for plant genes. *Science.* **236**: 1299-1302.

173. Datla, R.S.S.; Hammerlindl, J.K.; Panchuk, B.; Pelcher, L.E.; and Keller, W. 1992. Modified binary plant transformation vectors with the wild-type gene encoding *NptII*. *Gene.* **122**: 383-384.

174. Holsters, M.; de Waele, D.; Depicker, A.; Messens, E.; van Montagu, M.; Schell, J. 1978. Transfection and transformation of *Agrobacterium tumefaciens*. *Mol. Gen. Genet.* **163**: 181-187.

175. Horsch, R.B.; Fry, J.E.; Hoffmann, N.L.; Eichholtz, D.; Rogers, S.G.; and Fraley, R.T. 1985. A simple and general method for transferring genes into plants. *Science.* **227**: 229-1231.

176. Thangavelu, M.; Belostotsky, D.; Bevan, M.W.; Flavell, R.B.; Rogers, H.J.; Lonsdale, D.M. 1993. Partial characterization of the *Nicotiana tabacum* actin gene family: evidence for pollen-specific expression of one of the gene family members. *Mol. Gen. Genet.* **240**: 290-295.

177. Bagga, S.; Sutton, D.; Kemp, J.D.; Sengupta-Gopalan, C. 1992. Constitutive expression of the β-phaseolin gene in different tissues of transgenic alfalfa does not ensure phaseolin accumulation in non-seed tissue. *Plant Mol. Biol.* **19**: 951-958.

178. Christensen, A.B.; Thordal-Christensen, H.; Zimmermann, G.; Gjetting, T.; Lyngkjaer, M.F.; Dudler, R.; Schweizer, P. 2004. The germin-like protein GLP4 exhibits superoxide dismutase activity and is an important component of quantitative resistance in wheat and barley. *Mol. Plant Microbe In.* **17**: 109-117.

Appendix A

Sequences

Figure A.1 Wheat 7S globulin cDNA sequences. Complementary DNA sequences for Glb1- and Glb2-like sequences were obtained from a variety of sources: NCBI (National Center for Biotechnology Information), TIGR (The Institute for Genomic Research), GG2.0 (GrainGenes 2.0) and MacFarlane, 2004 (145), as explained in Chapter 2. When applicable, the accession numbers for these sequences are provided. Only the coding sequence is shown: full-length transcripts and contigs can be obtained at the source. The transcripts obtained from the TIGR and GG2.0 databases were designated according to the protein they most closely resembled: Glb1-like A through I and Glb2-like A through C. Gbl1 and WP5212 were named by their source. The open reading frames (ORF) of the coding sequences of Glb1-like A through E only covered a portion of the WP5212 full-length sequence. Glb1-like D and Glb1-like E are derived from different ORFs from the same expressed sequence tag (EST) contig. Glb1-like I is missing sequence data at its 3' end as it does not code for a stop codon. Glb2-like C is truncated relative to Glb2-like A and B. Some sequences contain N when the nucleotide identity is uncertain, instead of one of the four standard nucleotides (A, T, G, C).

>Gbl1 (Source:NCBI, Genbank acc: M81719, Litts, *et al.* 1991, unpublished)
ATGGCGACCCGAGCCAAAGCAACCATCCCTCTCCTCTTTCTCCTCGGCACCAGCCTTCTCTTCGCCGCGGCTGT
TTCCGCCTCCCATGACGACGAGGATGACAGGCGCGGCGGGCACTCGCTGCAGCAGTGCGTGCAGCGATGCCGGC
AGGAGCGGCCGCGGTACTCGCATGCCCGGTGCGTGCAGGAGTGCAGGGACGACCAGCAGCAGCACGGAAGGCAC
GAGCAGGAGGAGGAGCAGGGCCGTGGCCGTGGCTGGCACGGCGAGGGAGAGCGTGAGGAGGAGCACGGCCGTGG
CCGTGGCCGGCATGGCGAGGGAGAGCGTGAGGAGGAGCATGGCCGTGGCCGTGGCCGGCACGGCGAGGGAGAGC
GTGAAGAGGAGCGTGGCCGTGGCCATGGCCGGCATGGCGAGGGAGAGCGCGAGGAGGAGCGTGGCCGTGGCCGT
GGACGGCACGGCGAGGGAGAGCGTGAGGAGGAAGAGGGCCGTGGCCGAGGCCGGCGCGGCGAGGGAGAGCGTGA
TGAGGAGCAAGGGGACAGCCGCCGGCCGTACGTGTTCGGCCCGCGGAGCTTCCGTCGCATCATCCAGAGCGACC
ACGGGTTCGTGAGGGCCCTTCGCCCGTTCGACCAAGTGTCCAGGCTCCTCCGGGGCATCAGGGACTACCGTGTC
GCCATCATGGAGGTGAACCCGCGCGCGTTCGTCGTGCCGGGATTCACGGACGCGGACGGCGTCGGCTACGTCGC
TCAAGGCGAGGGGGTGCTGACGGTGATCGAGAACGGCGAGAAGCGGTCCTACACCGTCAAGGAAGGCGATGTCA
TCGTGGCGCCGGCGGGGTCGATCATGCACCTGGCTAACACCGACGGCCGGAGGAAGCTGGTCATCGCCAAGATC
CTCCACACCATCTCCGTGCCCGGCAAGTTCCAGTTTTTGTCGGTCAAGCCTCTGCTGGCGAGTTTGAGCAAACG
CGTGCTCCGAGCGGCTTTCAAGACCTCCGATGAGCGGCTTGAGAGGCTGTTCAACCAGCGCCAAGGCCAGGAGA
AGACGAGGTCCGTGTCCATCGTCCGCGCGTCCGAGGAGCAGCTCCGTGAGCTGCGTCGCGAGGCGGCCGAGGGC
GGGCAGGGCCACCGCTGGCCTCTCCCTCCGTTCCGCGGCGACTCCCGCGACACCTTCAACCTTCTGGAGCAGCG
CCCAAAGATCGCCAACCGCCATGGCCGCCTCTACGAGGCCGACGCCCGCAGCTTTCACGCTCTCGCCAACCAGG
ACGTCCGGGTCGCCGTTGCCAACATCACGCCGGGTTCTATGACGGCGCCGTACCTGAACACGCAGTCGTTCAAG
CTCGCCGTCGTGCTGGAAGGCGAGGGCGAGGTGCAGATCGTGTGTCCGCACCTTGGCCGCGAAAGCGAGAGCGA
GCGAGAGCACGGAAAGGGCAGGCGGCGCGAGGAAGAGGAGGACGACCAACGGCAGCAACGTCGTCGCGGGTCCG
AGTCCGAGTCGGAGGAGGAGGAGGAGCAGCAGAGGTACGAGACGGTCCGTGCGCGGGTGTCGCGCGGCTCGGCG
TTCGTGGTGCCGCCCGGCCATCCGGTGGTGGAGATCTCGTCGTCCCAAGGCAGCAGCAACCTCCAGGTGGTGTG
CTTCGAGATCAACGCCGAGAGGAACGAGAGGGTATGGCTCGCCGGGAGGAACAACGTGATCGGCAAGCTCGGCA
GCCCCGCCCAGGAGCTCACGTTCGGCAGGCCGGCGAGGGAGGTGCAAGAGGTGTTCCGCGCACAGGATCAGGAC
GAGGGCTTCGTCGCCGGACCCGAGCAGCAGAGCCGAGAGCAGGAGCAGGAGCAGGAGCGGCACCGCCGCCGTGG
TGACCGCGGGCGCGGCGACGAAGCCGTGGAGACGTTCCTGAGGATGGCAACCGGTGCGATCTGA

>WP5212 (Source:MacFarlane, 2004)
ATGGCGACCAGAGGCAGAGCAACCATCCCTCTCCTCTTCCTCCTGGGCACAAGCCTTCTCTTCGCCGCGGCTGT
TTCGGCCTCCCATGACGAGGAGGAGGACAGGCGCGGTGGGCGCTCGCTTCAGCGGTGCGTGCAGCGGTGCCAGC
AGGACCGGCCGCGGTACTCTCATGCCCGGTGCGTGCAGGAGTGCCGGGACGACCAGCAGCAGCACGGAAGGCAC
GAGCAGGAGGAGCAGGGCCGCGGGCATGGCCGGCACGGCGAGGGGGAGCGTGAGGAGGAGCAGGGCCGTGGCCG
TGGGCGGCGCGGCCAGGGAGAGCGTGAGGAGGAGCAGGGCCGTGGACGTGGGCGGCGCGGCGAGGGAGAGCGTG
ATGAGGAGCACGGGGATGGCCGGCGGCCGTACGTGTTCGGCCCGCGCAGCTTCCGCCGCATCATCCGGAGCGAC
CACGGGTTCGTCAAGGCCCTTCGCCCGTTCGACGAAGTGTCCAGGCTCCTCCGGGGCATCAGGAACTACCGTGT
CGCCATCATGGAGGTGAACCCGCGCGCGTTCGTCGTGCCGGGACTCACGGACGCAGACGGCGTCGGCTACGTCG
CTCAAGGCGAGGGGGTGCTGACGGTGATCGAGAACGGCGAGAAGCGGTCCTACACCGTCAGGCAAGGCGATGTG
ATCGTGGCGCCGGCGGGGTCCATCATGCACCTGGCCAACACCGACGGCCGGAGGAAGCTGGTCATCGCCAAGAT
TCTCCACACCATCTCCGTCCCCGGCAAGTTCCAGTATTTCTCGGCCAAGCCTCTCCTCGCTAGTTTGAGCAAAC
GCGTGCTCACAGCGGCGTTAAAGACCTCGGATGAGCGGCTGGGTAGTCTCTTGGGCAGCCGCCAAGGCAAGGAG
GAGGAGGAGAAGTCCATCTCCATCGTCCGCGCGTCAGAGGAGCAGCTCCGCGAGCTGCGTCGCCAGGCGTCCGA
GGGTGACCAGGGCCACCACTGGCCTCTCCCCCCGTTCCGCGGCGACTCGCGCGACACCTTCAACCTCCTGGAGC
AGCGCCCCAAGATCGCCAACCGCCATGGCCGCCTCTACGAGGCCGACGCCCGTAGCTTCCACGCCCTCGCCCAA
CACGACGTCCGCGTCGCCGTGGCCAACATCACGCCGGGTTCTATGACCGCGCCCTACCTGAACACCCAGTCGTT
CAAGCTCGCCGTCGTGCTGGAAGGCGAGGGCGAGGTGGAGATCGTCTGCCCGCACCTCGGCCGCGACAGCGAGC
GCCGCGAGCAAGAGCACGGCAAGGGCAGGTGGAGGAGCGAGGAAGAGGAGGACGACCGGCGGCAGCAACGCCGA
CGCGGGTCCGGCTCCGAGTCGGAGGAGGAGCAGGACCAGCAGAGGTACGAGACGGTCCGCGCGCGGGTGTCGCG
CGGCTCGGCGTTCGTGGTGCCCCCCGGCCACCCGGTGGTGGAGATCGCCTCGTCCCGCGGCAGCAGCAACCTCC
AGGTGGTGTGCTTCGAGATCAACGCCGAGAGGAACGAGCGGGTGTGGCTCGCCGGGAGGAACAACGTGATCGCC
AAGCTGGACGACCCCGCCCAGGAGCTCGCCTTCGGCAGGCCCGCGAGGGAGGTGCAGGAGGTGTTCCGCGCCAA
GGATCAGCAGGACGAGGGCTTCGTCGCCGGACCCGAGCAGCAGCAGGAGCATGAGCGCGGGGACCGCCGCCGTG
GTGACCGCGGGCGCGGCGACGAAGCCGTGGAGGCGTTCCTGAGGATGGCAACCGCCGCGCTCTGA

>Glb1-like A (Source:TIGR accession:TC234134)
ATGGCGACCAGAGCCAGAGTAACCATCCCTCTCCTCTTCCTCCTGGGCACAAGCCTTCTCTTCGCCGCGGCTGT
TTCGGCCTCCCATGACGAGGAGGAGGACAGGCGCGGTGGGCGCTCGCTTCAGCAGTGCGTGCAGCGGTGCCATC
AGGACCGGCCGCGGTACTCGCATGCCCGGTGCGTGCAGGAGTGCCGGGACGAGCAGCAGCAGCATGGAAGGCAC
GAGCAGGAGGAGCAAGGCCGTGGCCATGGCCGGCACGGCGAAGGGGGGCGTGAGGAGGAGCAGGGCCGTGGCCG
TGGCAGGCACGGCCAGGGGGAGCGTGAGGAGGAGGAGGGCCGTGGCCGTGGGCGGCACGGCGAGGGAGAGCGTG
AGGAGGAGGAGGGCCGTGGCCGTGGGCGGCACGGCGAGGGGGAGCGTGAGGAGCACGGAAAGCACGAGCAGGGC
CGTGGCCGGCGCGGCGAGGGAGAGCGTGATGAGGAGCACGGGGATAGCCGCCGGCCGTACGTGTTCGGCCCGCG
CAACTTTCGTAGCATCATCCGGAGCGACCACGGGTTCGTCAAGGCCCTTCGCCCGTTCGACGAAGTGTCCAGGC
TCCTCCGAGGCATCAGGAACTACCGCGTCGCCATCATGGAGGTGAACCCGCGCTCGTTCGTCGTGCCGGGACTC
ACGGACGCGGACGGCGTCGGCTACGTCGCTCAAGGCGAGGGCGTGCTAACGGTGATCGAGAACGGCGAGAGGCG
GTCCTACACCGTCAGGCAAGGCGATGTCATCGTGGCGCCGGCGGGGTCGATCATGCACCTGGCCAACACCGACG
GCCGGAGGAAGCTGGTCATCGCCAAGATTCTCCACACCATCTCCGTGCCCGGCAAGTTCCAGTATTTCTCGGCC
AAGCCTCTCCTCGCGAGTTTGAGCAAACGCGTGCTGAGAGCGGCTCTCAAGACCTCGGATGAGCAGCTGGACAG
GCTGCTGTTCGGCAGGCGCCAAGGCCAGGAGGAGGAGGTCCATCTCCATCGTCCGCGCGTCCGAGGAGCAGCTC
CGCGAGCTGCGTCGCGAGGCGTCCGAGGGCGGGCAGGGCCACCACTGGCCTCTCCCCCCGTTCCGCGGCGACTC
GCGCGACACCTTTAA

>Glb1-like B (Source:TIGR, accession:TC234046)
ATGACCGCGCCGTACTTGAACACCCAGTCATTCAAGCTCGCCGTCGTGCTGGAAGGCGAGGGAGAGGTGGAGAT
CGTGTGCCCGCACCTCGGCCGCGACAGCGAGCGCCGCGAGCACGGCAAGGGCAGGTGGAGCGAGGAAGAGGAGG
ACGACCGACGGCAGCAACGTCGACACGGGTCCGGCTCCGAGTCCGAGTCGGAGGAGGAGCAGGATCAGCAGAGG
TACGAGACGGTCCGCGCGCGGGTGTCGCGCGGCTCGGCGTTCGTGGTGCCCCCCGGCCACCCGGTCGTGGAGAT
CTCCTCGTCCCAAGGCAGCAGCAACCTCCAGGTGGTGTGCTTCGAGATCAACGCCGAGAGGAACGAGAGGGTGT
GGCTCGCCGGCAGGAACAACGTGATCGCCAAGCTCGACAGCCCCGCCCAGGAGCTGACCTTCGGCAGGCCCGCC
AGGGAGGTGCAGGAGGTGTTCCGCGCCAAGGATCAGCAGGACGAGGGCTTCGTCGCCGGACCCGAGCAGCAGAG
CCACGAGCAGGAGCAGGAGCGCGGGGACCGCCGCCGTGGTGACCGCGGGCGCGGTGACGACGCCGTGGGGGCGT
TCCTGAGGATGGCGACCGGCGCGTTCTGA

>Glb1-like C (Source:TIGR, accession:TC234094)
ATGGAGGTGAACCCGCGCGCGTTCGTCGTGCCAGGACTCACGGACGCGGACGGCGTCGGCTACGTCGCTCAAGG
CGAGGGGGTACTGACGGTGATTGAGAACGGCGAGAAGCGGTCCTACACCGTCAGGGAAGGCGATGTCATCGTGG
CGCCGGCGGGGTCTATCATGCACCTGGCCAACACCGACGGCCGGAGGAAGCTGATCATCGCCAAGATTCTCCAC
ACCATCTCCGTGCCCGGCATGTTCCAGTATTTCTCGGCCAAGCCTCTCTTGGCGAGTTTGAGCAAACGCGTGCT
GAGAGCGGCTCTCAAGACCTCCGATGAGCGCCTGGAGAGGCTGTTAGATCCGCGCCAAGGCCAGGAGAAGACAG
GGGGGTCCATGTCCATCGTCCGCGCGTCCGAGGAGCAGCTCCACGAGCTGAGTCGCCAGGCGTCCGAGGGCAGC
CAGGGCCATCACTGGCCCCTCCCCCCGTTCCGCGGCGACTCGCGCGACACCTACAACCTCCTGGAGCAGCGCCC
CAGGATCGCCAACCGCCATGGCCGCCTCTACGAGGCCGACGCCCGCAGCTTCCACGCCCTCGCCCAGCACGACG
TCCGTGTCGCCGTGGCCAACATCACGCCGGGTTCTATGACCGCGCCGTACTTGAACACCCAGTCGTTCAAGCTC
GCCGTCGTGCTGGAAGGCGAGGGCGAGGTGCAGATCGTGTGCCCGCACCTCGGCCGCGACAGCGAGCGCCGCGA
GCAAGGCAAGGGCAGGTGGAGCGAGGAAGAGGACGACGACCAGCGGCAGCAACGCCGACGCGGGTCCGGCTCCG
AGTCCGAGTCCGAGTCGGAGGAGCAGCAGGATCAGCAGAGATACCAGACCATCCGCGCGAGGGTGTCACGCGGC
TCGGCGTTCGTGGTGCCCCCCGGCCACCCGGTGGTGGAGATCGCCTCGTCCCAAGGCAGCAGCAACCTCCAGGT
GGTGTGCTTCGAGATCAACGCCGAGCGGAACGAGAGGGTGTGGCTCGCCGGGAGGAACAACGTGATCGGCAAGC
TGGACAACCCCGCCCAGGAGCTCACGTTCGGCAGGCCCGCCAGGGAGGTGCAGGAGGTGTTCCGCGCCAAGGAT
CAGCAGGACGAGGGCTTCGTCGCCGGTCCCGAGCAGCAGGAGCAGGAGCGCGGGGACCGCCGCCGTGGTGACCG
CGGGCGCGGCGACGAAGCCGTGGAGGCGTTCCTGAGGATGGCGACCGCCGCGCTCTGA

>Glb1-like D (Source:TIGR, accession:TC234172)
ATGGCGACCAGAGGCAGAGCAACCATCCCTCTCCTCTTCCTCCTGGGCACAAGCCTTCTCTTCGCCGCGGCTGT
TTCGGCCTCCCATGACGAGGAGGAGGACAGGCGCGGTGGGCGCTCGCTTCAGCGGTGCGTGCAGCGGTGCCAGC
AGGACCGGCCGCGGTACTCTCATGCCCGGTGCGTGCAGGAGTGCCGGGACGACCAGCAGCAGCACGGAAGGCAC
GAGCAGGAGGAGCAGGGCCGCGGGCATGGCCGGCACGGCGAGGGGGAGCGTGAGGAGGAGCAGGGCCGTGGCCG
TGGGCGGCGCGGCCAGGGAGAGCGTGAGGAGGAGCAGGGCCGTGGACGTGGGCGGCGCGGCGAGGGAGAGCGTG
ATGAGGAGCACGGGGATGGCCGGCGGCCGTACGTGTTCGGCCCGCGCAGCTTCCGCCGCATCATCCGGAGCGAC
CACGGGTTCGTCAAGGCCCTTCGCCCGTTCGACGAAGTGTCCAGGCTCCTCCGGGGCATCAGGAACTACCGTGT
CGCCATCATGGAGGTGAACCCGCGCGCGTTCGTCGTGCCGGGACTCACGGACGCAGACGGCGTCGGCTACGTCG
CTCAAGGCGAGGGGGTGCTGACGGTGATCGAGAACGGCGAGAAGCGGTCCTACACCGTCAGGCAAGGCGATGTG
ATCGTGGCGCCGGCGGGGTCCATCATGCACCTGGCCAACACCGACGGCCGGAGGAAGCTGGTCATCGCCAAGAT
TCTCCACACCATCTCCGTGCCCGGCAAAGTTTCCAGTATTTTCTCGGCCCAAGCCTCTCCTCGCCGAGTTTGA

>Glb1-like E (Source:TIGR, accession:TC234172)
ATGACCGCGCCGTACCTGAACACGCAGTCGTTCAAGTTCGCCGTCGTGCTGGAAGGCGAGGGCGAGGTGCAGAT
CGTGTGCCCGCACCTAGCCCGCGACAGCGAGCGCCGCGAGCACGAGCACGGCAAGGGCAGGCGGAGCGAGGAAG
AGGACGACGACCAGCGGCAGCAACGCCGACGACGCGGGTCCGGCTCCGAGTCCGAGTCCGAGTCCGAGGAGGAG
CAGGACCAGCAGAGGTACGAGACGGTCCGCGCGCGGGTGTCGCGCGGCTCGGCGTTCGTGGTGCCGCCCGGCCA
CCCGGTGGTGGAGATCGCCTCGTCCCGCGGCAGCAGCAACCTCCAGGTGGTGTGCTTCGAGATCAACGCCGAGA
GGAACGAGCGGGTGTGGCTCGCCGGGAGGAACAACGTGATCGGCAAGCTGGACAGCCCCGCCCAGGAGCTCACC
TTCGGCAGGCCCGCCAGGGAGGTGCAGGAGGTGTTCCGCGCCAAGGACCAGCAGGACGAGGGCTTCGTCGCCGG
GCCCGAGCAGCAGAGCCGCCACGAGCAGGAGCAGGAGCGCCGCCGCGGTGACCGCGGGCGTGGTGACGACGCCG
TGGGGGCGTTCCTGAGGATGGCGACCGGCGCGTTCTGA

>Glb1-like F (Source:TIGR, accession:TC234045)
ATGGCGATCAGAGCAACCATTCCTCTCCTCTTCCTCCTGGGAACCAGCCTTCTCTTTGCCGCGGCTGTTTCGGC
CTCCCGTGACGACGAGGAGGACAGGCGCGGTGGGCACTCGCTGCAGCAGTGCGTGCAGCGGTGCCAGCAGGACC
GGCCACGGTACTCGCATGCCCGGTGCGTGCAGGAGTGCAGGGAGGACCAGCAGCAGCACGGGAGGCACGAGCAG
GAGGAGCAGGGCCACAGCCATGGCCGGCACGGCGAAGGGGGGCGTGAGGAGGAGCAGGGCCGTGGCCGTGGGCG
GCACGGCGAGGGAGAGCGTGAGGAGGAGGAGGGCCGTGGCCGTGGGCGGCACGGCGAGGGGGAGCGTGAGGAGC
ACGGAAGGCACGAGCAGGGCCGTGGCCGGCGCGGCGAGGGAGAGCGTGATGAGGAGCACGGGGATAGCCGCCGG
CCGTACGTGTTCGGCCCGCGCAGCTTCCGTAGCATCATCCGGAGCGACCACGGGTTCGTCAAGGCCCTTCGCCC
GTTCGACGAAGTGTCCAGGCTCCTCCGGGGCATCAGGAACTACCGCGTCGCCATCATGGAGGTGAACCCGCGCG
CGTTCGTCGTGCCGGGACTCACGGACGCGGACGGCGTCGGCTACGTCGCTCAAGGCGAGGGGGTGCTGACGGTG
ATCGAGAACGGCGAGAAGCGGTCCTACACCGTCAGGCAAGGCGATGTGATCGTGGCGCCGGCGGGGTCCATCAT
GCACCTGGCCAACACCGACGGCCGGAGGAAGCTGGTCATCGCCAAGATTCTCCACACCATCTCCGTCCCCGGCA
AGTTCCAGTATTTCTCGGCCAAGCCTCTCCTCGCTAGTTTGAGCAAACGCGTGCTCACAGCGGCGTTAAAGACC
TCGGATGAGCGGCTGGGTAGTCTCTTGGGCAGCCGCCAAGGCAAGGAGGAGGAGGAGAAGTCCATCTCCATCGT
CCGCGCGTCAGAGGAGCAGCTCCGCGAGCTGCGTCGCCAGGCGTCCGAGGGTGACCAGGGCCACCACTGGCCTC
TCCCCCCGTTCCGCGGCGACTCGCGCGACACCTTCAACCTCCTGGAGCAGCGCCCCAAGATCGCCAACCGCCAT
GGCCGCCTCTACGAGGCCGACGCCCGTAGCTTCCACGCCCTCGCCCAACACGACGTCCGCGTCGCCGTGGCCAA
CATCACGCCGGGTTCTATGACCGCGCCCTACCTGAACACCCAGTCGTTCAAGCTCGCCGTCGTGCTGGAAGGCG
AGGGCGAGGTGGAGATCGTCTGCCCGCACCTCGGCCGCGACAGCGAGCGCCGCGAGCAAGAGCACGGCAAGGGC
AGGTGGAGGAGCGAGGAAGAGGAGGACGACCGGCGGCAGCAACGCCGACGCGGGTCCGGCTCCGAGTCGGAGGA
GGAGCAGGACCAGCAGAGGTACGAGACGGTCCGCGCGCGGGTGTCGCGCGGCTCGGCGTTCGTGGTGCCCCCCG
GCCACCCGGTGGTGGAGATCGCCTCGTCCCGCGGCAGCAGCAACCTCCAGGTGGTGTGCTTCGAGATCAACGCC
GAGAGGAACGAGCGGGTGTGGCTCGCCGGGAGGAACAACGTGATCGCCAAGCTGGACGACCCCGCCCAGGAGCT
CGCCTTCGGCAGGCCCGCGAGGGAGGTGCAGGAGGTGTTCCGCGCCAAGGATCAGCAGGACGAGGGCTTCGTCG
CCGGACCCGAGCAGCAGCAGGAGCATGAGCGCGGGGACCGCCGCCGTGGTGACCGCGGGCGCGGCGACGAAGCC
GTGGAGGCGTTCCTGAGGATGGCAACCGCCGCGCTCTGA

>Glb1-like G (Source:GG2.0, accessions:Ta.154.1.S1_a_at/Ta.154.1.S1_x_at)
NNGGCGACNAGAGCCAGAGTAACCATCCCTCTCCTCTTCCTCCTGGGCACAAGCCTTCTCTTCGCCGCGGCTGT
TTCGGCCTCCCATGACGAGGAGGAGGACAGGCGCGGTGGGCGCTCGCTTCAGCAGTGCGTGCAGCGGTGCCATC
AGGACCGGCCGCGGTACTCGCATGCCCGGTGCGTGCAGGAGTGCCGGGACGAGCAGCAGCAGCATGGAAGGCAC
GAGCAGGAGGAGCAAGGCCGTGGCCATGGCCGGCACGGCGAAGGGGGGCGTGAGGAGGAGCAGGGCCGTGGCCG
TGGCAGGCACGGCCAGGGGGAGCGTGAGGAGGAGGAGGGCCGTGGCCGTGGGCGGCACGGCGAGGGAGAGCGTG
AGGAGGAGGAGGGCCGTGGCCGTGGGCGGCACGGCGAGGGGGAGCGTGAGGAGCACGGAAAGCACGAGCAGGGC
CGTGGCCGGCGCGGCGAGGGAGAGCGTGATGAGGAGCACGGGGATAGCCGCCGGCCGTACGTGTTCGGCCCGCG
CAACTTTCGTAGCATCATCCGGAGCGACCACGGGTTCGTCAAGGCCCTTCGCCCGTTCGACGAAGTGTCCAGGC
TCCTCCGAGGCATCAGGAACTACCGCGTCGCCATCATGGAGGTGAACCCGCGCTCGTTCGTCGTGCCGGGACTC
ACGGACGCGGACGGCGTCGGCTACGTCGCTCAAGGCGAGGGGGTGCTGACGGTGATCGAGAACGGCGAGAGGCG
GTCCTACTCCGTCAGGCAAGGCGATGTCATCGTGGCGCCGGCGGGGTCCATCATGCACCTGGCCAACACCGACG
GCCGGAGGAAGCTGGTCATCGCCAAGATCCTCCACACCATCTCCGTGCCCGGCAAGTTCCAGTATTTCTCGGCC
AAGCCTCTTCTCGCGAGTTTGAGCAAACGCGTGCTGAGAGCGGCTCTCAAGACCTCGGATGAGCAGCTGGACAG
GCTGCTGTTCGGCAGGCGCCAAGGCCAGGAGGAGGAGNCNNNCATCTCCATCGTCCGCGCGTCCGAGGAGCAGC
TCCGCGAGCTGCGTCGCGAGGCGTCCGAGGGCGGGCAGGGCCACCACTGGCCTCTCCCCCCGTTCCGCGGCGAC
TCGCGCGACACCTTTAACCTCCTGGAGCAGCGCCCCAAGATCGCCAACCGCCATGGCCGCCTCTTCGAGGCCGA
CGCCCGGAGCTTCCACGCCCTNGCCCAACACGACGTCCGTGTCGCCGTGGCCAACATCACNCCGGGATCTATGA
CCGCGCCTTACTTGAACACCCAGTCGTTCAAGCTCGCCGTCGTGCTGGAAGGCGAGGGCGAGGTGCAGATCGTG
TGCCCGCACCTTGNCCNNGACAGCGAGCGCNNNCACNAGCANGGCAAGNNNGGCAGGNGGAGCGAGGAAGANGA
NGACGNCCAGCGGCAGCAACGCCGACGCGGCTCCGGCTCCGAGTCNNNNNNNNNGGAGGAGGAGCAGGATCAGC
AGAGGTACGAGACGGTCCGCGCGCGGGTGTCGCGCGGCTCGGCGTTCGTGGTGCCCCCCGGCCACCCGGTCGTG
GAGATCGCCTCGTCCCGCGGCAGCAGCAACCTCCAGGTGGTGTGCTTCGAGATCAACGCCGAGAGGAACGAGAG
GGTGTGGCTCGCCGGGAGGAACAACGTGATCGGCAAGCTGGACAACCCCGCGCAGGAGCTCACGTTCGGCAGGC
CCGCCAGGGAGGTGCAGGAGGTGTTCCGCGCCAAGGATCAGCAGGACGAGGGCTTCGTCGCCGGACCCGAGCAG
CAGGAGCAGGAGCGCGGGGACCGCCGCCGTGGTGACCGCGGCCGCGGCGACGAAGCCGTGGAGGCCTTCCTGAG
AATGGCAACCGCCGCGCTCTGA

>Glb1-like H (Source:GG2.0, accession:8718.1)
ATGGCGACCAGAGGCAGAGCAACCATCCCTCTCCTCTTCCTCCTGGGCACAAGCCTTCTCTTCGCCGCGGCTGT
TTCGGCCTCCCATGACGAGGAGGAGGACAGGCGCGGTGGGCGCTCGCTTCAGCGGTGCGTGCAGCGGTGCCAGC
AGGACCGGCCGCGGTACTCTCATGCCCGGTGCGTGCAGGAGTGCCGGGACGACCAGCAGCAGCACGGAAGGCAC
GAGCAGGAGGAGCAGGGCCGCGGGCATGGCCGGCACGGCGAGGGGGAGCGTGAGGAGGAGCAGGGCCGTGGCCG
TGGGCGGCGCGGCCAGGGAGAGCGTGAGGAGGAGCAGGGCCGTGGACGTGGGCGGCGCGGCGAGGGAGAGCGTG
ATGAGGAGCACGGGGATGGCCGGCGGCCGTACGTGTTCGGCCCGCGCAGCTTCCGCCGCATCATCCGGAGCGAC
CACGGGTTCGTCAAGGCCCTTCGCCCGTTCGACGAAGTGTCCAGGCTCCTCCGGGGCATCAGGAACTACCGTGT
CGCCATCATGGAGGTGAACCCGCGCGCGTTCGTCGTGCCGGGACTCACGGACGCAGACGGCGTCGGCTACGTCG
CTCAAGGCGAGGGGGTGCTGACGGTGATCGAGAACGGCGAGAAGCGGTCCTACACCGTCAGGCAAGGCGATGTG
ATCGTGGCGCCGGCGGGGTCCATCATGCACCTGGCCAACACCGACGGCCGGAGGAAGCTGGTCATCGCCAAGAT
TCTCCACACCATCTCCGTCCCCGGCAAGTTCCAGTATTTCTCGGCCAAGCCTCTCCTCGCTAGTTTGAGCAAAC
GCGTGCTCACAGCGGCGTTAAAGACCTCGGATGAGCGGCTGGGTAGTCTCTTGGGCAGCCGCCAAGGCAAGGAG
GAGGAGGAGAAGTCCATCTCCATCGTCCGCGCGTCAGAGGAGCAGCTCCGCGAGCTGCGTCGCCAGGCGTCCGA
GGGTGACCAGGGCCACCACTGGCCTCTCCCCCCGTTCCGCGGCGACTCGCGCGACACCTTCAACCTCCTGGAGC
AGCGCCCCAAGATCGCCAACCGCCATGGCCGCCTCTACGAGGCCGACGCCCGTAGCTTCCACGCCCTCGCCCAA
CACGACGTCCGCGTCGCCGTGGCCAACATCACGCCGGGTTCTATGACCGCGCCCTACCTGAACACCCAGTCGTT
CAAGCTCGCCGTCGTGCTGGAAGGCGAGGGCGAGGTGGAGATCGTCTGCCCGCACCTCGGCCGCGACAGCGAGC
GCCGCGAGCAAGAGCACGGCAAGGGCAGGTGGAGGAGCGAGGAAGAGGAGGACGACCGGCGGCAGCAACGCCGA
CGCGGGTCCGGCTCCGAGTCGGAGGAGGAGCAGGACCAGCAGAGGTACGAGACGGTCCGCGCGCGGGTGTCGCG
CGGCTCGGCGTTCGTGGTGCCCCCCGGCCACCCGGTGGTGGAGATCGCCTCGTCCCGCGGCAGCAGCAACCTCC
AGGTGGTGTGCTTCGAGATCAACGCCGAGAGGAACGAGCGGGTGTGGCTCGCCGGGAGGAACAACGTGATCGCC
AAGCTGGACGACCCCGCCCAGGAGCTCGCCTTCGGCAGGCCCGCGAGGGAGGTGCAGGAGGTGTTCCGCGCCAA
GGATCAGCAGGACGAGGGCTTCGTCGCCGGACCCGAGCAGCAGCAGGAGCATGAGCGCGGGGACCGCCGCCGTG
GTGACCGCGGGCGCGGCGACGAAGCCGTGGAGGCGTTCCTGAGGATGGCAACCGCCGCGCTCTGA

>Glb1-like I (Source:GG2.0, accession:8849.1)

ATGGCGATCAGAGCAACCATTCCTCTCCTCTTCCTCCTGGGAACCAGCCTTCTCTTTGCCGCGGCTGTTTCGGC
CTCCCGTGACGACGAGGAGGACAGGCGCGGTGGGCACTCGCTGCAGCAGTGCGTGCAGCGGTGCCAGCAGGACC
GGCCACGGTACTCGCATGCCCGGTGCGTGCAGGAGTGCAGGGAGGACCAGCAGCAGCACGGGAGGCACGAGCAG
GAGGAGCAGGGCCACAGCCATGGCCGGCACGGCGAAGGGGGGCGTGAGGAGGAGCAGGGCCGTGGCCGTGGGCG
GCACGGCGAGGGAGAGCGTGAGGAGGAGGAGGGCCGTGGCCGTGGGCGGCACGGCGAGGGGGAGCGTGAGGAGC
ACGGAAAGCACGAGCAGGGCCGTGGCCGGCGCGGCGAGGGAGAGCGTGATGAGGAGCACGGGGATAGCCGCCGG
CCGTACGTGTTCGGCCCGCGCAACTTTCGTAGCATCATCCGGAGCGACCACGGGTTCGTCAAGGCCCTTCGCCC
GTTCGACGAAGTGTCCAGGCTCCTCCGAGGCATCAGGAACTACCGCGTCGCCATCATGGAGGTGAACCCGCGCT
CGTTCGTCGTGCCGGGACTCACGGACGCGGACGGCGTCGGCTACGTCGCTCAAGGCGAGGGCGTGCTAACGGTG
ATCGAGAACGGCGAGAGGCGGTCCTACACCGTCAGGCAAGGCGATGTCATCGTGGCGCCGGCGGGGTCGATCAT
GCACCTGGCCAACACCGACGGCCGGAGGAAGCTGGTCATCGCCAAGATTCTCCACACCATCTCCGTGCCCGGCA
AGTTCCAGTATTTCTCGGCCAAGCCTCTCCTCGCGAGTTTGAGCAAACGCGTGCTCACAGCGGCTTTGAAGACC
TCGGATGAGCAGCTGGGCAGGCTGCTGTTCGGCAGGCGCCAAGGCCAGGAGGAGGAGTCGTCCATCTCCATCGT
CCGCGCGTCCGAGGAGCAGCTCCGCGAGCTGCGTCGGCAGGCGTCCGAGGGCGGGCAGGGCCACCACTGGCCTC
TCCCCCCGTTCCGCGGCGACTCGCGCGACACCTTTAACCTCCTGGAGCAGCGCCCCAAGATCGCCAACCGCCAT
GGCCGCCTCTTCGAGGCCGACGCCCGGAGCTTCCACGCCCTTGCCCAACACGACGTCCGTGTCGCCGTGGCCAA
CATCACACCGGGATCTATGACCGCGCCTTACTTGAACACCCAGTCGTTCAAGCTCGCCGTCGTGCTGGAAGGCG
AGGGCGAGGTGCAGATCGTGTGCCCGCACCTTGGCCAGGACAGCGAGCGCGAGCACGAGCACGGCAAGGGCAGG
CGGAGCGAGGAAGAAGAGGACGACCAGCGGCAGCAACGCCGACGCGGCTCCGGCTCCGAGTCGGAGGAGGAGCA
GGATCAGCAGAGGTACGAGACGGTCCGCGCGCGGGTGTCGCGCGGCTCGGCGTTCGTGGTGCCCCCCGGCCACC
CGGTCGTGGAGATCGCCTCGTCCCGCGGCAGCAGCAACCTCCAGGTGGTGTGCTTCGAGATCAACGCCGAGAGG
AACGAGAGGGTGTGGCTCGCCGGGAGGAACAACGTGATCGGCAAGCTGGACAACCCCGCGCAGGAGCTCACGTT
CGGCAGGCCCGCCAGGGAGGTGCAGGAGGTGTTCCGCGCCAAGGATCAGCAGGACGAGGGCTTCGTCGCCGGAC
CCGAG

>Glb2-like A (Source:TIGR, accession:TC246874)

ATGAAGTCCGCAGTAAGATCACCATGGCTAGTGCTAGCCATCGTCCTCTCCCTCTGCCTCTCCCTCTCGTTCGC
GTCGTGGGATGCCGAGGACGTAGGCAGGGGTAGTAGGAGGTGGCAAGAAGGGGGCGACGAAGGGCGGTCCGGCG
GAAGTGGCCGGCCGTACCACTTCGGCCAGGAGAGCTACCGGGAGTGGGCCAAGTCGCGGCACGGCCACTTTAAG
GTGCTGGAGCGGTTCGACCACGAGCTGCTCCGGGGTTCCATCGGCGACTACCGCGTCGCGTACCTGGACGCGGC
GCCGCGCGCGTTCCTGCAGCCCAGTCACCACGACGCAGACGAGATCGCCTTCGTCAGGGAAGGCGAGGGCGTGC
TCGTGTTGCTGAGGAACGGGAAGCGGGAGTCGTTCTGCATCAGGGAGGGCGACGTCATCGTGATCCCGGCCGGA
TCCATCGTGTACTCCGCCAACACGCACCGGTCCAAATGGCTCCGCGTCGTCATGTTCATCAACCCCGTCTCCAC
GCCGGGCCGCTTCCAGGAGTTCTTCCTTATTGGATCTGGAGACGAGCGCCCGCAGTCCTTCTTGAGCGTCTTCA
GCGACGAGGTTATCCAGGCCGCATTAAACACTCGGCGGGAGGATGTGGACAGAGTGTTTGAGAGCAAAAGCAAG
GGTGAGGGTGAGATATATGAGGCGTCGGAGGAGCAGATACGGGAGCTGAGCAGGTCGTGCTCCAGGGGAGGACG
CGGCGGCGGCGGCTCGGGTTCCGAGAAGGAGGACATCCAGCCGCGCAGCCTCACCGGCGAGAAGCCGCGCTACT
CGAACAAGCACGGCAGGTTCCACCAGATCACCGGAGACCAGTGCCACCACCTCCGCAAGCTCGACATGGATGTG
ACCCTCGTCAACATCACCCGGGGCTCGATGACGGCACTGAAGTACACCACCCGGTCGACCAGGATCTACGTCGT
CGTGGAGGGGCGCGACGGCTACTTCGAGATGGCGTGCCCGCACATCTCCAGCTCCGGCCGTTCTGAACGCCGTG
AGCACGAGCAAGAGCGCGAGCGCGAGCACGGGCAGGGCAGGAGGAGCGAGGAGCGCGAGCGTGAGCAGGGGCGG
GGCAGGAGGAGCGAGGAGCGCGAGCAGGAGCAGGGAAGGCAGGAGGAGGAGCAGGGCCACGGCCGGGAGCAGGA
GAAATCGAGGGGCTACAGGCAGGTGAGGGCCCAGATCAAGGTGGGGTCGGTGATCGTGCTCCCCGCGGGCCACC
CGGCGACGTTCGTGGCCGGGAACGACGGGAACCTCGCCCTGCTGTCCTTCGGCGTGGGCGCCAACAACGACGAG
GAGGTGTTCGTGACCGGCGGGAACAGCGTCCTGAAGCAGCTGGACGAGGCGGCCAAGGCGCTGTCGTTCCCCCA
GCAGGCGAGGGAGCTGGCGGACAGGGTCATCCGCGCGCAGCCGGAGTCCGTGTTCGTCGCCGGCCCGCAGCAGC
AGCGCCGCGTCGCCGACATGTGA

>Glb2-like B (Source:TIGR, accession:TC246703)
ATGAAGTCCACGGTAGTAAGATCGCCATGGCTAGCGCTAGCCCTCGTCCTCTCCCTGTGCCTCTCCCTCTCGTT
CGCGTCGTGGGATGCCGAGGACGTAGGTAGGGGTAGTAGGAGGTGGCAGGAAGGGGGCGACGACGAAGGGCGGT
CCGGCTCCGGCTCCGGCCGGCCGTACCACTTCGGCGAGGAGAGCTTCCGGGAGTGGGCCAAGTCGCGGCACGGC
CACTTCAAGGTGCTGGAGCGTTTCGACCACGAGCTGCTCCGGGGCTCCATCGGCGACTACCGCGTCGCGTGCCT
GGACGCGGCGCCGCGCGCGTTCCTGCACCCCAGCCACTACGACGCCGACGAGATCGCCTTCGTGAGGGAAGGCG
AGGGCGTGCTGGTGCTGCTGAGGAACGGGAAGCGGGAGTCGTTCTGCGTCAGGGAGGGCGACGTGTTCGTGATC
CCGGCTGGGTCCATCGTGTACTCCGCCAACACGCACCGCTCCAAGTGGTTCCGGGTCGTCATGCTCCTCAACCC
CGTCTCCACGCCGGGCAGCTTCCAGGAGTTCTCCCCTATTGGGTTTGGAGGCGAGCAGCCACAGTCCTTCTTCA
GCGTCTTCAGCGACGAGGTTATCCAGGCGGCATTCAACACTCGGCAGCGGGAGGATGTGGACAGAGTGTTCCAG
AGGAAGAGCAGAGGTGAGGGTCCGATATCTGAGGGGTCGGAGGAGCAGATACGGGAGCTGAGCAGGTCGTGCTC
CAGGGGAGGACGCGGCGGCGGCGGCGGGTCGGGTTCCGAGAAGGAGGACATCCAGCCGCGCAGCCTCACCGGCG
AGAAGCCCCGCTACTCGAACAAGCACGGCAGGTTCCACCAGATCACCGGCGACCAGTGCCACCACCTCCGCAAG
CTCGACATGGATGTCACCCTCGTCAACATCACCCGGGGCTCGATGACGGCGCTGAGATACGCCACCCGGTCGAC
CAGGATCTACATCGTCGTGGAGGGGCGCGACGGCTACTTCGAGATGGCGTGCCCGCACGTCTCCAGCTTCGGCC
GTTCTGAGCGCCGGGAGCACGAGCAGGAGCGCGAGCGCGAGCACGGACACGGCAGGAGGAGCGAGGAGCGCGAG
CGTGAGCATGGGCAGGGCAGGAGGAGCGAGGAGCGCAAGGACGAGCAGGGAAGACAGGAGGAGGAGCAGGGCCG
CGGCCAGGAGCAGGAGAAATCGAGGGGCTACAGGCAGGTGAGGGCCCAGATCAAGGTGGGGTCGGTGATCGTGC
TCCCCGCGGGCCACCCGGCGACGTTCGTGGCCGGGAACGAGGGGAACCTCGCCCTGCTGTCCTTCGGCGTGGGC
GCCAACAACGACGAGGAGGTGTTCGTGACCGGCGGGAACAGCGTGCTGAAGCAGCTGGACGACGCGGCCAAGGC
GCTGGCGTTCCCCCAGCAGGCGAGGGAGCTGGCGGACAGGGTCATCCGCGCGCAGCCGGAGTCCGTGTTCGTCG
CCGGCCCGCAGCAGCAGCGCCGCGTCGCCGACATGTGA

>Glb2-like C (Source:TIGR, accession:TC246759)
ATGGCTCCGCGTCGTCATGTTCATCAACCCCGTCTCCACGCCGGGCCGCTTCCAGGAGTTCTTCCTTATTGGAT
CTGGAGACGAGCGCCCGCAGTCCTTCTTGAGCGTCTTCGGCGAAGAAAACTCCGGCGGGAGGATGTGGACAGAG
TGTTTGAGAGCAAAAGCAAGGGTGAGGGTGAGATATATGAGGCGTCGGAGGAGCAGATACGGGAGCTGAGCAGG
TCGTGCTCCAGGGGAGGACGCGGCGGTGGCGGCGGGTCGGGTTCCGAGAAGGAGGACATCCAGCCGCGCAGCCT
CACCGGCGAGAAGCCCCGCTACTCGAACAAGCACGGCAGGTTCCACCAGATCACCGGCGACCAGTGCCACCACC
TCCGCAAGCTCGACATGGATGTCACCCTCGTCAACATTACCCGGGGCTCGATGACGGCGCTGAGGTACACCACC
CGGTCGACCAGGATCTACATCGTCGTGGAGGGGCGCGACGGCTACTTCGAGATGGCGTGCCCGCACGTCTCCAG
CTCCGGCCGTTCTGAACGCCGGGAGCACGAGCAGGAGCGCGAGCGCGAACACGGACACGGCAGGAGAAGCGAGG
AGCGCGGGCAGGAGCACGGCAGGAGGAGCGAGGAGGAGGAGCACGGCCACGGCGGCGAGCAGGAGAAATCGAGG
GGCTACAGGCAGGTGAGGGCCCAGATCAAGGTGGGGTCGGTGATCGTGCTCCCCGCGGGCCACCCGGCGACGTT
CGTGGCCGGGAACGAGGGGAACCTCGCCCTGCTGTCCTTCGGCGTGGGCGCCAACAACGACGAGGAGGTGTTCG
TGACCGGCGGGAACAGCGTGCTGAAGCAGCTGGACGAGGCGGCCAAGGCGCTGGCGTTCCCCCAGCAGGCGAGG
GAGCTGGCGGACAGGGTCATCCGCGCGCAGCCGGAGTCCGTGTTCGTCCCCGGCCCGCAGCAGCAGCGCCGCGT
CGCCGACATGTGA

Figure A.2 Wheat 7S globulin protein sequences. Protein sequences, presented in FASTA format, were obtained from one of four sources: NCBI (National Center for Biotechnology Information); MacFarlane, 2004; TIGR or GG2.0. Corresponding accession numbers are provided when applicable. Coding sequences corresponding to proteins are shown in Figure A.1. The conceptual translation products obtained from TIGR and GG2.0 were designated according to the known protein they most closely resemble. As a result, these proteins were either Glb1-like or Glb2-like. The Glb1-like A through E open reading frames (ORF) only covered a portion of the WP5212 full-length sequence. Glb1-like D and Glb1-like E are derived from different ORFs from the same EST contig. Glb1-like I is missing sequence data C-terminus relative to the other sequences. Glb2-like C is truncated relative to Glb2-like A and B. Some sequences contain X when the residue identity is uncertain. Sequences identified by N-terminal sequencing (Table 3.3 of this study) are bolded in the corresponding proteins; those identified by mass spectrometry (Table 3.2 of this study) are double underlined.

>Gbl1 (Genbank: accession AAA34269)
MATRAKATIPLLFLLGTSLLFAAAVSASHDDEDDRRGGHSLQQCVQRCRQERPRYSHARCVQECRDDQQQHGRH
EQEEEQGRGRGWHGEGEREEEHGRGRGRHGEGEREEEHGRGRGRHGEGEREEERGRGHGRHGEGEREEERGRGR
GRHGEGEREEEEGRGRGRRGEGERDEEQGDSRRPYVFGPRSFRRIIQSDHGFVRALRPFDQVSRLLRGIRDYRV
AIMEVNPRAFVVPGFTDADGVGYVAQGEGVLTVIENGEKRSYTVKEGDVIVAPAGSIMHLANTDGRRKLVIAKI
LHTISVPGKFQFLSVKPLLASLSKRVLRAAFKTSDERLERLFNQRQGQEKTRSVSIVRASEEQLRELRREAAEG
GQGHRWPLPPFRGD**SRDTFNLL**EQRPKIANRHGRLYEADARSFHALANQDVRVAVANITPGSMTAPYLNTQSFK
LAVVLEGEGEVQIVCPHLGRESESEREHGKGRRREEEEDDQRQQRRRGSESESEEEEEQQRYETVRARVSRGSA
FVVPPGHPVVEISSSQGSSNLQVVCFEINAERNERVWLAGRNNVIGKLGSPAQELTFGRPAREVQEVFRAQDQD
EGFVAGPEQQSREQEQEQERHRRRGDRGRGDEAVETFLRMATGAI

>WP5212 (MacFarlane, 2004)
MATRGRATIPLLFLLGTSLLFAAAVSASHDEEEDRRGGRSLQRCVQRCQQDRPRYSHARCVQECRDDQQQHGRH
EQEEQGRGHGRHGEGEREEEQGRGRGRRGQGEREEEQGRGRGRRGEGERDEEHGDGRRPYVFGPRSFRRIIRSD
HGFVKALRPFDEVSRLLRGIRNYRVAIMEVNPRAFVVPGLTDADGVGYVAQGEGVLTVIENGEKRSYTVRQGDV
IVAPAGSIMHLANTDGRRKLVIAKILHTISVPGKFQYFSAKPLLASLSKRVLTAALKTSDERLGSLLGSRQGKE
EEEKSISIVRASEEQLRELRRQASEGDQGHHWPLPPFRGD**SRDTFNLL**EQRPKIANRHGRLYEADARSFHALAQ
HDVRVAVANITPGSMTAPYLNTQSFKLAVVLEGEGEVEIVCPHLGRDSERREQEHGKGRWRSEEEEDDRRQQRR
RGSGSESEEEQDQQRYETVRARVSRGSAFVVPPGHPVVEIASSRGSSNLQVVCFEINAERNERVWLAGRNNVIA
KLDDPAQELAFGRPAREVQEVFRAKDQQDEGFVAGPEQQQEHERGDRRRGDRGRGDEAVEAFLRMATAAL

>Glb1-like A (TIGR: accession TC234134)
MATRARVTIPLLFLLGTSLLFAAAVSASHDEEEDRRGGRSLQQCVQRCHQDRPRYSHARCVQECRDEQQQHGRH
EQEEQGRGHGRHGEGGREEEQGRGRGRHGQGEREEEEGRGRGRHGEGEREEEGRGRGRHGEGEREEHGKHEQG
RGRRGEGERDEE**HGDSRR**PYVFGPRNFRSIIRSDHGFVKALRPFDEVSRLLRGIRNYRVAIMEVNPRSFVVPGL
TDADGVGYVAQGEGVLTVIENGERRSYTVRQGDVIVAPAGSIMHLANTDGRRKLVIAKILHTISVPGKFQYFSA
KPLLASLSKRVLRAALKTSDEQLDRLLFGRRQGQEEEVHLHRPRVRGAAPRAASRGVRGRAGPPLASPPVPRRL
ARHL

>Glb1-like B (TIGR: accession TC234046)
MTAPYLNTQSFKLAVVLEGEGEVEIVCPHLGRDSERREHGKGRWSEEEEDDRRQQRRHGSGSESESEEEQDQQR
YETVRARVSRGSAFVVPPGHPVVEISSSQGSSNLQVVCFEINAERNERVWLAGRNNVIAKLDSPAQELTFGRPA
REVQEVFRAKDQQDEGFVAGPEQQSHEQEQERGDRRRGDRGRGDDAVGAFLRMATGAF

>Glb1-like C (TIGR: accession TC234094)
MEVNPRAFVVPGLTDADGVGYVAQGEGVLTVIENGEKRSYTVREGDVIVAPAGSIMHLANTDGRRKLIIAKILH
TISVPGMFQYFSAKPLLASLSKRVLRAALKTSDERLERLLDPRQGQEKTGGSMSIVRASEEQLHELSRQASEGS
QGHHWPLPPFRGDSRDTYNLLEQRPRIANRHGRLYEADARSFHALAQHDVRVAVANITPGSMTAPYLNTQSFKL
AVVLEGEGEVQIVCPHLGRDSERREQGKGRWSEEEDDDQRQQRRRGSGSESESESEEQQDQQRYQTIRARVSRG
SAFVVPPGHPVVEIASSQGSSNLQVVCFEINAERNERVWLAGRNNVIGKLDNPAQELTFGRPAREVQEVFRAKD
QQDEGFVAGPEQQEQERGDRRRGDRGRGDEAVEAFLRMATAAL

>Glb1-like D (TIGR: accession TC234172)
MATRGRATIPLLFLLGTSLLFAAAVSASHDEEEDRRGGRSLQRCVQRCQQDRPRYSHARCVQECRDDQQQHGRH
EQEEQGRGHGRHGEGEREEEQGRGRGRRGQGEREEEQGRGRGRRGEGERDEEHGDGRRPYVFGPRSFRRIIRSD
HGFVKALRPFDEVSRLLRGIRNYRVAIMEVNPRAFVVPGLTDADGVGYVAQGEGVLTVIENGEKRSYTVRQGDV
IVAPAGSIMHLANTDGRRKLVIAKILHTISVPGKVSSIFSAQASPRRV

>Glb1-like E (TIGR: accession TC234172)
MTAPYLNTQSFKFAVVLEGEGEVQIVCPHLARDSERREHEHGKGRRSEEEDDDQRQQRRRRGSGSESESESEEE
QDQQRYETVRARVSRGSAFVVPPGHPVVEIASSRGSSNLQVVCFEINAERNERVWLAGRNNVIGKLDSPAQELT
FGRPAREVQEVFRAKDQQDEGFVAGPEQQSRHEQEQERRRGDRGRGDDAVGAFLRMATGAF

>Glb1-like F (TIGR: accession TC234045)
MAIRATIPLLFLLGTSLLFAAAVSASRDDEEDRRGGHSLQQCVQRCQQDRPRYSHARCVQECREDQQQHGRHEQ
EEQGHSHGRHGEGGREEEQGRGRGRHGEGEREEEEGRGRGRHGEGEREEHGRHEQGRGRRGEGERDEE**HGDSRR**
PYVFGPRSFRSIIRSDHGFVKALRPFDEVSRLLRGIRNYRVAIMEVNPRAFVVPGLTDADGVGYVAQGEGVLTV
IENGEKRSYTVRQGDVIVAPAGSIMHLANTDGRRKLVIAKILHTISVPGKFQYFSAKPLLASLSKRVLTAALKT
SDERLGSLLGSRQGKEEEEKSISIVRASEEQLRELRRQASEGDQGHHWPLPPFRGD**SRDTFNLL**EQRPKIANRH
GRLYEADARSFHALAQHDVRVAVANITPGSMTAPYLNTQSFKLAVVLEGEGEVEIVCPHLGRDSERREQEHGKG
RWRSEEEEDDRRQQRRRGSGSESEEEQDQQRYETVRARVSRGSAFVVPPGHPVVEIASSRGSSNLQVVCFEINA
ERNERVWLAGRNNVIAKLDDPAQELAFGRPAREVQEVFRAKDQQDEGFVAGPEQQQEHERGDRRRGDRGRGDEA
VEAFLRMATAAL

>Glb1-like G (GG2.0: accessions Ta.154.1.S1_a_at/Ta.154.1.S1_x_at)
XATRARVTIPLLFLLGTSLLFAAAVSASHDEEEDRRGGRSLQQCVQRCHQDRPRYSHARCVQECRDEQQQHGRH
EQEEQGRGHGRHGEGGREEEQGRGRGRHGQGEREEEEGRGRGRHGEGEREEEGRGRGRHGEGEREEHGKHEQG
RGRRGEGERDEE**HGDSRR**PYVFGPRNFRSIIRSDHGFVKALRPFDEVSRLLRGIRNYRVAIMEVNPRSFVVPGL
TDADGVGYVAQGEGVLTVIENGERRSYSVRQGDVIVAPAGSIMHLANTDGRRKLVIAKILHTISVPGKFQYFSA
KPLLASLSKRVLRAALKTSDEQLDRLLFGRRQGQEEEXXISIVRASEEQLRELRREASEGGQGHHWPLPPFRGD
SRDTFNLLEQRPKIANRHGRLFEADARSFHALAQHDVRVAVANITPGSMTAPYLNTQSFKLAVVLEGEGEVQIV
CPHLXXDSERXHXXGKXGRXSEEXXDXQRQQRRRGSGSESXXXEEEQDQQRYETVRARVSRGSAFVVPPGHPVV
EIASSRGSSNLQVVCFEINAERNERVWLAGRNNVIGKLDNPAQELTFGRPAREVQEVFRAKDQQDEGFVAGPEQ
QEQERGDRRRGDRGRGDEAVEAFLRMATAAL

>Glb1-like H (GG2.0: accession 8718.1)
MATRGRATIPLLFLLGTSLLFAAAVSASHDEEEDRRGGRSLQRCVQRCQQDRPRYSHARCVQECRDDQQQHGRH
EQEEQGRGHGRHGEGEREEEQGRGRGRRGQGEREEEQGRGRGRRGEGERDEEHGDGRRPYVFGPRSFRRIIRSD
HGFVKALRPFDEVSRLLRGIRNYRVAIMEVNPRAFVVPGLTDADGVGYVAQGEGVLTVIENGEKRSYTVRQGDV
IVAPAGSIMHLANTDGRRKLVIAKILHTISVPGKFQYFSAKPLLASLSKRVLTAALKTSDERLGSLLGSRQGKE
EEEKSISIVRASEEQLRELRRQASEGDQGHHWPLPPFRGD**SRDTFNLL**EQRPKIANRHGRLYEADARSFHALAQ
HDVRVAVANITPGSMTAPYLNTQSFKLAVVLEGEGEVEIVCPHLGRDSERREQEHGKGRWRSEEEEDDRRQQRR
RGSGSESEEEQDQQRYETVRARVSRGSAFVVPPGHPVVEIASSRGSSNLQVVCFEINAERNERVWLAGRNNVIA
KLDDPAQELAFGRPAREVQEVFRAKDQQDEGFVAGPEQQQEHERGDRRRGDRGRGDEAVEAFLRMATAAL

>Glb1-like I (GG2.0: accession 8849.1)
MAIRATIPLLFLLGTSLLFAAAVSASRDDEEDRRGGHSLQQCVQRCQQDRPRYSHARCVQECREDQQQHGRHEQ
EEQGHSHGRHGEGGREEEQGRGRGRHGEGEREEEEGRGRGRHGEGEREEHGKHEQGRGRRGEGERDEE**HGDSRR**
PYVFGPRNFRSIIRSDHGFVKALRPFDEVSRLLRGIRNYRVAIMEVNPRSFVVPGLTDADGVGYVAQGEGVLTV
IENGERRSYTVRQGDVIVAPAGSIMHLANTDGRRKLVIAKILHTISVPGKFQYFSAKPLLASLSKRVLTAALKT
SDEQLGRLLFGRRQGQEEEESSISIVRASEEQLRELRRQASEGGQGHHWPLPPFRGD**SRDTFNLL**EQRPKIANRH
GRLFEADARSFHALAQHDVRVAVANITPGSMTAPYLNTQSFKLAVVLEGEGEVQIVCPHLGQDSEREHEHGKGR
RSEEEEDDQRQQRRRGSGSESEEEQDQQRYETVRARVSRGSAFVVPPGHPVVEIASSRGSSNLQVVCFEINAER
NERVWLAGRNNVIGKLDNPAQELTFGRPAREVQEVFRAKDQQDEGFVAGPE

>Glb2-like A (TIGR: accession TC246874)
MKSAVRSPWLVLAIVLSLCLSLSFASWDAEDVGRGSRRWQEGGDEGRSGGSGRPYHFGQESYREWAKSRHGHFK
VLERFDHELLRGSIGDYRVAYLDAAPRAFLQPSHHDADEIAFVREGEGVLVLLRNGKRESFCIREGDVIVIPAG
SIVYSANTHRSKWLRVVMFINPVSTPGRFQEFFLIGSGDERPQSFLSVFSDEVIQAALNTRREDVDRVFESKSK
GEGEIYEASEEQIRELSRSCSRGGRGGGGSGSEKEDIQPRSLTGEKPRYSNKHGRFHQITGDQCHHLRKLDMDV
TLVNITRGSMTALKYTTRSTRIYVVVEGRDGYFEMACPHISSSGRSERREHEQERERREHGQGRRSEEREREQGR
GRRSEEREQEQGRQEEEQGHGREQEKSRGYRQVRAQIKVGSVIVLPAGHPATFVAGNDGNLALLSFGVGANNDE
EVFVTGGNSVLKQLDEAAKALSFPQQARELADRVIRAQPESVFVAGPQQQRRVADM

>Glb2-like B (TIGR, accession TC246703)
MKSTVVRSPWLALALVLSLCLSLSFASWDAEDVGRGSRRWQEGGDDEGRSGSGSGRPYHFGEESFREWAKSRHG
HFKVLERFDHELLRGSIGDYRVACLDAAPRAFLHPSHYDADEIAFVREGEGVLVLLRNGKRESFCVREGDVFVI
PAGSIVYSANTHRSKWFRVVMLLNPVSTPGSFQEFSPIGFGGEQPQSFFSVFSDEVIQAAFNTRQREDVDRVFQ
RKSRGEGPISEGSEEQIRELSRSCSRGGRGGGGGSGSEKEDIQPRSLTGEKPRYSNKHGRFHQITGDQCHHLRK
LDMDVTLVNITRGSMTALRYATRSTRIYIVVEGRDGYFEMACPHVSSFGRSERREHEQERERREHGHGRRSEERE
REHGQGRRSEERKDEQGRQEEEQGRGQEQEKSRGYRQVRAQIKVGSVIVLPAGHPATFVAGNEGNLALLSFGVG
ANNDEEVFVTGGNSVLKQLDDAAKALAFPQQARELADRVIRAQPESVFVAGPQQQRRVADM

>Glb2-like C (TIGR, accession TC246759)
MAPRRHVHQPRLHAGPLPGVLPYWIWRRAPAVLLERLRRRKLRREDVDRVFESKSKGEGEIYEASEEQIR
ELSRSCSRGGRGGGGGSGSEKEDIQPRSLTGEKPRYSNKHGRFHQITGDQCHHLRKLDMDVTLVNITRGS
MTALRYTTRSTRIYIVVEGRDGYFEMACPHVSSSGRSERREHEQERERREHGHGRRSEERGQEHGRRSEEE
EHGHGGEQEKSRGYRQVRAQIKVGSVIVLPAGHPATFVAGNEGNLALLSFGVGANNDEEVFVTGGNSVLK
QLDEAAKALAFPQQARELADRVIRAQPESVFVPGPQQQRRVADM

AppendixB

Expression of recombinant WP5212 in transgenic tobacco

Introduction

Wheat proteins are proposed environmental antigens contributing to the onset and development of type 1 diabetes (T1D). A recent study sought to identify specific wheat proteins involved in this process. Screening of a wheat seed cDNA expression library with polyclonal IgG antibodies identified a wheat clone, WP5212, which was not only antigenic in diabetic BB rats, but could also be linked to islet damage (144, 145). In order to assess the effects of WP5212 on the development of T1D, it is desirable to produce significant quantities of purified protein that could be used in feeding trials.

The development of a plant expression system for the production of WP5212 protein was investigated. Plants offer a number of advantages over other heterologous expression systems such as the ability to carry out post-translational modifications in an environment free of mammalian pathogens, lower production costs and scale-up ability (169). Successful expression and purification of WP5212 from tobacco would permit large-scale production of the protein which could be used in feeding trials to assess its potential as a diabetogen. Tobacco was chosen over other plant expression systems because of its well-established techniques for gene transfer, its ease of transformation and regeneration, and its large biomass (170).

Objective

To subclone the His-tagged *WP5212* sequence into a plant expression vector and generate transgenic tobacco lines expressing the recombinant protein.

Materials and Methods

Materials

Primers were synthesized on site by the UOCDSSF (University of Ottawa Core DNA Sequencing and Synthesis Facility, Ottawa, ON) or were purchased from IDT (Integrated DNA Technologies, Coralville, IA). Restriction, modifying and PCR enzymes were obtained from Invitrogen unless otherwise noted. Purified 6H-WP5212, the product of *6XHis-WP5212* expressed in *E. coli*, and WP5212 epitope-specific polyclonal rabbit antiserum were provided by Dr. F. Scott (OHRI, Ottawa). Secondary and tertiary antibodies came from Cell Signalling Technology, Inc. (Danvers, MA), as described in chapter 2.1.

Expression cassette construction and tobacco transformation

The *WP5212* cds in the pET17b expression vector, flanked at its 5' end by a polyhistidine tag (6XHis-WP5212/pET17b) was provided by Dr. F. Scott (OHRI, Ottawa) and is shown in Figure B.1. To construct a plant expression cassette, the *6XHis-WP5212* cds, renamed *ntHWP*, was subcloned into pBR322 (New England Biolabs, Pickering, ON). The subsequent subcloning steps and the restriction enzymes employed are illustrated in Figure B.1. Briefly, the ex-nosZ sequence was excised from pPCRScript-d35S-ex-nosZ-NOSter (171). The *ntHWP* sequence was cloned into the remainder of the plasmid under the control of the cauliflower mosaic virus double 35S promoter (d35S) for high constitutive transgene expression with the nopaline synthase termination of transcription signal, NOSter (172). The entire expression cassette was cloned into the *Agrobacterium* binary expression vector pRD400 (173). The resultant plant expression vector, pRD-ntHWP, was sent to the Ontario Genomics Innovation Centre (OGIC, Ottawa, ON) to confirm the sequence and integrity of the cds using forward primers WPNtS 5' CCAACATCACGCCGGGTTCTATG 3'; seqWPF1 5' ACCACGGGTTCGTCAAGG 3' and reverse primer seqWPR2 5' CTCGATCACCGTCAGCAC 3' (Table B.1). The freeze-thaw method (174) was used to introduce the pRD-ntHWP vector into *Agrobacterium tumefaciens* strain LBA4404 (Gibco/BRL, Mississauga,

Figure B.1 Construction of the ntHWP expression cassette. The *WP5212* cds (purple), including its 5'6XHis-tag (pink), was renamed *ntHWP* and was subcloned into the pBR322 vector using restriction endonucleases, *Pvu* II and *Sph* I. A second subcloning step using endonucleases, *Bam*H I and *Xba* I, was used to replace the ex-*nosZ* cds (blue) in the pPCRScript-d35S-ex-nosZ-NOSter with the *ntHWP* cds. The completed expression cassette includes the *ntHWP* cds flanked at its 5'end by the cauliflower mosaic virus double 35S (d35S) promoter (green) and at its 3'end by the NOSter, nopaline synthase gene polyadenylation signal (orange). The ntHWP expression cassette was excised from pPCR-ntHWP using restriction endonucleases, *Eco*R I and *Kpn* I. Cassettes were subcloned into the multiple cloning site of the transfer (T) region of binary expression vector pRD400 (not shown). The pRD400 vector also carries within its T-region the kanamycin resistance gene *nptII*, neomycin phosphotransferase II driven by its nopaline synthase promoter (nosP), and 3'-bounded by the NOSter. T-region carrying the ntHWP expression cassette is shown, delimited by the T-DNA left and right borders (LB and RB, respectively) shown in dark blue. Expression construct, pRD-ntHWP, carries the d35S promoter (green), the *WP5212* cds (purple) and the NOSter (orange). The pRD-ntHWP cassette bears a 5'6XHis-tag immediately upstream and in frame with the *WP5212* cds.

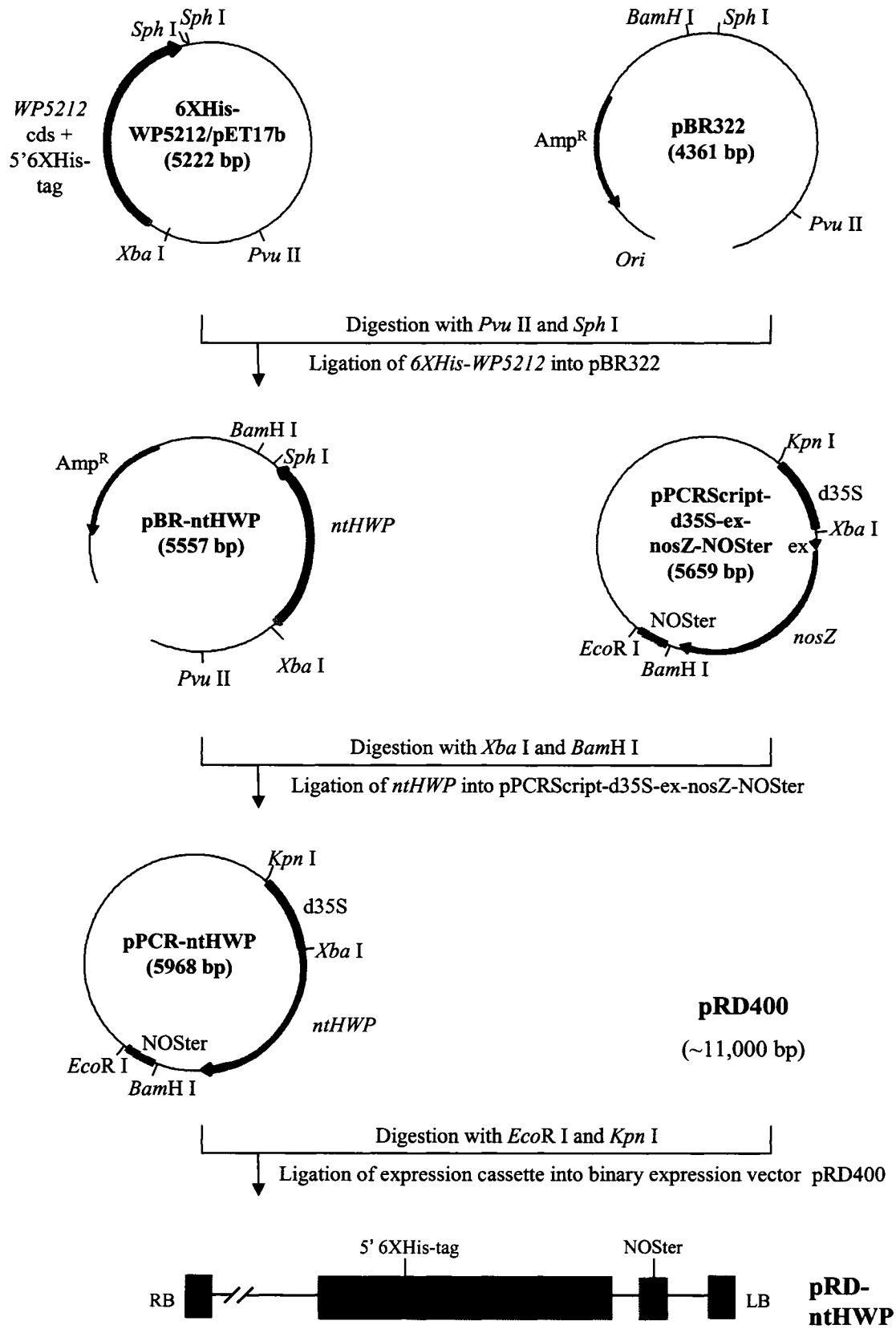

Sph I
Sph I
WP5212 cds + 5'6XHis-tag
6XHis-WP5212/pET17b (5222 bp)
Xba I
Pvu II
BamH I
Sph I
Amp^R
pBR322 (4361 bp)
Pvu II
Ori
Digestion with Pvu II and Sph I
Ligation of 6XHis-WP5212 into pBR322
BamH I
Amp^R
Sph I
pBR-ntHWP (5557 bp)
ntHWP
Pvu II
Xba I
Kpn I
pPCRScript-d35S-ex-nosZ-NOSter (5659 bp)
d35S
Xba I
ex
NOSter
EcoR I
BamH I
nosZ
Digestion with Xba I and BamH I
Ligation of ntHWP into pPCRScript-d35S-ex-nosZ-NOSter
Kpn I
d35S
pPCR-ntHWP (5968 bp)
Xba I
ntHWP
NOSter
EcoR I
BamH I
pRD400
(~11,000 bp)
Digestion with EcoR I and Kpn I
Ligation of expression cassette into binary expression vector pRD400
5' 6XHis-tag
NOSter
RB
LB
pRD-ntHWP

Table B.1 Detailed description of primers used in transgenic study. Specifies primer name, orientation and sequence. Indicates the sequence the primer was designed to bind. Start codon, ATG, is in bolded font.

Primer Name	Primer orientation	Primer sequence	Sequence of primer design	Primer purpose
seqWPF1	Sense	5' ACCACGGGTTCGTCAAGG 3'	WP5212	Sequencing
seqWPR2	Antisense	5' CTCGATCACCGTCAGCAC 3'	WP5212	Sequencing
WPNtS	Sense	5' CCAACATCACGCCGGGTTCTATG 3'	WP5212	PCR amplification and sequencing
WPC	Antisense	5' TGATGGTGGTGGTGGAGCGCGGCGGTTGCCATCC 3'	WP5212	PCR amplification
WP5p1F	Sense	5' **ATG**GCGACCAGAGGCAGAGC 3'	WP5212	PCR amplification
WP3p1R	Antisense	5' CCCGGCGAGCCACAC 3'	WP5212	PCR amplification
TactinS	Sense	5'GGACTGGAATGGTCAAGG 3'	Tobacco actin gene	PCR amplification
TactinA	Antisense	5' AAGTCACGACCAGCTAAATC 3'	Tobacco actin gene	PCR amplification

ON). *Nicotiana tabacum* cv. Xanthi leaf sections were transformed with modifications to the leaf-disc method (175) as previously described (171) and were regenerated on kanamycin (300mg/L) selection media. Transformed tobacco plants (T0 generation) were grown to maturity with weekly fertilizer applications, daily watering and self-fertilization. Desiccated seed pods were collected and seeds were grown on germination medium containing 300mg/L kanamycin, to obtain the first transgenic progeny line, T1.

Extraction of DNA and PCR screening of transformed plants

New tobacco leaves were harvested from transgenic and non-transformed (NT) plants, ground under liquid nitrogen and stored at -80°C until use. DNA for PCR screening was extracted using the DNeasy Plant Mini Kit (Qiagen Inc., Mississauga, ON) following the prescribed instructions. Presence of the transgene was assessed using PCR program MEL2F and primers: WPNtS and WPC. For PCR reaction components and conditions, see Table B.2. NT tobacco DNA served as a negative control. A fragment of the tobacco actin gene, Genbank accession: X63603 (176), was amplified by PCR to show the quality of the NT plant genomic DNA. PCR program used: MELACT7 with primers TactinS and TactinA (Table B.2).

Extraction of RNA and Reverse Transcriptase PCR (RT-PCR)

RNA was extracted from freshly ground tobacco leaf tissue and from milky-endosperm stage wheat seeds using the RNeasy Plant Mini Kit (Qiagen Inc., Mississauga, ON), following the instructions provided. On-column DNA digestion by DNase was performed. The SuperScript First-Strand Synthesis System for RT-PCR (Invitrogen Life Technologies, Burlington, ON) was employed according to the instructions for random hexamers to obtain reverse transcripts. For each RT reaction, a negative control (devoid of SuperScript II reverse transcriptase) was performed to monitor genomic DNA contamination of the RNA extraction. First strand products were quantified in the Beckman DU600 spectrophotometer. Equal amounts of cDNA from transformed and NT plants were employed for RT-PCR. Amplification of the 5' and 3' regions of the *WP5212* cds was conducted

Table B.2 PCR reaction conditions for minicycler used in transgenic study. Specifies program names (stored in cycler's memory), amplification products, primers, PCR reaction components and cycling conditions used in the transgenic study. Dimethylsulfoxide (DMSO) was added, in situations where PCR products were not abundant, to destabilise secondary structure, facilitating denaturation of DNA.

Program name	Product	Primers	Reaction components	Cycle conditions
MELACT1	Tobacco actin cDNA fragment (517 bps)	TactinS and TactinA	1U Taq DNA polymerase; 1μM of TactinA and TactinS; 1xPCR buffer (Invitrogen); 2mM $MgCl_2$; 0.25mM dNTP mix; 5% DMSO; 50ng RT product (cDNA)	**Denaturation**: 3min at 94°C; **32 cycles**: 45s at 94°C, 45s at 48.5°C, 45s at 72°C; **Elongation**: 3min at 72°C
MELACT7	Tobacco actin genomic DNA fragment (1241 bps)	TactinS and TactinA	1U Taq DNA polymerase; 1μM of TactinA and TactinS; 1xPCR buffer (Invitrogen); 2mM $MgCl_2$; 0.25mM dNTP mix; 5% DMSO; 100ng genomic DNA	**Denaturation**: 3min at 94°C; **32 cycles**: 1min at 94°C, 1min at 64°C, 1.5min at 72°C; **Elongation**: 3min at 72°C
MEL2F	Amplification of the 3' terminus of the *WP5212* cds (646bps) from tobacco genomic DNA	WPNtS and WPC	1U Taq DNA polymerase; 1μM WPNtS and WPC; 1X PCR buffer (Invitrogen); 2mM $MgCl_2$; 0.25mM dNTP mix; 5% DMSO; 100ng DNA	**Denaturation**: 5min at 95°C; **35 cycles**: 2min at 94°C, 1min at 65°C, 1min at 72°C; **Elongation**: 5min at 72°C
MELRT8	Amplification of the 5' terminus of the *WP5212* cds (1536bps)	WP5p1F and WP3p1R	1U Taq DNA polymerase; 1μM WP5p1F and WP3p1R; 1X PCR buffer (Invitrogen); 2mM $MgCl_2$; 0.25M dNTP mix; 5% DMSO; 50ng DNA	**Denaturation**: 5min at 95°C; **30 cycles**: 1min at 94°C, 1min at 70.5°C, 2min at 72°C; **Elongation**: 5min at 72°C

using the PCR programs: MELRT8 and MEL2F, respectively (see Table B.2). Plasmid DNA and milky endosperm wheat cDNA were used as positive controls. The tobacco actin gene product was also amplified (program MELACT1, in Table B.2) to ensure quality of the NT plant cDNA and for semi-quantitative purposes.

Extraction of proteins, SDS-PAGE and western blot analysis

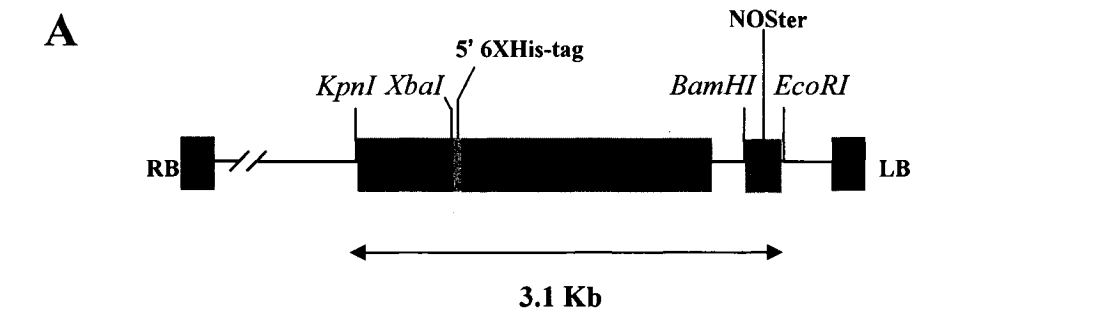
Tobacco leaves from selected transformants were ground with a mortar and pestle under liquid nitrogen and homogenized in extraction buffer (0.1 M Tris pH 6.8, 1 mM ethylenediaminetetraacetic acid (EDTA), 6% SDS, 20% glycerol) in a 1:1 (v/v) ratio. Tubes were vortexed vigorously then heated to 95°C for 10 min with frequent mixing. The suspension was centrifuged at 15,000 x *g* for 20 min. Supernatants were recovered, quantified by Bradford method (147) and used in SDS-PAGE and western blot analyses. *Triticum aestivum* cv. AC Barrie globulins were extracted as in chapter 2.2. Tobacco leaf extracts were supplemented with 5% (v/v) β-mercaptoethanol and 0.01% bromophenol blue. Protein extracts were resolved under SDS-PAGE reducing conditions on discontinuous (5% stacking, 10% resolving) 1.5 mm SDS polyacrylamide gels (149). Approximately 150 μg of tobacco leaf protein were loaded on gels. Electrophoresis was carried out until the dye front reached the bottom of the gel. Proteins from SDS-PAGE were transferred under semi-dry conditions onto nitrocellulose membranes (Bio-Rad, Mississauga, ON). Western blot preparation and analysis was carried out as in chapter 2.

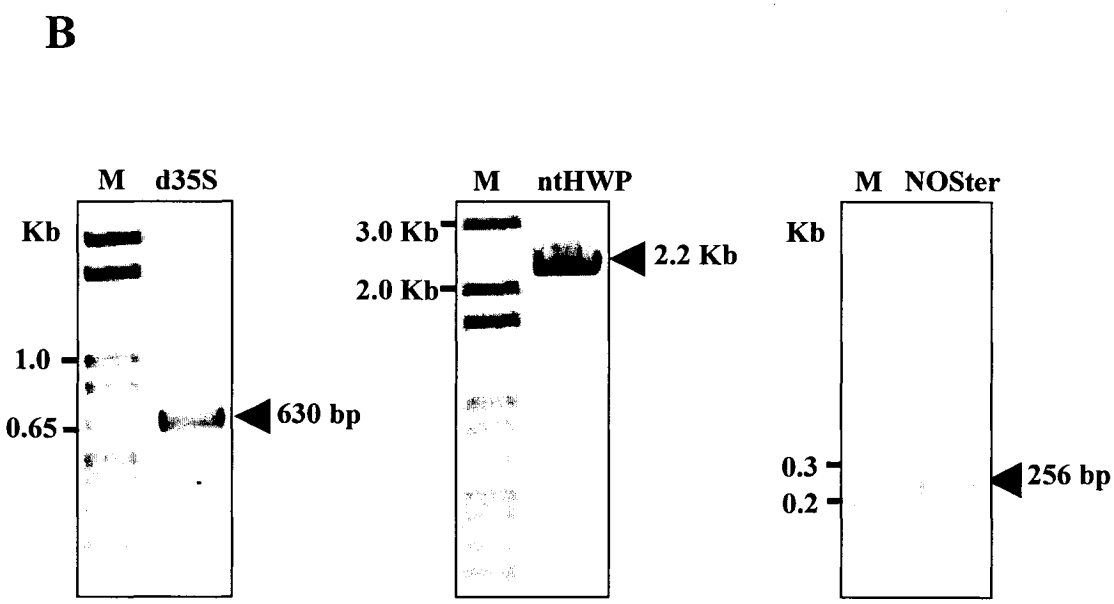
Results & Discussion

Generation and identification of transgenic tobacco lines

Successful construction of the d35S-ntHWP-NOSter expression cassette is shown in Figure B.2. DNA sequencing (data not shown) of the full-length *ntHWP* cds from binary vector pRD-ntHWP indicated that no mutations were present (145). The regenerated tobacco plants (T0) were selected using kanamycin, and confirmed positive by PCR amplification of the 3' region of the *WP5212* cds (646 bp). No amplification was observed in the NT plants. (Figure B.3).

Figure B.2 Restriction endonuclease and PCR analysis of expression construct pRD-ntHWP. **(A)** Depiction of the transfer region (T-DNA) of pRD-ntHWP binary expression vector, delimited by the left and right borders (LB and RB, respectively). Expression cassette (~3.1 kb) includes the d35S promoter, a 5'6XHis-tag, the *WP5212* coding sequence (cds) and the NOSter regulatory sequence. Restriction endonucleases used to excise components of the expression cassette are also identified. **(B)** Restriction digests of the pRD-ntHWP vector, isolated from *E. coli*, show the individual components of the expression cassette. *Kpn* I and *Xba I* were used to excise the d35S promoter (630 bp); *Xba* I and *Bam*H I were used to excise the component containing the nt*HWP* cds (2.2 Kb); *Bam*H I and *Eco*R I were used to excise the NOSter (256 bp). **(C)** Restriction endonucleases, *Kpn* I and *Eco*R I were used to excise the entire ntHWP expression cassette (3.1 Kb) from pRD-ntHWP, isolated from *A. tumefaciens*. **(D)** Primers WPNtS and WPC were used with PCR program MEL2F to show amplification of the 3'region of the *ntHWP* cds (646 bp) from expression vector pRD-ntHWP, isolated from *A. tumefaciens*. The pRD400 vector, devoid of the ntHWP expression cassette, was used as a negative control. M: 1Kb Plus DNA Ladder (Invitrogen).


A
NOSter
5' 6XHis-tag
KpnI XbaI
BamHI EcoRI
RB
LB
3.1 Kb


B
M d35S
Kb
1.0
0.65
630 bp
M ntHWP
3.0 Kb
2.0 Kb
2.2 Kb
M NOSter
Kb
0.3
0.2
256 bp

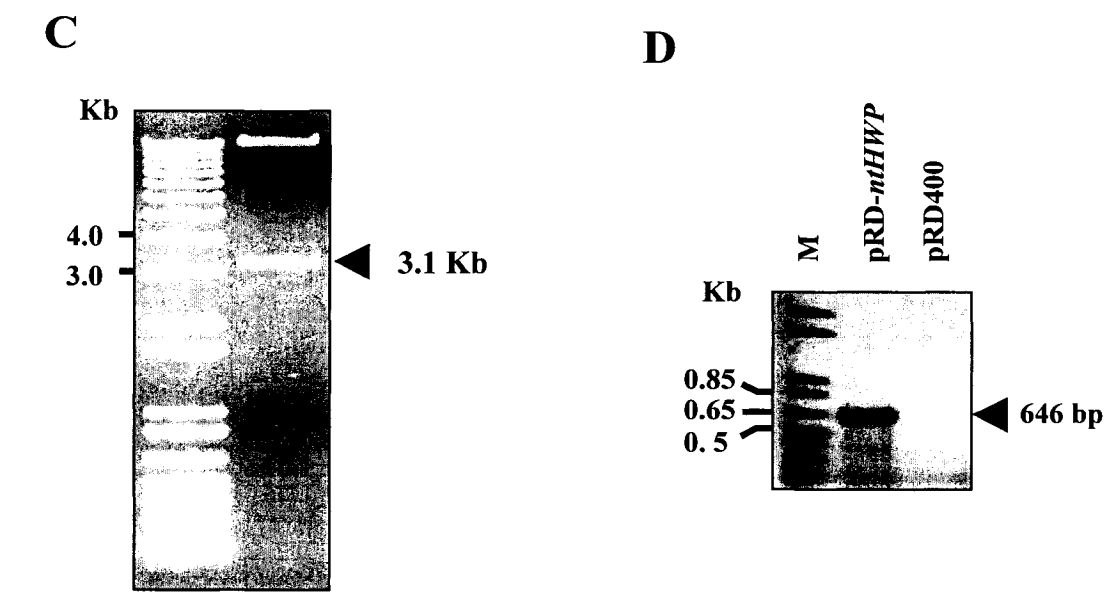
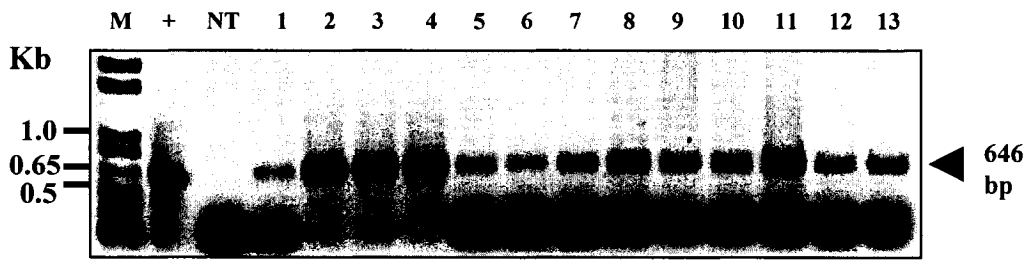
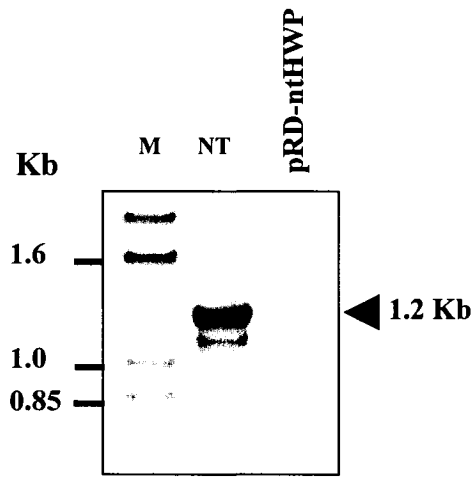

C
Kb
4.0
3.0
3.1 Kb
D
M
pRD-ntHWP
pRD400
Kb
0.85
0.65
0. 5
646 bp

Figure B.3 Identification by PCR screening of ntHWP T0 transgenic tobacco. (A) PCR amplification of the 3' region of the *WP5212* cds (646 bp) from T0 generation transgenic tobacco genomic DNA extracted from leaf tissue. PCR reactions used primers WPNtS (sense) and WPC (antisense) and PCR program MEL2F. Lanes 1 through 13 represent transgenic tobacco lines; +: positive control, pDR-ntHWP; NT: non-transformed tobacco; M: 1Kb Plus DNA Ladder. **(B)** To show that the negative control, NT tobacco DNA was of PCR quality, a fragment of the tobacco actin genomic gene (1.2 Kb) was amplified using primers TactinS and TactinA and PCR program MELACT7. NT: non-transformed tobacco DNA; pRD-ntHWP: negative control; M: 1Kb Plus DNA Ladder (Invitrogen).

A


M + NT 1 2 3 4 5 6 7 8 9 10 11 12 13
Kb
1.0
0.65
0.5
646 bp

B


M NT pRD-ntHWP
Kb
1.6
1.0
0.85
1.2 Kb

Identification of T1 generation plants by PCR analysis confirmed stable integration of the transgene (data not shown).

Expression of ntHWP in transgenic tobacco lines

Full length transcript was identified in all transformants by RT-PCR of the 3' and 5' regions of the transgene (Figure B.4, panels A and B, respectively). Actin cDNA was used as a control for semi-quantitative RT-PCR (Figure B.4, panel B). Plant 7 appeared to have lower *ntHWP* expression than the other plants (Figure B.4, panel B) and was thus omitted from western blots. Immunoblots using optimized dilutions of anti-WP5212 polyclonal antibodies (Figure B.4, panel C) and anti-6XHis-tag antibodies (data not shown) did not detect expression of ntHWP at the protein level. Transformed tobacco extracts failed to show expression of protein corresponding to any of the positive controls: AC Barrie globulin fraction (~64-67kDa, ~50kDa and ~30-35kDa), purified 6H-WP5212 from *E. coli* (~65kDa) and NT protein extract spiked with 6H-WP5212 (Figure B.4, panel C).

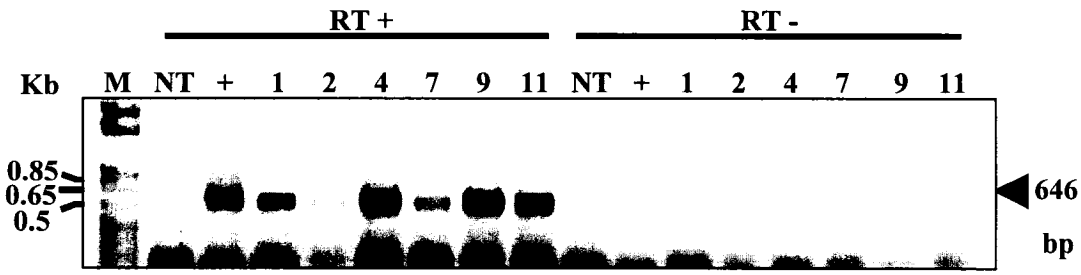
One explanation for the lack of detectable recombinant protein is the presence of a 6XHis-tag N-terminal to the WP5212 hydrophobic signal sequence, which directs the nascent protein into the ER. The hydrophilic hexahistidine tract could prevent entry into the ER resulting in degradation of the ntHWP product. Even if the protein were successfully translocated to the ER, stable accumulation might not occur. A previous study expressing phaseolin, a WP5212 homologue, in alfalfa non-seed storage tissues failed to result in detectable levels of protein (177).

Conclusion

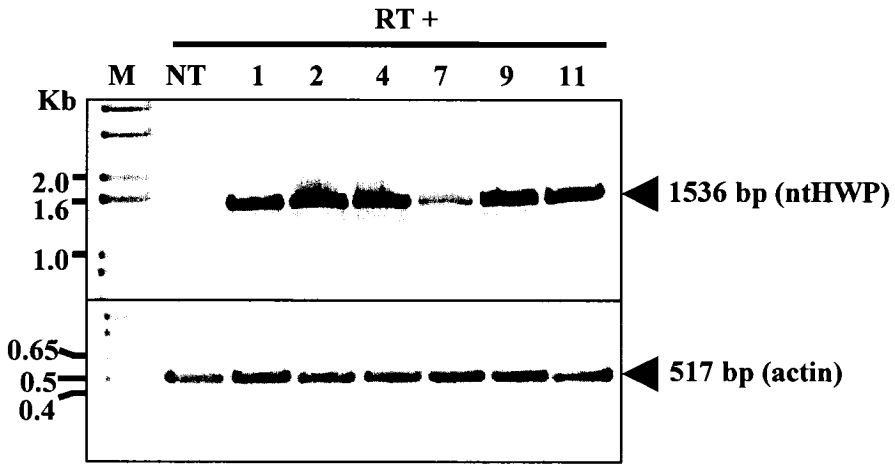
It is clear from the present study that the heterologous expression of wheat seed protein, WP5212 must be more carefully approached. The current N-terminal 6XHis-tag must be removed. However, adding a His-tag immediately C-terminal to the signal sequence may not result in a His-tagged protein. WP5212 is expected to undergo post-translational endoproteolytic cleavage events at its

Figure B.4 Expression of *ntHWP* in T0 generation transgenic tobacco. **(A)** RT-PCR amplification of the 3' region of the *ntHWP* cds (646 bp) from transcripts isolated from T0 generation transgenic tobacco leaf tissue. The reactions were performed with (RT+) or without (RT-) the reverse transcriptase step to confirm amplification of cDNA. **(B)** First-Strand products were quantified prior to PCR and equal amounts were used in each reaction. Upper panel shows semi-quantitative RT-PCR of the 5' region of *ntHWP* (1536 bp). PCR amplification of *ntHWP* was performed using primers WP5p1F and WP3p1R, and cycling program MELRT8. As a control, the bottom panel shows the amplification of a portion of the tobacco actin cDNA (517 bp) using equal amounts of RT product. PCR amplification of actin was performed using primers TactinS and TactinA and PCR program MELACT1. M: 1Kb Plus DNA Ladder (Invitrogen); NT: non-transformed tobacco; lanes 1 through 11: transgenic tobacco plants. **(C)** Immunoblot of total protein extracts from plants expressing *ntHWP*. Following blocking, blots were treated with polyclonal WP5212 antiserum at a dilution of 1:10,000; secondary antibodies (1:200,000) and tertiary antibodies (1:2,000). M: Biotinylated Protein Ladder (Invitrogen); AC Barrie glob: 40 µg salt-soluble (globulin) fraction from wheat; 6H-WP5212: 2 µg *6XHis-WP5212* translated product from *E. coli*; NT+6H-WP5212: 75 µg total protein from NT tobacco spiked with 1.5 µg 6H-WP5212; NT: total protein from non-transformed tobacco; lanes 1 through 11: total protein extracts from transformed plants.

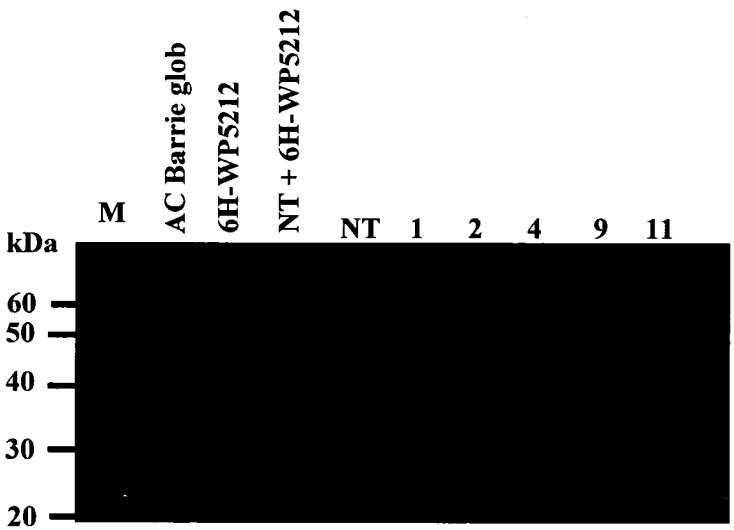
A


RT +
RT -
Kb
M NT + 1 2 4 7 9 11 NT + 1 2 4 7 9 11
0.85
0.65
0.5
646
bp

B


RT +
M NT 1 2 4 7 9 11
Kb
2.0
1.6
1.0
1536 bp (ntHWP)
0.65
0.5
0.4
517 bp (actin)

C


M
AC Barrie glob
6H-WP5212
NT + 6H-WP5212
NT 1 2 4 9 11
kDa
60
50
40
30
20

N-terminus (see chapter 4 of the present study). Given our knowledge of the three-dimensional structure of WP5212 homologues (40, 44-46) and possible WP5212 post-translational events (see chapter 4 of the present study), the use of a polyhistidine tag should be limited to regions of the protein that can tolerate insertions. One such region may be the loop located between the E and F β-barrels in the C-terminal module (23, 25, 46) as shown in Figure 4.2. A C-terminal His-tag is another possibility, though it could result in disruption of the trimeric complex as the terminus is a site of monomer-monomer interaction (44).

While His-tagged proteins allow simplified purification procedures, the presence of a His-tag can disrupt proper folding and accumulation of recombinant proteins (178). If WP5212 is to be used in feeding trials, expression in a plant system producing an edible seed, such as rice, would allow high expression levels to be obtained, without necessitating a purification step. This would eliminate the requirement of a His-tag. Furthermore, since native WP5212 is a seed specific protein, such a production system is more likely to result in a native-like protein conformation and stable accumulation.

Contributors and Collaborators

Agriculture and Agri-Food Canada, Indian Head Research Farm and Seed Increase Unit, Indian Head, SK, provided the wheat, *Triticum aestivum*, cultivar AC Barrie seeds.

Dr. Zemin Yao, Professor and Chair, Department of Biochemistry, Microbiology and Immunology, University of Ottawa, allowed us to borrow his PROTEAN IEF Cell.

Dr. Fraser Scott, Associate Professor, Department of Biochemistry, Microbiology and Immunology, University of Ottawa & OHRI) provided purified globulin WP5212 protein and pre-immune and polyclonal rabbit anti-WP5212 sera.

Julian Vasilescu, Facility Manager, Ottawa Systems Biology Institute, performed the LC-MS/MS and Mascot searches.

Mr. John Steve Smith, Assistant Scientist, Biomolecular Resource Facility Core, Department of Human Biological Chemistry and Genetics, University of Texas Medical Branch, performed the N-terminal sequencing.

Monsanto Company, St. Louis, MO, financially supported this project.

Copyright permission

Request:

Permission to use:
Figure 3 from pg 484
Figure 7 from pg 556
In: Seed Proteins; ISBN 0412815702; Ed: P.R. Shewry and R. Casey; 1999

Reply: 7/6/2006

Dear Ms. McNulty,

With reference to your request (above) to re-use material on which Springer controls the copyright, our permission is granted free of charge, on the following conditions:

* it concerns original material which does not carry references to other sources,
* if material in question appears with credit to another source, authorization from and reference to that source is required as well, and permission is also obtained from the author (address is given on the imprint page or with the article);
* allows you non-exclusive reproduction rights throughout the world,
* permission includes use in an electronic form, on the condition that content is
- password protected,
- at Intranet or
- in CD-ROM/E-book;
* full credit (book/journal title, volume, year of publication, page, chapter/article title, name(s) of author(s), figure number(s), original copyright notice) is given to the publication in which the material was originally published by adding: With kind permission of Springer Science and Business Media.

Permission free of charge does not prejudice any rights we might have to charge for reproduction of our copyrighted material in the future.

With best regards,

Alice Essenpreis
Springer
Rights and Permissions

Tiergartenstrasse 17 | 69121 Heidelberg GERMANY
FAX: +49 6221 487 8223
Alice.Essenpreis@springer.com
WWW.springer.com/rights