THE UNIVERSITY OF CALGARY

Production of the blood anticoagulant, hirudin, in plants

by

Dana L. Parmenter

A DISSERTATION

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE

DEGREE OF DOCTOR OF PHILOSOPHY

DEPARTMENT OF BIOLOGICAL SCIENCES

CALGARY, ALBERTA,

FEBRUARY, 1998

ABSTRACT

This thesis presents a study on the use of oil body proteins, or oleosins, as "carriers" for seed-localized recombinant proteins in plants. Oleosins are unique proteins which decorate the surface of oil bodies in seeds. They are co-translated on the endoplasmic reticulum and target specifically to oil bodies. Starting with a gene encoding an oleosin derived from *Arabidopsis*, a model system for protein production in seeds was created. The model protein used was hirudin, a pharmaceutically valuable blood anticoagulant originally derived from medicinal leeches.

Genetic constructs comprising a synthetic, mature hirudin variant 2 cDNA fused, in-frame, to the 3' end of an *Arabidopsis* oleosin promoter and coding region and terminated by a nopaline synthase terminator were created and introduced into the crop plant, *Brassica napus* (rapeseed, canola). An *Arabidopsis* oleosin promoter was shown to direct seed-specific expression of the transgene. The fusion protein accumulated in tobacco and *B. napus* seed and targeted primarily to the oil body fraction. The hirudin moiety resides on the cytosolic side of the oil body and is released following site-specific proteolysis with Factor Xa. Hirudin variant 2 proteolytically released from its oleosin "carrier" was recovered in a biologically active form. Hirudin was purified to near-homogeneity via floatation centrifugation and two chromatographic steps. The oleosin-hirudin variant 2 fusion protein is stable in seeds for long periods of time (years).

The above experiments clearly demonstrated that oleosins could be used as fusional "carriers" for the production of biologically active hirudin. In an effort to enhance expression and protein accumulation of the oleosin-hirudin fusion gene/protein, the above construct was modified. A longer *Arabidopsis* oleosin promoter was used to drive the expression of a *B. napus* oleosin-hirudin variant 1 fusion gene harbouring, or lacking, an alfalfa mosaic virus RNA4 (AMV) leader sequence. Preliminary results indicate that the fusion protein accumulated to higher levels.

In a third set of experiments, three constructs, having 1, 2, or 4 hirudin-encoding concatameric sequences fused, in-frame, to a *B. napus* cDNA were created. Analysis of seed extracts from transformants harbouring these constructs showed that the fusion protein accumulated and targeted to the oil body.

ACKNOWLEDGMENTS

My journey into the world of science is just beginning, and I have many people to thank for making my voyage so far such an enjoyable one.

First of all, I would like to thank my supervisor, Dr. Maurice M. Moloney for giving me the opportunity to work in his lab, to gain experience through his knowledge, and for allowing me the opportunity to speak at conferences, including an international conference in Amsterdam. Not only did he provide insight which guided me into becoming a better researcher, speaker, and writer, but he was a very good friend. Thank you.

I would also like to thank a number of people in Dr. Moloney's lab and the Botany Division. First I would like to thank Dr. Gijs van Rooijen, who provided me with a solid knowledge base when I first came to the lab as a very "green" undergraduate. He helped take the sting out of first-year inexperience, and I managed to reap a lot of valuable data through his expertise. I would also like to thank Dr. Joe Boothe and Jill Saponja, the FPLC and HPLC wizards of the lab, who helped make the chromatographic purification of recombinant hirudin, such as that shown in Fig. 3.12, possible. Thanks to Dr. Ed Yeung, who guided me through an independent project which resulted in the immunofluorescent localization of recombinant hirudin shown in Fig. 3.10. Thank you to Dr. Thorpe, who not only served on my committee, lo these many years, but provided insightful discussion in the field of embryogenesis, an area of study which I will enter as a post-doctoral fellow. Thank you to my committee members, Drs. Enno Krebbers, Bill Crosby, Henning Muendel, and Larry Kawchuk who agreed to travel and be on my committee, despite their hectic schedules. I would like to thank Ciba Geigy for funding my research project. Of course, I would also like to thank the members of the Moloney lab, for their willingness to share their expertise, and their friendship. Finally, I would like to thank my best friend, Steve Zaplachinski, whose patience, clarity of thought, intelligence and reason have helped me to be a better researcher and person.

Thank you, Steve

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

°C	degrees Celsius
2.4-D	2.4 dichlorophenoxy acetic acid
^{32}P	phosphorous 32
^{35}S	sulfur 35
AMV	alfalfa mosaic virus
ATP	adenosine triphosphate
BA	6-benzylamino purine
bp	base pair
BSA	bovine serum albumin
cDNA	complementary deoxyribonucleic acid
Ci	Curie
cOL	oleosin cDNA
cpm	counts per minute
d	day(s)
dATP	2 - deoxyadenosine-5'-triphosphate
dCTP	2 - deoxycytosine-5'-triphosphate
ddH_2O	distilled, deionized water
DNA	deoxyribonucleic acid
dNTP	2-deoxyribonucleotide-5'-triphosphate
dpa	days post anthesis
dpm	disintegrations per minute
dTTP	2-deoxythymidine-5'-triphosphate
EDTA	ethylenediamine tetraacetic acid
FXa	Factor Xa
g	gram(s)

GUS	ß-glucuronidase
h	hour(s)
H_20	water
HV1	hirudin variant 1
HV2	hirudin variant 2
IAA	indole-3-acetic acid
IBA	indole-3-butyric acid
kb	kilobase
kbp	kilobase pair(s)
kDa	kiloDalton
L	liter
LB	Luria broth
m	meter
M	Molar
mA	milliamperes
min	minute(s)
mg	milligram
ml	milliliter
mm	millimeter
mM	millimolar
mRNA	messenger ribonucleic acid
MS	Murashige-Skoog
NAA	naphthalenacetic acid
ng	nanogram
nm	nanometer
nmol	nanomol

NOS	nopaline synthase
OB	oil body
OBP	oil body protein
OD_x	optical density at wavelength x
OLB	oligolabelling buffer
oleo	oleosin
PBS	phosphate buffered saline
PCR	polymerase chain reaction
pmol	pico mol
PMSF	phenylmethylsulphonylfluoride
PVDF	polyvinylidene difluoride
RNA	ribonucleic acid
rpm	revolutions per minute
rRNA	ribosomal RNA
RT	room temperature
s	second(s)
SDS	sodium dodecyl sulfate
SDS-PAGE	SDS- polyacrylamide gel electrophoresis
sHV1	HV1 having an N-terminal secretion leader peptide
SSC	150 mM NaCl, 15 mM sodium citrate
SSPE	150 mM NaCl, 10 mM sodium phosphate, 1 mM EDTA
TAG	triacylglycerol
TBS	Tris buffered saline
Tris-Cl	Tris(hydoxymethyl)aminoethane chloride
tRNA	transfer ribonucleic acid
U	international unit

μCi	microcurie
μg	microgram
μl	microliter
μM	micromolar
UV	ultraviolet
V	volts
vol	volume

Chapter 1:

Introduction

1.1 General Introduction

The production and marketing of biologically active peptides and proteins having pharmaceutical, industrial or diagnostic applications represents a large and lucrative industry. By mid- 1997, total market capitalization of the top 50 biotechnology and top 10 pharmaceutical companies exceeded $40 billion, and $596 billion, respectively (Anonymous, 1997a). Continued advancements in molecular biology technology have resulted in the economical and large-scale production of a number of biotechnology-derived therapeutic proteins (biopharmaceuticals). Currently, there are at least 29 biopharmaceuticals available on the world market (Thomas, 1995), an additional 280 biotechnology-derived products in clinical trials, and 65 products in the third and final phase of clinical trials (Anonymous, 1997b). These proteins include hormone replacements, monoclonal antibodies, vaccines, blood clotting and anticlotting agents, and cytokines (e.g. interleukins, interferons, tumor necrosis factors). Total sales of these products reached $10 billion in 1996 (Anonymous, 1997b).

Encouraged by the lucrative possibility that molecular biology can create a new generation of superior biopharmaceuticals, approximately 1000 biopharmaceutical companies exist in the U.S., alone (Abelson, 1996). Even countries such as Germany, which had been exceedingly distrustful of this industry, are now eager to join the biotechnology race (Dickman, 1996). The financial strength of the biotechnology industry has been fortified by the partnership of biotechnological and pharmaceutical companies (Persidis and Persidis, 1996). This strength has also resulted in heightened competition in the biopharmaceutical drug market. A recent example of this is the rivalry between drug companies to produce thrombolytic agents (blood thinners) such as tissue plasminogen activator (tPa). Genentech's tPa, Activase, which has monopolized this $315 million market, has been challenged by Boehringer Mannheim with the development of a

potentially superior variant of tPA, a deletion mutant called Retavase (Edgington, *et al.*, 1997). The competition in the biotechnology industry has provided greater incentive for the improvement of bioprocesses, the development of more efficient expression systems, and lower production costs of recombinant proteins.

The characterization of proteins of interest has been facilitated/achieved through advancements in *in vitro* bioassays, PCR analysis, immunological tests, such as ELISAs, epitope mapping, immunoblotting, chromatographic techniques, 3-D structural analysis, and mass spectrometry. Although advancements in molecular biology have significant economic implications, an important result of this technology is its impact on fundamental research. Improving the ease by which proteins can be expressed, isolated and purified requires an understanding of fundamental protein biochemistry.

To be commercially feasible, the large-scale production and isolation of proteins of interest must be inexpensive. This will be dependent on expression levels and the ease of isolation and purification of these proteins from the host organism. The natural source of such proteins can only rarely accommodate such demands, necessitating the production of recombinant proteins. It has been possible to produce economically important proteins by recombinant DNA technology for about 20 years (Itakura *et al.*, 1977: Goeddel *et al.*, 1979, 1980). This technology not only permits high yields of target proteins at economically attractive costs, but also facilitates production of proteins not found in nature or variants superior to naturally occurring product.

Biologically active peptides or proteins may be produced through chemical synthesis, or via biological expression systems such as bacterial or yeast fermentation, baculovirus-infected insect cells, or mammalian cell cultures. Chemical synthesis of peptides is not amenable to large-scale production for a number of reasons. Though relatively pure preparations of peptides can be synthesized chemically, it costs

approximately $50-$75 per amino acid to give 25 mg of peptide and it is extremely difficult to synthesize peptide chains greater than 60 amino acids. In addition, synthetic coupling can be complicated by the secondary structure of the growing peptide chain (McMaster, 1993). As a result, large-scale production of biologically important proteins is typically accomplished by exploiting biological expression systems. A number of factors must be considered when choosing an expression system for the production of heterologous proteins. These include the ease of manipulation of the host organism, the range of posttranslational modifications possible, the site of synthesis and deposition of the heterologous protein, the effect of heterologous gene expression on growth of the host, and the level of protein accumulation required. The quality of the protein produced is of particular importance if it is to be used therapeutically *vs.* diagnostically. In addition, compatibility of the production system and the quantity of the protein required should be considered. The optimal production system for a protein which is required in small quantities (i.e. 1kg/year) will be much different than one required at 10^5kg/year.

As a process is scaled-up, it may result in a disjunction between quantities which can be produced and the quality of the product. Production economics are heavily influenced by the ease of recovery and purification of the desired protein. As technology has advanced, a wide variety of expression systems have become available. Each has its own singular advantages. Most have significant limitations in their scope. To provide a context for critical discussion of plant-based expression, the following is a brief description of the most common expression systems used for the production of heterologous proteins.

1.2 Expression Systems For Recombinant Proteins

1.2.1 *E. coli* as a host for recombinant protein production

E. coli is the most widely used organism for the expression of recombinant proteins. There are a number of reasons why *E. coli* lends itself well to expression technology. Extensive knowledge regarding the genetics and molecular biology of *E. coli* exists, and a wide variety of well-defined *E. coli* mutants are readily available. This prokaryote is easily transformed by plasmid DNA using one of many fast and simple techniques. Since the 1970's, a plethora of plasmid DNA vectors have been developed (rev. by Makrides, 1996). Vectors may vary with regard to their copy number, selectable markers, and promoter/inducer systems. Expression vectors have been developed which have strong, inducible promoters exhibiting minimal basal transcriptional activity (rev. by Makrides, 1996). Transformed cells may be recovered using one of many selection markers. This organism is easy to grow in large quantities at high cell density, lending itself well to large-scale heterologous protein production in fermentation systems. It is not uncommon for heterologous proteins expressed in *E. coli* to represent 10-20% of the total cell protein (King *et al.*, 1989). The advantages of *E. coli* led it to become the "system of choice" in the 1980's, and little attention was focused on other expression systems, such as yeast and *B. subtilis*, at that time. However, when attempts to express larger, more complex proteins, such as tissue plasminogen activator (tPA), failed, the limitations of this system became apparent, stimulating interest in other expression hosts.

Prokaryotes lack the ability to glycosylate proteins. This can be advantageous if the desired protein is to be used for crystallographic analysis, as the carbohydrate moieties of glycoproteins are often heterogeneous. This heterogeneity often results in a structurally disordered protein which is inhibitory to crystallographic investigation (McPherson, 1982). Unfortunately, many eukaryotic proteins require glycosylation for biological activity or stability (King *et al.*, 1989), making *E. coli* an unsuitable host for production of several important eukaryotic proteins. In addition, proteins made in bacteria are often deposited as

highly concentrated, denatured aggregates (inclusion bodies) when the expression level exceeds 10% of cytoplasmic proteins (rev. by Heinrikson and Tomasselli, 1991). Although this may protect the heterologous protein from proteolysis and facilitate its separation from the rest of the bacterial contents by differential centrifugation, solubilization may require strongly denaturing detergents. As a result, it may be necessary to restore (refold) the proteins to their native and biologically active form. This is a time consuming, difficult, and an expensive process which often results in significant losses. In addition to some fundamental problems with protein synthesis and deposition, gene sequences of interest which originate from a eukaryotic organism must be modified before they can be appropriately expressed by prokaryotic systems. This is because bacteria lack the ability to recognize eukaryotic promoters and to splice intervening sequences from an RNA transcript. Production of foreign proteins in prokaryotes may also stimulate synthesis of proteolytic heat shock proteins, resulting in degradation of the heterologous protein (Goff and Goldberg, 1985, 1987). Large-scale protein production using bacterial fermentation systems can also be impractical because of the cost associated with prokaryotic fermentation procedures, protein toxicity towards its bacterial host, and difficulties associated with protein purification and recovery (Sherwood, 1991).

Although there can be a number of difficulties associated with expression of heterologous proteins in large-scale bacterial cultures, a large number of recombinant proteins have been successfully produced in *E. coli* and are commercially viable. These include hormones (most notably, insulin), immunomodulators such as interferon-alpha, interferon-gamma, and interleukins, growth factors, blood proteins and vaccines (Hodgson, 1993).

The production of heterologous proteins in bacteria is not limited to *E. coli*. *B. subtilis* is a non-pathogenic, genetically manipulable prokaryote capable of secreting

proteins into the culture medium (rev. by Wong, 1995). *Bacillus subtilis* is currently used for the production of industrial enzymes, most notably, the alkaline protease, subtilisin (rev. by Ferrari *et al.*, 1993). Subtilisin has a wide variety of industrial applications, including baking, brewing, cheese making, leather processing, and meat tenderizing. By 1992, commercial sales of this enzyme exceeded $300 million. The typical yield of enzymes secreted during *Bacillus* fermentation is 20g/L, however, significantly higher yields have been achieved. For example, Genecor International Inc., one of the major companies producing subtilisin, has reported yields of up to 350g/L (Dr. S-L Wong, personal communication). Other industrial enzymes produced in *Bacillus* include amylase, cellulase and lipase. In addition to industrial enzyme production, this organism has been used for the expression of medically important proteins such as human interleukin and anitidigoxin single-chain antibody fragment (rev. by Wong, 1995).

1.2.2 Yeast as a host for recombinant protein production

Yeast-derived production of heterologous proteins originally employed the budding yeast *Saccharomyces cerevisiae*. This is a nonpathogenic organism which has been used in the baking and brewing industries for centuries. The large body of knowledge concerning the genetics and physiology of this organism has assisted the exploitation of this organism as a protein production system. The use of this organism on a laboratory scale for the production of proteins of interest is enormous. In addition, yeast production systems are readily adapted for use in large-capacity fermenters.

Like many bacteria, yeast is easily manipulated and can be grown in large-scale fermenters. However, unlike prokaryotes, it is capable of post-translational protein modifications, such as disulfide bond formation, proteolysis of signal peptides and N- and O- linked glycosylation (rev. in Buckholz and Gleeson, 1991). In addition, yeast has a

sophisticated endomembrane and secretory system which can be manipulated for the extracellular production of heterologous proteins. This is advantageous because secretion of proteins into the medium can greatly facilitate their purification. *S. cerevisiae* secretes very few native proteins, thus simplifying purification of secreted heterologous proteins (King *et al.*, 1989). As a result, a large number of other proteins are currently being produced in yeast fermentation systems for commercial sales. These include interferon, interleukin, blood proteins, and a variety of growth factors (Hodgson, 1993).

As with any expression system, the production of heterologous proteins in yeast has drawbacks. The secretion of proteins can be problematic because they are secreted into a large extracellular volume of growth medium, making it necessary to concentrate the proteins before they can be purified. Inefficient secretion may also occur, resulting in protein retention in the cell and periplasmic space, complicating protein purification and decreasing yields (De Nobel and Barnett, 1991). Yields of heterologous protein production in *S. cerevisiae*, with a few exceptions, can be disappointingly low. Even with a strong promoter, levels reach a maximum of 1-5% of total protein (Buckholz and Gleeson, 1991). As with bacterial expression systems, large-scale fermentative growth is costly, and the production of foreign proteins can provide stress on the organism, resulting in slower growth rates, especially if the protein exhibits toxicity towards its host. In addition, yeast often hyperglycosylate recombinant molecules with carbohydrate molecules distinct from those which are added by the source organism. *S. cerevisiae* may add more than 40 mannose residues to proteins it produces (Grinna and Tschopp, 1989). This is one of the major disadvantages of using *S. cerevisiae* as a protein expression system because it may result in difficulties in purification, decreased activity and/or increased protein immunogenicity (Heinrikson and Kezdy, 1991). Increased immunogenicity can pose a serious problem, especially if the protein is intended for pharmaceutical use. It has also

been found that the stability of autonomously replicating plasmids transformed into yeast (integrative transformation is a relatively rare event) is often reduced under fermentation conditions (Da Silva and Bailey, 1991).

Within the last decade, there has been a dramatic rise in the industrial use of other yeast species, most notably, the methylotrophic yeast *Pichia pastoris* and *Hansenula polymorpha.* Methylotrophic yeast grow on a simple, defined media utilizing methanol as their sole carbon source. These yeast, possess all of the desirable traits of *S. cerevisiae*, but also produce recombinant proteins at reliably higher levels (Faber *et al.*, 1995). Furthermore, these methylotrophic species typically glycosylate proteins with 8-14 units of mannose, a significant decrease to that found in *S. cerevisiae* (Grinna and Tschopp, 1989; Hodgson, 1993). In addition, production yields of >1g/L have been reported using *P. pastoris* and *H. polymorpha* (Laroche *et al.*, 1994; Weydemann *et al.*, 1995). Some proteins, such as single chain antibodies, are produced at significantly higher levels in methylotrophic yeast as compared to *E. coli.* For example, monoclonal antibodies secreted in shake flask cultures of *P. pastoris* yielded 100mg/L of active protein, a 100-fold increase to the yields obtained from *E. coli* (Ridder *et al,* 1995). Yields of this protein in *S. cerevisiae* were also disappointingly low. Although the large-scale production of recombinant proteins in methylotrophic yeast appears to have a number of advantages over that in *S. cerevisiae,* it should be noted that this novel expression system has not yet achieved the wide-ranging acceptance and the safety rating of its well-established counterpart (Sudbery, 1996).

1.2.3 Baculovirus-infected insect cells as hosts for recombinant protein production

The manipulation of baculovirus-infected insect cells for the production of heterologous proteins was pioneered by Max Summers' laboratory at Texas A&M University, Texas, in the late 1980's (Luckow and Summers, 1988). This was brought about by the realization that insect cells *(Spodoptera frugiperda)* infected with a baculovirus vector could produce milligram to gram quantities of soluble, biologically active, heterologous proteins per liter.

The best-characterized baculovirus which is used in expression systems is the *Autographa californica* nuclear polyhedrosis virus (AcNPV). The ability to produce recombinant baculovirus vectors is due to the fact that sequences encoding two viral polyhedron proteins may be replaced with foreign DNA sequences of interest without affecting viral particle formation. The baculovirus expression vector is a helper-independent, double-stranded, eukaryotic viral DNA which mainly infects lepidopterans (i.e. butterflies and moths) and their isolated cells. The most common cell lines are derived from the fall armyworm, *Spodoptera frugiperda,* and the cabbage looper, *Trichoplusia ni.*

Insect cells infected with the baculovirus expression vector may be grown adherently on flask surfaces or in shake flasks or fermenters (Gossen, 1992). The transformation of insect cells with recombinant baculovirus on a laboratory scale is receiving increased attention due to the availability of commercial kits and the existence of comprehensive laboratory manuals (Gruenwald and Heitz, 1996; King and Possee, 1992). By early 1995, approximately 500 recombinant proteins had been produced using this system (Patterson *et al,* 1995).

On a larger scale, this system has been used successfully to produce subunit vaccines, recombinant antibodies and a number of enzymes. MicroGeneSys (Meriden, CT) is using this technology to produce a number of different vaccines which are currently in clinical trial (Hodgson, 1993). The French biotechnology company, Proteine Performance,

assessed the yield of more than 130 proteins produced via baculovirus-infected insect cells and found it was up to 100-200 fold higher compared to their production in Chinese Hamster Ovary (CHO) cells (Hodgson, 1993).

Like yeast and mammalian systems, proteins expressed using this system may undergo post-translational modifications such as proteolytic cleavage (i.e. removal of signal peptides), glycosylation, phosphorylation, fatty acid acylation, and amidation (King and Possee, 1992; Luckow and Summers, 1988). In addition, improvements in vectors has allowed recombinant protein expression to occur intracellularly, or to be secreted into the culture medium (King and Possee, 1992). However, the reliability of some of these processes are often questionable, and the development of this novel system for the production of therapeutic proteins is going to require more rigorous experimentation to ensure the safety of heterologous proteins produced by it.

1.2.4 Mammalian cell culture for recombinant protein production

Numerous mammalian cell culture systems exist, such as mouse L cells, various myeloma cell lines, baby hamster kidney cells, and human embryonic kidney cells. The most widely used cell culture system employs the use of Chinese Hamster Ovary (CHO) cells. This fast-growing cell line can be maintained in either suspension culture or on a monolayer. A number of expression vectors are available for use in CHO cells. Most contain strong, constitutive promoters, typically of viral origin, to ensure maximal expression of the recombinant DNA. CHO cells have been characterized extensively at the molecular and physiological level. This familiarity is a valuable trait to biotechnology companies that want to market a recombinant protein, because it is often conducive to regulatory acceptance. One of the most popular CHO cell lines for recombinant protein production are those deficient in dihydrofolate reductase ($DHFR^-$) (Urlaub and Chasin,

1980). Dihydrofolate reductase catalyzes the conversion of folate to tetrahydrofolate. This cell line can be stably transformed (most commonly via calcium-phosphate co-precipitation and cation lipids) with a construct containing a DHFR expression vector. Cells containing this construct may be selected by incubating the cells in methotrexate (MTX), a folate analog which inhibits DHFR activity. MTX selects for cells in which the DHFR expression vector (containing the gene of interest) undergoes gene amplification (resulting in higher levels of expression of DHFR and the gene of interest) (Schimke, 1988). Only cells expressing high levels of vector-derived DHFR will survive in MTX. In practice, however, intermediate amounts of the selective drug are used in selectable marker systems because cells which must amplify their DNA grow more slowly, and the amount of heterologous protein produced in latter rounds of selection are smaller. Thus, drug levels which select for cells containing fewer gene copies can be more advantageous (Simonsen and McGrogan, 1994).

Other amplification systems having selection capabilities exist, such as the glutamine synthetase (GS) system. Glutamine synthetase is an enzyme which is produced only in low levels in murine myeloma cells. Initially, GS is used as a selectable marker in host cell selection. Amplification of copy number in cells transfected with a vector containing the gene of interest/ glutamine synthetase gene can be achieved by selection with the GS inhibitor, methionine sulfoximine (Bebbington *et al.*, 1992). Celltech (Slough, U.K.) first developed this system in the late 1980's and claims it has a number of advantages over the CHO/DHFR system. Using the GS system, they have achieved optimal amplification of gene copy number after only one round of amplification, compared to the 3-4 rounds typically required using the DHFR system, which saves approximately 3-4 months in cell line development. Additionally, use of the GS system results in high protein yields (500-1000 mg antibody/ml) (Bebbington *et al.*, 1992; Hodgson, 1993).

In 1993, an interesting comparison between the production of tissue plasminogen activator (tPA) produced via animal cell culture *vs.* bacterial fermentation systems was analyzed. This drug is used to treat acute myocardial infarction, deep vein thrombosis, pulmonary embolism and stroke. Its heterologous expression was originally developed in *E. coli*, although mammalian cell-derived tPA is now also being manufactured and marketed. Recombinant tPA produced in Chinese Hamster Ovary (CHO) cells was essentially identical to that found in humans and production in a growth fermenter yielded 33.5 mg/L. *E-coli*-derived tPA accumulated to 5-10% of total cell protein, yielding 300-700 mg/L, however, this was in an insoluble, denatured, form (Datar *et al.*, 1993).

Mammalian cell cultures are an attractive system for the production of heterologous proteins because they are capable of specific post-translational modifications and secretion of folded, biologically active proteins (Bebbington and Hentschel, 1985). However, these cells are difficult to manipulate genetically and produce recombinant proteins often at a much higher cost relative to yeast and bacterial expression systems (Heinrikson and Kezdy, 1991). In addition, heterologous proteins are often produced at much lower quantities compared to these systems. Nevertheless, the expression of a number of proteins in Chinese Hamster Ovary , mouse, and murine myeloma cells is being developed, including immunomodulators (i.e. interferon-beta and -gamma), growth factors (i.e. epidermal growth factor), human growth hormone, tissue plasminogen activator (tPA), and hepatitis B vaccine (Hodgson, 1993).

1.2.5 Animal expression systems

Recently, the use of transgenic mammals as a viable alternative for the large-scale production of heterologous proteins has been explored. The most common animal used for recombinant protein production is mice, although larger animals, such as sheep, goats, pigs

and cows, having the potential to produce greater, and more accessible amounts of protein (via milking) have been used. Besides their ability to produce large amounts of easily obtainable recombinant protein, the use of animals as bioreactors is advantageous because they are capable of performing a wide variety of post-translational modifications, and farms, compared to industrial high-volume fermenters, have a low capital cost.

There are two main systems of recombinant protein expression in animals, via the mammary glands, or in the blood of transgenic animals. The production of very high levels of milk-borne recombinant protein in large animals has been attained. Wright *et al.* (1991) generated transgenic sheep producing greater than 60g/L of active human α-1-antitrypsin in their milk. Human deficiency of α-1-antitrypsin, a common hereditary disorder, causes life-threatening emphysema. The recombinant α-1-antitrypsin produced in these sheep appeared to be fully glycosylated with a biological activity indistinguishable from that derived from human plasma. The production of a wide variety of other proteins, often to high levels (> 1mg/ml), in the milk of transgenic animals has also been achieved. These include active human urokinase, human serum albumin, human tissue plasminogen activator, human and bovine growth hormone, human CD4, human extracellular superoxide dismutase, human lactoferrin, human surfactant protein C, and human factor IX (rev. in Echelard, 1996).

As with other protein expression systems, however, the use of animals as bioreactors can be problematic. The generation of transgenic animals is a slow, labour-intensive and costly process. The production of transgenic animals is also inefficient. For example, in the experiments of Wright *et al.* (1991), the generation of transgenic sheep expressing human α-1-antitrypsin required the microinjection of 549 eggs with donor DNA, and their implantation into 152 recipients. Of the 113 lambs born, only 5 were

transformed with the human α-1-antitrypsin construct. In addition to these limiting factors, careful consideration must be given to the choice of heterologous protein produced, such that protein toxicity or other deleterious effects are not suffered by its host. Besides technical challenges, the use of animals as bioreactors has a greater potential to be on the receiving end of negative public opinion than the other expression systems described.

1.2.6 Plant-based expression systems

Over the last decade, plant biotechnology has grown into a large and commercially valuable industry. Crop plants are being produced which have improved or unique agronomic qualities, such as herbicide tolerance, insect resistance, and improved protein, oil, and carbohydrate composition (rev. in Goddijn and Pen, 1995; Murphy, 1996). The economic impact of these types of plants is significant. For example, the war against insect pests has been fought by approximately $10 billion dollars worth of chemical insecticides per year (Oerke, 1994). The production of plants resistant to destructive insect pests would undoubtedly be met with significant financial rewards. In addition, the use of transgenic plants is currently being explored as a means of reclaiming land contaminated with industrial waste. Phytoremediation, using a variety of naturally-occurring plants has been used successfully for the removal of trichloroethylene, 2,4,6-trinitrotolulene (TNT) and hexahydro-1,3,5-trinitro-1,3,5-triazine (RDX) (rev. in Boyajian and Carreira, 1997). Recently, *Arabidopsis* transformed with an altered mercuric ion reductase (*merA*) gene was found to remove mercury from the soil and release it into the atmosphere (Rugh *et al.*, 1996). In addition, a wide variety of novel recombinant proteins are being produced in plants, such as vaccines, pharmaceuticals, antibodies and industrial enzymes. Currently there are almost 2000 projects, having direct applications to crop biotechnology, in progress in 15 European countries. The estimated growth of the agrobiotechnology

industry of more than 30% per annum and is expected to be worth more than 1 billion dollars (US) by the year 2000 (Lloyd-Evans and Barfoot, 1996).

The ability to transform and regenerate plants with DNA sequences of interest has been instrumental to the production of foreign proteins in plants. Plants represent a unique biological system because many are capable of regenerating into whole organisms from single cells. Such totipotency is observed in a wide variety of plants, including those of economic importance, such as tobacco, petunia, tomato, *Brassica*, petunia, and soybean. A wide variety of techniques currently exist that facilitate the introduction and expression of transgenes in plant cells. The most notable of these is *Agrobacterium*-mediated transformation, which results in the stable and transmissible integration of heterologous DNA into plant cells. This method has been successfully used in the transformation of a wide variety of dicotyledonous plants (rev. in Klee *et al.*, 1987). Other methods include the use of electroporation, chemicals, such as PEG, viral vectors, microinjection, and particle bombardment. Recent technical advances have allowed the transformation of formerly recalcitrant species, such as cereals, grains, legumes and conifers. For example, the efficiency of *Agrobacterium*-mediated transformation of rice has increased over 100-fold, from approximately 0.2% to approximately 20% (Hiei *et al.*, 1994). *Agrobacterium*-mediated transformation of maize is also being developed. Microparticle bombardment-mediated transformation of barley has increased in efficiency by more than 10-fold within the last few years (Wan and Lemaux, 1994). Corn, wheat, oats, sugar cane, and soybean have also been successfully transformed using this technique (rev. by Walden and Wingender, 1995). A large number of facilities, such as the Plant Biotechnology Institute (PBI), Saskatoon, Texas A&M University and the universities of Georgia, Kentucky, Iowa, Nebraska, etc., which specialize in the transformation of target crops, currently exist. The ease by which DNA sequences may be transferred and stably integrated into the

plant genome, the ability of plants to be regenerated from single cells, and the ease in selection of regenerated plant transformants has been integral to the understanding of a wide range of areas in plant physiology and molecular biology. This, coupled with the fact that plants are an inexpensive source of biomass, are capable of performing post-translational modifications, and the machinery, land, and expertise needed to plant and harvest large quantities of crop plants already exists, has aroused the interest of biotechnology companies.

The production of transgenic plants having agronomically attractive qualities is now a commercial reality. Plants are now receiving attention as alternative systems for the production of high-value heterologous proteins, due to the limitations of other expression systems, as discussed above (sections 1.2.1-1.2.5). One of the earliest examples of recombinant protein production in plants was the expression of antibodies in transgenic tobacco. Hiatt and colleagues (1989) produced two lines of tobacco transformants expressing full-length cDNAs encoding mouse gamma or kappa immunoglobin chains (including the mouse signal sequence). Upon crossing of the two plant lines, functional, correctly assembled recombinant gamma-kappa complexes accumulated to 1.3% of the total leaf protein. Specific antibody binding by the plant-derived immunoglobin gamma-kappa complexes was equivalent to that obtained with the corresponding mouse monoclonal antibody. Since then, numerous other examples of plant-derived antibody production, including the production of multimeric secreted antibodies, have been published (rev. in Ma and Hein, 1996). Plants represent an attractive system for antibody production because they are capable of performing post-translational modifications such as ER-assisted proteolytic processing of signal sequences, folding, assembly, glycosylation and secretion. This results in the production of full-length antibodies which is not achievable in bacterial expression systems. Based on 1990 costs, the production of antibody expressed in

soybean at a level of 1% of total plant protein, would cost approximately $100 U.S./kg of antibodies (Hiatt, 1990). This estimate could decrease significantly with improved isolation/purification procedures.

The identification of the immunogenic agents known to cause disease is important in the development of effective and non-infectious vaccines. Plants have recently become an emerging tool in the expression of these immunogenic proteins. A short linear epitope of the VP2 capsid protein of mink enteritis virus (MEV) occurs in three highly infectious animal virus species- MEV, canine parvovirus, and feline panleukopenia virus. Dalsgaard and colleagues (1997) fused a gene encoding this epitope to a gene encoding a plant pathogenic virus coat protein (cowpea mosaic virus). This fusion construct was inserted into an infectious cowpea mosaic virus cDNA clone, and used to infect the bean *Vigna unguiculata*. The recombinant virus, which replicated *in planta*, was isolated and injected into animals susceptible to the virus. Animals inoculated with plant-derived, chimeric virus particles developed no sign of clinical disease when challenged with the infectious form of the virus. It has also been demonstrated that hepatitis B surface antigen expressed in, and isolated from, transgenic tobacco, elicits an immunogenic response in mice which mimics the effect of commercial vaccine (Thanavala *et al.*, 1995).

An alternative approach to plant-based production of antigens for the purpose of oral immunization is to express the protein of interest in edible plants. The cost of purification of recombinant proteins or peptides has generally been considered a limiting factor for the commercial application of plant-based expression systems. Expression of immunogens in this manner would negate any requirements for protein purification and reduce production costs considerably. Transgenic tobacco and potato expressing the binding subunit of *E. coli* enterotoxin (which is structurally, functionally and antigenically

similar to cholera toxin) have been shown to be an effective oral vaccine in mice (Haq *et al.*, 1995).

Pen *et al.* (1993) also circumvented the need to purify recombinant proteins by expressing phytase in seeds to be used for animal feed. Phytate, the main seed storage form of phosphorus, is essentially indigestible by monogastric animals, thus, phosphate must be added as a dietary supplement. Phytase was expressed in transgenic tobacco and shown to release phosphate in trials mimicking conditions of the chicken gut. Unfortunately, however, the number of proteins which are useful in an unpurified form is limited.

Although the production of foreign proteins in plant cells has been accomplished, the expression and accumulation of heterologous proteins in the cell cytosol can have deleterious effects. For example, enzymes responsible for polyhydroxybutyrate (PHB) synthesis were transformed into *Arabidopsis*. The resulting plants produced polyhydroxybutyrate, however, the levels of this polyester were low (20-100μg/g fresh weight), accumulated in the nucleus, cytosol, and vacuoles, and the plants were severely stunted (Poirier *et al.*, 1992). However, when the enzymes responsible for PHB formation were targeted to plastids, where the precursor for PHB (acetyl-CoA) is abundant, transgenic plant growth and fertility was unaffected and PHB accumulation increased 100-fold (Nawrath *et al.*, 1994). These experiments illustrate the need to give careful consideration of the spatial accumulation of heterologous proteins.

In order to avoid complications resulting from heterologous accumulation in plants, one strategy has been to target foreign proteins to storage organs, such as seeds and tubers. Seeds possess a number of characteristics which make them highly suitable as targets for the expression of recombinant proteins. Plant seeds represent an extremely inexpensive source of protein. For example, oilseed rape seed, which is approximately 30% protein by dry weight, costs approximately $250 per metric tonne (Ewins, 1992) to produce. Upon

maturation, seeds lose greater than 95% of their water content (Goldberg *et al.*, 1989). therefore seed proteins are already in a very concentrated form. In addition, seeds have low hydrolytic activity (Wilson, 1986), minimizing protein degradation and seeds can be stored for long periods of time without deterioration.

The suitability of seed storage proteins for use as fusional "carriers" of target polypeptides has been explored. In 1989, Vandekerckhove and colleagues replaced a region of an *Arabidopsis* 2S albumin gene with a sequence encoding the neuropentapeptide, enkephalin. *Brassica napus* and *Arabidopsis* plants transformed with this internal gene fusion yielded up to 50 and 200 nmol enkephalin/gram seed, respectively. Although these results were promising, it is probable that this system is inappropriate for the expression of larger proteins. Most seed storage proteins must be co-translationally transported to the lumen of the endoplasmic reticulum (ER) where signal peptides are cleaved (Chrispeels, 1984). From here they are passed to the Golgi complex and eventually segregated into protein bodies (Chrispeels *et al.*, 1982; Herman *et al.*, 1986). These proteins may also be glycosylated (Herman *et al.*, 1986) and frequently undergo further proteolytic cleavage at some stage in the process (Krebbers *et al.*, 1988). Obviously, these processes could adversely affect the biological activity of the heterologous protein. Furthermore, the fusion of a heterologous protein to storage proteins could interfere with normal protein processing. For example, experiments by Hoffman *et al.* (1988), showed that a modified version of the seed storage protein, phaseolin, was retained within the ER and/or Golgi and did not accumulate in the protein bodies of transgenic tobacco. Protein retention was possibly due to changes in the secondary or tertiary structure of the protein. As a result, degradation of this modified protein occurred, most likely in the Golgi apparatus, or, possibly, within the protein bodies. In addition to these potential limitations,

extraction and purification of recombinant proteins made as seed storage protein fusions is impractical for large-scale peptide production (Vandekerckhove *et al.*, 1989).

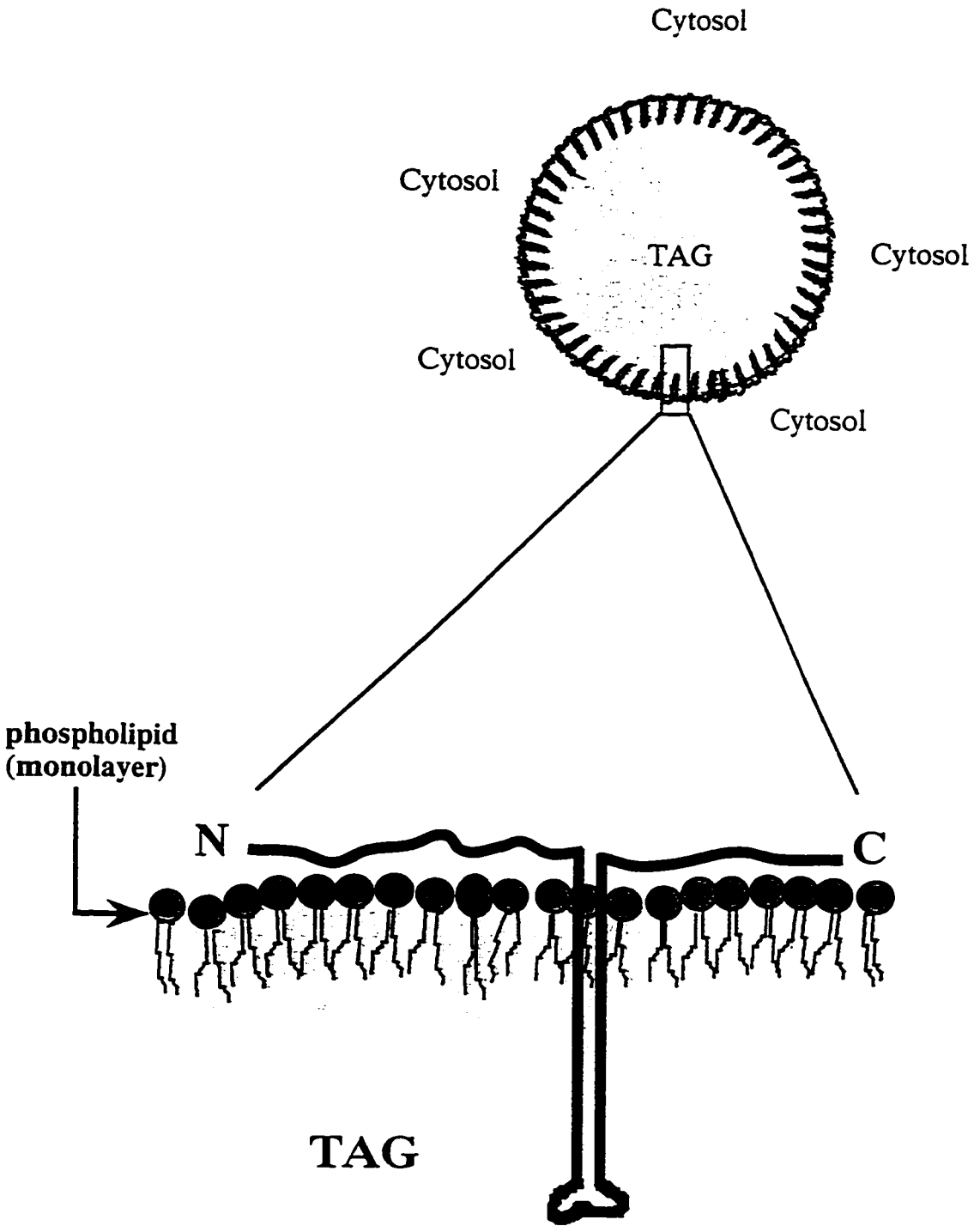
If plant systems are to become viable alternatives for the large-scale production of recombinant proteins, methods must be developed to ensure ease of purification and high yield of the recombinant protein, while providing a flexible expression system. The experiments outlined in this thesis are designed to test the suitability of a seed-specific protein, oleosin, for use as a fusional "carrier" of the blood anticoagulant, hirudin.

1.3 Oleosins

Seed triacylglycerols (TAGs) provide an essential energy reserve for germination and early growth of many plant species. Free fatty acids are synthesized in plastids then transferred, as fatty acyl-CoA, to the ER. Here they are assembled by acyltransferases onto a glycerol-3-phosphate backbone (via the Kennedy pathway) to form TAGs (Kennedy, 1961). Stored seed TAGs can account for 5-40% of the total dry seed weight. TAGs may be stored primarily in the cotyledons of the embryo (e.g. oilseeds such as canola (*B. napus*), sunflower, and linseed), the endosperm (i.e. castor bean, carrot, coriander), or both (e.g. tobacco). In all of these plants, the TAGs are not stored as naked lipid droplets, but are instead sequestered in subcellular organelles called oil bodies. Isolation of oil bodies from 18 different monocot and dicotyledonary species has shown that these organelles are spherical and have an average diameter of approximately 0.65-2μm, depending on the species (Tzen *et al.*, 1993).

It is believed the structural integrity of oil bodies is maintained by proteins, called oleosins, which are embedded in the half-unit phospholipid monolayer which surrounds the TAG (Fig. 1.1). Subcellular fractionation and immunohistochemistry has shown that seed oleosins associate exclusively with the oil body (Herman, 1987; Murphy and

Figure 1.1 An oil body (shown at top) magnified to reveal the conformation of an oleosin embedded in the phospholipid monolayer/TAG (triacylglycerol) matrix (diagram not to scale). The N- and C- terminal domains of oleosins face the seed cytosol.


Cytosol
Cytosol
TAG
Cytosol
Cytosol
Cytosol
phospholipid
(monolayer)
N
C
TAG

Cummins, 1989; Murphy *et al.*, 1989). Oleosins are believed to provide a physical barrier which prevents the coalescence of neighbouring oil bodies. Tzen and Huang (1992) showed that oil bodies isolated from maize were resistant to coalescence even after treatment with phospholipase A2 or C. The structural integrity of the subcellular organelles was disrupted, however, if the oil bodies were treated with trypsin. This latter treatment resulted in the digestion of oleosins and coalescence of oil bodies. Trypsin-treated oil bodies were also susceptible to lipase action. These experiments demonstrated the importance of oleosin in maintaining oil body integrity. By maintaining the small size of the oil bodies, a high surface:volume ratio is preserved. This is important for rapid oil mobilization to occur during seed germination.

In addition to providing a protective barrier, oleosins are also believed to be important to the stabilization of lipid bodies during seed desiccation and may act as a lipase docking site for oil mobilization (Vance and Huang, 1987; Murphy *et al.*, 1991). Insight into the role of oleosins in seed storage TAG can be inferred by comparing it to storage TAG in the mesocarp of avocado, olive, and oil palm. Oil stored in the mesocarp of these plants is not used as an energy reserve for the germinating seed, nor does it undergo desiccation like seeds. In these examples, TAG is stored as large lipid particles (>10-20μm) which are not surrounded by detectable oleosins (Mohankumar *et al.*, 1990: Ross *et al.*, 1993).

Over the past few years, oleosin proteins and genes have been characterized at the biochemical, cellular and molecular levels in gymnosperms and angiosperms, including numerous monocot and dicotyledonary species (rev. by Murphy, 1993). Oleosins are small proteins, ranging in size from 16-26 kDa among diverse plant species. Common to all seed oleosins described to date, is the presence of three distinct domains. The N- and C-terminal domains are highly divergent in both amino acid sequence and length, accounting for most

of the variability in oleosin protein size. Both these domains are exposed to the cytosolic side of the oil body (Abell *et al.*, 1997). The overall hydrophilicity and predicted secondary structure of these three domains was compared between seed oleosins of 9 different plant species (Tzen et al., 1992). The N-terminal domain is amphipathic and has no consistent secondary structure. The C-terminal domain is also amphipathic, but forms an α-helix stretch containing alternating basic and acidic amino acids. The basic amino acids are believed to interact with the negatively charged phosphate moieties of the oil body membrane, while the acidic amino acids face the cytosolic side of the oil body, resulting in a negatively charged surface. Thus, in addition to providing a physical barrier, negatively charged oleosins on the surface of the oil body provide electrical repulsion, which is believed to prevent oil body coalescence in the cell cytosol. It is these factors which are also believed responsible for the ease of oil body recovery. Intact oil bodies may be isolated by grinding seeds in an aqueous buffer (pH 7-8), and centrifuging the mixture (flotation centrifugation). This results in the flotation of a "fat pad" (an emulsified oil body layer). If the pH is lowered to 5, however, oil body aggregation becomes apparent (Kühnel *et al.*, 1996), presumably due to the protonation of histidine residues which results in a net neutral charge (Tzen *et al.*, 1992).

Common to seed oleosins from diverse plant species is the presence of a highly conserved, hydrophobic stretch of amino acids in the central domain. This region is composed of approximately 70 amino acids which are believed to form an anti-parallel β-structure which lies amid the hydrophobic interior of the oil body. The anti-parallel β-structure is believed to be stabilized by hydrogen bonding between opposite β-strands (Tzen *et al.*, 1992).

Another common feature of all oleosin proteins analyzed to date is the absence of a cleavable N-terminal signal sequence or ER targeting signal. Nevertheless, it has been demonstrated that oleosins are synthesized on polyribosomes bound to the ER membrane (Qu *et al,*. 1986). It has been suggested that a region in the N-or C-terminal domain may form a secondary structure which signals targeting to the polyribosomes bound to the ER. Alternatively, it could be the conserved, hydrophobic central domain which directs oleosins to the ER. It is interesting to note that maize oleosins correctly target to the oil body membrane of transgenic *B. napus* (Lee *et al.*, 1991). This occurred, despite the fact that *Brassica* and maize oleosins share little sequence identity at their N- and C- domains, and that these two species are dicotyledonous and monocotyledonous, respectively. More recently, oleosins from maize were shown to correctly target to lipid bodies of transformed yeast (Ting *et al.*, 1997). The ability to correctly target oleosin to lipid bodies in such heterologous systems suggests that oleosin targeting to the ER membrane is universal. This may occur via an undefined, uncleaved signal sequence or may be directed to the ER due to the nature of its secondary structure, charge distribution, or hydrophilicity/hydrophobicity.

Insight into the role each domain plays in subcellular targeting of oleosins to the oil body was recently investigated (van Rooijen and Moloney, 1995a). Deletion mutants of the coding region of an *Arabidopsis* oleosin were fused upstream of a reporter gene (GUS; β-glucuronidase) and transformed into *B. napus*. GUS activity from these mutant oleosin fusions, which lacked either the N-, central, or C-terminal domain of oleosin, was measured and compared to a complete oleosin-GUS construct. Deletion of the N-terminal oleosin domain resulted in 10-fold loss of GUS activity compared to the complete oleosin-GUS fusion protein, while deletion of the central hydrophobic domain resulted in a 20-fold loss of GUS activity. Deletion of the C-terminal domain resulted in 2-fold loss of GUS activity. Determination of GUS activity which remained associated with isolated oil bodies

showed 60% was targeted to the oil body of N-terminal oleosin deletions while only 40% of the central domain deletions were targeted to the oil body. This demonstrated that both the N-terminal and central domains are important for the stability and localization of oleosin to the oil body.

At the centre of the central hydrophobic domain of oleosins are three proline residues which form the turn of the anti-parallel β-strands. This "proline knot", which spans 12 residues, is the most conserved region of the oleosin. The importance of the proline knot for localization/stability of oleosins to oil bodies was explored by Abell *et al.*, (1997). The coding region of an *Arabidopsis* oleosin was modified such that the proline-encoding residues of the proline knot were substituted with leucines. *In vitro* translation experiments using this mutagenized oleosin mRNA, and wild type oleosin mRNA, demonstrated that the nascent protein accumulated to equivalent levels for both mRNA types. In addition, fractionation experiments from transformed embryos expressing a wild-type oleosin-GUS and proline knot-mutant oleosin-GUS showed that both accumulated to equivalent levels in the microsomal fraction. This demonstrated that the proline knot-mutant oleosin is translocated to the ER efficiently and the proline knot is not essential for protein targeting to the ER. However, transient and stable expression of a gene encoding the proline-mutant oleosin fused to GUS showed a marked decrease in oleosin accumulation and targeting to the oil body compared to wild type oleosin-GUS. These results demonstrated that the proline knot plays a role in the stability and/or insertion of oleosins on oil bodies.

It is agreed that the enzymes responsible of TAG formation, such as diacylglycerol acyltransferase (DAGAT), exist in the endoplasmic reticulum (ER). However, the site of deposition of the TAG destined for the oil body, in addition to the time at which oleosins are inserted into the half-unit phospholipid membrane which surrounds the TAGs is the

source of considerable debate. Two hypotheses have been proposed. One suggests that oil accumulates outside of the ER as naked droplets, after which time phospholipids and oleosins surround and embed into the TAG droplet. The other hypothesis suggests that the hydrophobicity of the TAGs results in their deposition between the two phospholipid layers of the ER membrane. Continued TAG accumulation between the membranes and cotranslational insertion of oleosins leads to the budding off of a lipid particle surrounded by a phospholipid monolayer (derived from the ER).

The first hypothesis, is based on the results of two papers (Murphy *et al.*, 1989; Cummins *et al.*, 1993). *B. napus* embryos were analyzed for their accumulation of TAG and oleosins from 2-12 weeks after flowering, and in dry seeds. While TAG accumulation occurred 3 weeks after flowering and continued to increase to seed desiccation, only trace amounts of oleosin was detected 6 weeks after flowering, with rapid accumulation occurring after 10 weeks. Here, the investigators found oleosin accumulation to follow that of the storage proteins napin and cruciferin. Immunogold labeling of 8 week old embryos showed only weak immunogold staining with anti-oleosin antibodies, although it was uncertain whether these "naked" oil bodies possessed a phospholipid coat. Oil bodies isolated from young embryos contained very little oleosin protein, and these oil bodies coalesced after dehydration, a result which did not occur when oil bodies from dry seeds were used.

In contrast to the above findings, other researchers have shown that oleosins accumulate much earlier in development (at the globular-heart stage) in *B. napus* microspore-derived embryos (Holbrook *et al.*, 1991), and this accumulation appears to occur concomitantly with that of TAGs in maize and *B. napus* (Qu *et al.*, 1986; Tzen *et al.*, 1993). Denis Murphy's group has argued that the experiments of Qu *et al.* (1986) did not include maize embryos beyond 7 weeks after flowering, when the bulk of oleosin gene

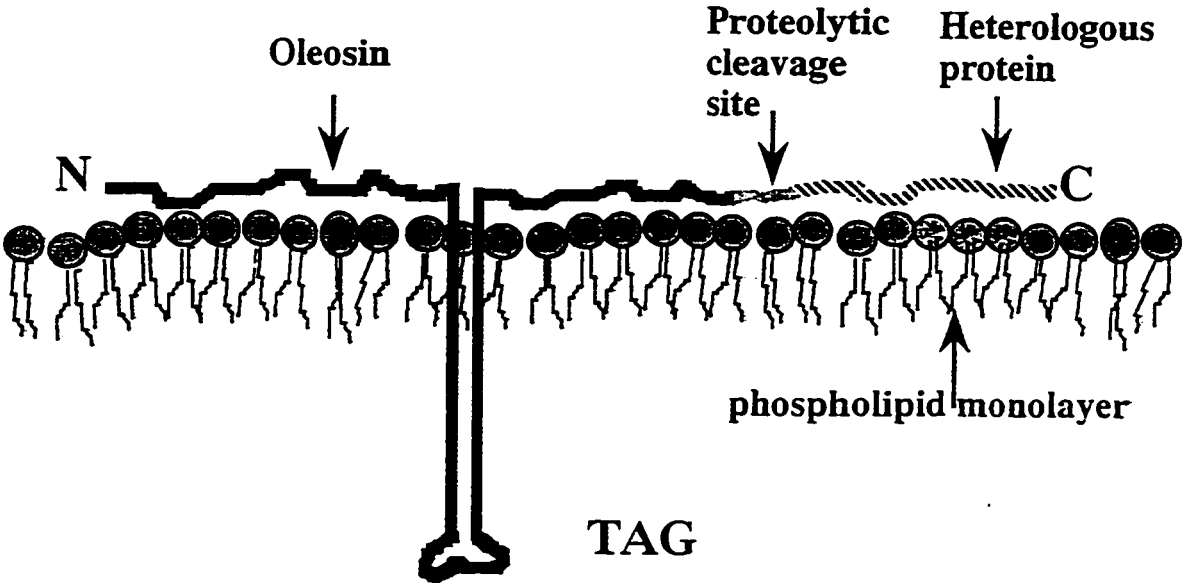
expression and protein accumulation occurs. Murphy contends that this group looked at only a small window of oleosin expression and accumulation, overlooking the bulk of oleosin which accumulates in the seed.

Ultrastructural analysis has shown physical contacts between the ER and oil bodies (Frey-Wyssling *et al.*, 1963; Schwarzenbach, 1971; Wanner *et al.*, 1981). This suggests that oil bodies are derived from budding of the ER membrane and not by assembly of naked oil droplets with a complement of phospholipids and oleosins. Unfortunately, the nature of these contacts, whether casual or indicative of oil body biogenesis, is unclear. Indeed, other research groups have failed to find convincing evidence that such contact even exists (Bergfield *et al*,. 1978; Herman, 1987).

To maximize stability and efficiency of purification and localization of a foreign protein expressed in a biological system, the gene of interest may be fused to all or part of a host gene. Gene fusions were first used for heterologous bacterial expression of small peptides, such as somatostatin (Itakura *et al.*, 1977) and insulin (Goeddel *et al.*, 1979) and more recently, in plant expression systems (Vandekerckhove *et al.*, 1989; Sijmons *et al.*, 1990). If a native protein is desired, site-specific cleavage of the fusion protein can be performed by either chemical or enzymatic methods.

As stated above, the oleosin protein has an N-terminal hydrophilic region, an internal hydrophobic domain, and a C-terminal amphipathic or hydrophilic region which is expected to reside in the cytosol (Fig. 1.1). It was reasoned that it might be possible to attach proteins (as fusions) to oleosins without adversely affecting their targeting (Fig. 1.2). If a heterologous protein is expressed as a C-terminal fusion to oleosin, it is likely it will reside on the cytosolic side of the oil body. In order to test the suitability of oleosins as a carrier protein for the large-scale production of high value proteins, an appropriate heterologous "target" protein gene must be chosen. Given the current interest in blood

Figure 1.2 Schematic representation of an oleosin-heterologous protein fusion inserted into the phospholipid monolayer of an oil body. The heterologous protein (hatched line) is fused to the C-terminal end of oleosin (solid black line), which lies on the cytosolic side of the oil body. This fusion protein is separated by a proteolytic cleavage site (dotted line). Abbreviations: TAG, triacylglycerol; N, N-terminus of oleosin; C, C-terminus of oleosin.


Oleosin
Proteolytic cleavage site
Heterologous protein
N
C
phospholipid monolayer
TAG

anticoagulants as biopharmaceuticals, hirudin was used as a model protein to evaluate this expression system.

1.4 Hirudin

Blood clotting (thrombosis) occurs as a cascade of proteolytic events involving the sequential conversion of zymogens (proenzymes) into active serine proteases (rev. by Mann, 1987). The final enzyme in this cascade, α-thrombin, converts fibrinogen into the insoluble, aggregate-forming protein, fibrin (K_i= 1μM). Alpha–thrombin consists of a small A chain linked by a disulfide bond to the B chain (Fenton *et al.*, 1977). The B chain is important for the specificity of α–thrombin to fibrinogen (Fenton et *al.*, 1988). In addition to fibrin formation, α–thrombin stimulates platelet aggregation and activates a number of proenzymes higher up in the thrombogenesis cascade (i.e. via positive feedback control). The medical significance of thrombosis, which can cause arterial and vascular occlusion, coupled with the central role thrombin plays in the blood coagulation pathway, illustrates the therapeutic value of anti-thrombin agents.

The anti-thrombin agent, heparin, is widely used in clinical therapy (Hirsh, 1991). This carbohydrate binds to the endogenous blood protein antithrombin III, enhancing its antiprotease activity and causing the rapid inactivation of thrombin (factor IIa), in addition to factor Xa and a number of other coagulation factors (rev. by Bettigole, 1992). The pleiotropic effects heparin has on the blood coagulation pathway is problematic, as it can result in severe side effects, such as hemorrhaging. Heparin is unable to inhibit thrombin bound to fibrin (Hogg and Jackson, 1989). As a result, patients may experience rethrombosis after angioplasty. In addition, heparin binds to, and is subsequently neutralized by, several plasma proteins which are released by activated platelets (rev. by Johnson, 1994). The neutralization of heparin in this manner may account for its

unpredictable dose response. These, and many other problems, have been a driving force in the search for a more specific thrombin inhibitor.

Leeches had been used and misused by the medical community of the 18th century as tool for blood-letting. Over the past two decades, research has shown that these animals possess qualities of true medical and pharmaceutical value. In order for continued feeding to occur, these animals must respond to and circumvent the hosts' natural blood clotting processes. In the late 1950's, the active agent responsible for the anticoagulant effect, hirudin, was isolated from the salivary glands of the leech *Hirudo medicinalis* (Markwardt, 1956). Hirudin is a small (approx. 7000 Da), highly acidic protein (pI= 3.9) (Markwardt and Walsmann, 1967) which forms a stable, 1:1 noncovalent complex with thrombin. The high specificity (Brown *et al*, 1980) and exceedingly low dissociation constant of hirudin-thrombin complexes (K_i= 27 fM) (Stone and Hofsteenge, 1986) makes it the most potent thrombin inhibitor known. Advances in recombinant DNA technology afforded the detection of hirudin isoforms and the availability of recombinant hirudin (rHir) for more rigorous biophysical and biochemical analysis. In the 1980's, the amino acid sequence of three naturally occurring hirudin variants (HV), designated HV1 (Dodt *et al.*, 1984), HV2 (Harvey *et al.*, 1986), and HV3 (Dodt *et al.*, 1986) were elucidated. These 65-66 amino acid polypeptides share ≥85% sequence identity and contain six cysteine residues at conserved positions in the N-terminal portion (residues 1-47) of the hirudin molecule. Elucidation of hirudin-thrombin conformation has been facilitated by X-ray crystallographic studies (Rydel *et al.*, 1990; Grutter *et al.*, 1990). The formation of three disulfide bridges between residues Cys^{6}-Cys^{14}, Cys^{16}-Cys^{28} and Cys^{22}-Cys^{39} creates a compact, apolar, N-terminus. The major role of the N-terminal domain is to orient amino acids 1-3, valine-valine-tyrosine (HV1) or isoleucine-threonine-tyrosine (HV2 and HV3) such that it hydrogen bonds with the active site serine of thrombin. This positioning is believed to

prevent substrate access, resulting in inhibition of thrombin activity. The C-terminus of hirudin (residues 49-69) appears to have no favored conformation when in the free state. This domain binds to a groove on the surface of thrombin, the fibrinogen-recognition exosite, through numerous electrostatic and hydrophobic interactions (Rydel *et al.*, 1990). Together, these two hirudin domains effectively inhibit all major thrombin activities.

Hirudin is of great medical interest, not only because of its high specificity for thrombin, but because it possesses a number of other characteristics which make it superior to the currently used blood anticoagulant, heparin. Pharmacological studies have shown that hirudin, apart from its anti-thrombin activity, is pharmacodynamically inert (rev. by Markwardt, 1994; Johnson, 1994). Unlike heparin, hirudin shows lower sustained bleeding after treatment (Markwardt *et al.*, 1991), does not require a cofactor for activation, and prevents reformation of blood clots after angioplasty or thrombolysis (Kaiser et al., 1990). The potential of foreign proteins to elicit an immune response is well known and of considerable concern in the development of new therapeutic drugs. Subcutaneous injections of various doses of recombinant HV1 to hundreds of volunteers and patients elicited very weak, or no, allergic response (Verstraete *et al*, 1993; Close *et al*, 1994). Hirudin is eliminated by the kidneys, showing no deposition in any organs, and has low toxicity (LD_{50}=500, 000 antithrombin units/kg in mice) (Markwardt *et al.*, 1982).

The limited availability of hirudin in leeches (one leech head contains approximately 20μg of hirudin) (Markwardt, 1970), has necessitated its production in heterologous expression systems. Biologically active recombinant hirudin (rHir) variants have been cloned and expressed in *E. coli* (Dodt *et al.*, 1986; Fortkamp *et al.*, 1986; Harvey *et al.*, 1986), yeast (Bischoff *et al*,. 1989; Lehman *et al*,. 1993; Loison *et al.*, 1988; Mendoza-Vega *et al.*, 1994; Riehl-Bellon *et al.*, 1989;), baculovirus-infected insect cells (Benatti *et al.*, 1991) and *Streptomyces lividans* (Bender *et al.*, 1990). Milligram to multigram

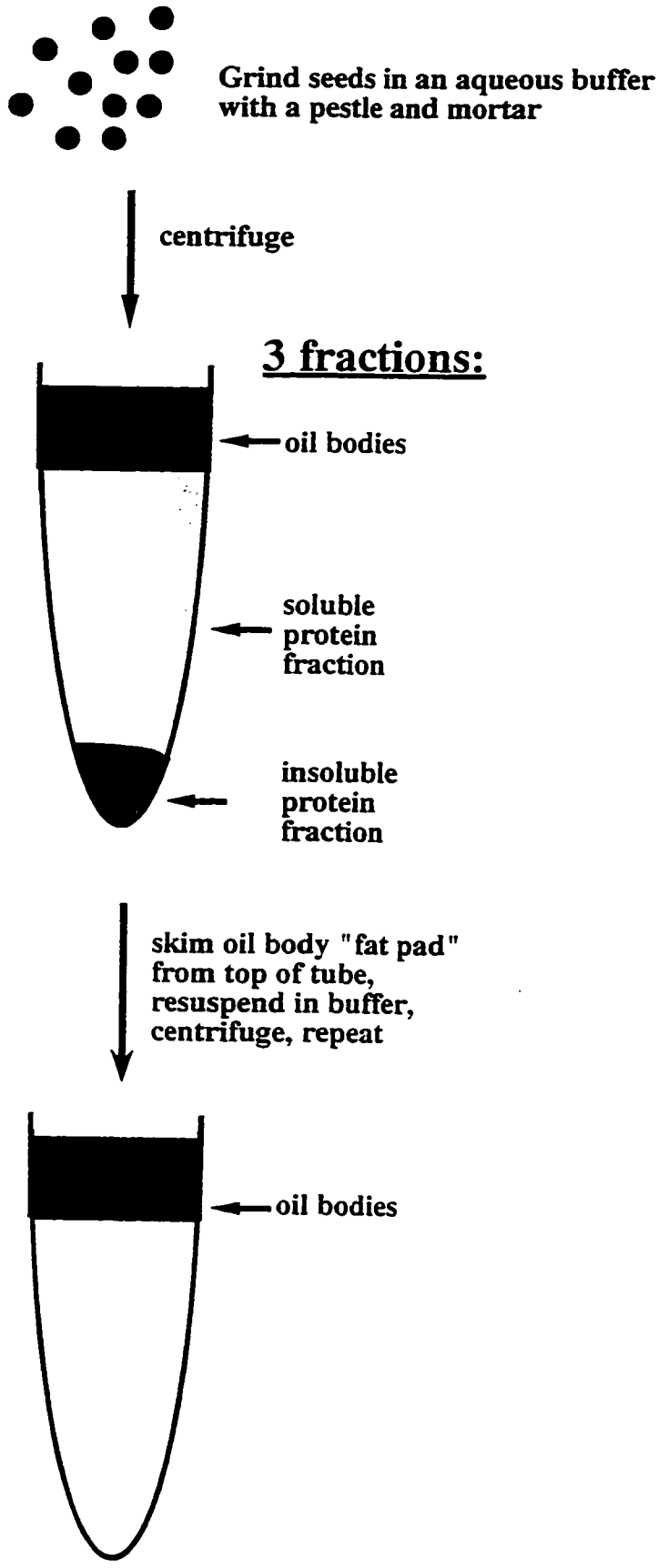
quantities of recombinant hirudin have been produced by these systems and purified to up to 99% homogeneity. Approximately 90% of leech-derived hirudin contains a sulfated tyrosine at position 63 (Dodt *et al.*, 1984; Chang, 1983). When desulfurylated, the affinity of hirudin for thrombin decreases 10-fold (Stone and Hofsteenge, 1986). To date, no heterologous expression system has produced recombinant hirudin in a sulfurylated form. Nevertheless, recombinant desulfatohirudin maintains its high specificity for thrombin, exhibits potent antithrombotic activity *in vivo* and *in vitro*, and is nonimmunogenic (Klocking *et al*, 1988; Johnson, 1994).

Hirudin is stable in the presence of strong denaturants, extreme pH, and high temperatures (Markwardt, 1970; Chang, 1991). The N-terminal core domain of hirudin is highly resistant to proteolytic digestion (Chang, 1991) and reduced and denatured hirudin spontaneously refolds into a biologically active protein in buffered solution (pH 8-8.7) (Chatrenet and Chang, 1992, 1993). In addition, the ability of hirudin to inhibit thrombin activity is easily quantitated by a colorimetric assay (Chang, 1983). The stability and ease of detecting its anti-amidolytic activity makes hirudin an ideal model protein for the assessment of novel expression systems.

1.5 Specific goals

Seed oleosins are proteins which embed into the phospholipid monolayer of subcellular organelles, called oil bodies, which house seed storage triacylglycerols. The following experiments are designed to test the suitability of seed oleosins to act as fusional "carrier" for the production of hirudin in seeds. Oleosins possess a number of important qualities which make them suitable for fusion with a heterologous protein. Firstly, seed oleosins are expressed solely in this organ, and no other part of the plant (Murphy *et al.*, 1989; Herman, 1987; Vance *et al.*, 1987). Thus, the expression of heterologous proteins, such as hirudin is unlikely to affect the growth and development of the plant. Secondly, oleosin expression can be induced by osmotic stress and abscisic acid (van Rooijen *et al.*, 1991). Thirdly, oil bodies with their associated proteins are easily separated from other seed components as an immiscible fat pad by flotation centrifugation (Qu *et al.*, 1986). As shown in Fig. 1.3, oil bodies are obtained by grinding seeds in an aqueous buffer, and repeatedly isolating, resuspending, and recentrifuging the oil body fraction, which is recovered as a "fat pad" from the top of a centrifugation tube. Thus, it is possible that the expression of an oleosin-hirudin fusion protein in seeds will not only stabilize the accumulation of the foreign protein, but facilitate its purification. The ability to isolate oil bodies using this method is owed largely to the extreme stability of oil bodies at pH 7-8, and the tight association of oleosins to the phospholipid monolayer of this subcellular organelle. Purification of heterologous proteins expressed as oleosin fusions may prove to be superior to the use of seed storage proteins as "carriers" since the latter are enclosed within protein bodies and are far more difficult to recover (Vandekerckhove *et al.*, 1989). Finally, oleosins account for approximately 10% of the total seed protein (Murphy and Cummins, 1989). This abundance suggests that heterologous proteins fused to them will be synthesized in large quantities.

Figure 1.3 Flow diagram of oil body isolation from seeds. Seeds are ground in an aqueous buffer using a pestle and mortar. The seed extract, upon centrifugation (13,000-14,000 g), is separated into three fractions: the insoluble protein/seed debris fraction (pelleted fraction), the soluble seed fraction (composed primarily of water-soluble storage proteins) (supernatant), and the oil body "fat pad" (uppermost fraction). The oil body fraction can be skimmed off the top of the supernatant with a spatula, resuspended in buffer, and recentrifuged. This process ("washing" of oil bodies) enriches this protein fraction such that it is comprised primarily of oil body associated proteins (oleosins).


Grind seeds in an aqueous buffer with a pestle and mortar
centrifuge
3 fractions:
oil bodies
soluble protein fraction
insoluble protein fraction
skim oil body "fat pad" from top of tube, resuspend in buffer, centrifuge, repeat
oil bodies

The importance of hirudin for use in the medical industry motivates research into its large-scale production via heterologous expression systems. The increasing knowledge of the oleosin gene family, their expression, and their gene products, in addition to the transformability of the well-studied plants, *Brassica napus* and tobacco, facilitate their use in expression systems. Based on this, my research has a number of specific objectives:

- to fuse, in frame, the gene encoding hirudin to the 3'end of the oleosin coding region
- to express this gene fusion, under the control of an *Arabidopsis thaliana* oleosin promoter in *Brassica napus* and tobacco in a seed-specific manner
- to determine the spacial expression, accumulation, and targeting of the fusion mRNA and protein in the above expression system
- to purify the fusion protein from the other seed components
- to efficiently cleave hirudin from its oleosin "carrier" by site-specific proteolysis
- to assess the accuracy of folding of the recombinant hirudin
- to measure the antithrombin activity of hirudin, both as a fusion to oleosin, and after proteolytic cleavage from this carrier protein.
- to increase the level of expression of the oleosin-hirudin fusion mRNA and protein through modifications of the promoter and 5' untranslated sequences.
- to increase the level of hirudin produced in *B. napus* seeds by expressing hirudin as concatamers fused to the C-terminus of oleosin

Chapter 2:

Materials and Methods

2.1 MATERIALS

2.1.1 Chemicals

Unless otherwise specified, all chemicals were of analytical grade and obtained from one of the following suppliers: BDH Inc. (Edmonton, Alberta); Boehringer Mannheim (Laval, Quebec); VWR Canlab (Mississauga, Ontario); Sigma (St. Louis, MO., USA); Bio-Rad Laboratories (Hercules, California); Gibco/BRL (Burlington, Ontario). Sucrose and bleach used in experiments was of the quality available in grocery stores.

2.1.2 Enzymes, nucleotides, radionucleotides, and growth regulators

Restriction endonucleases were obtained from Pharmacia Biotech Inc. (Baie d'Urfé, Quebec) and New England Biolabs Inc. (Mississauga, Ontario). Klenow DNA polymerase, T4 DNA ligase, and nucleotides were purchased from Pharmacia Biotech Inc. Sequenase™ and radiolabelled nucleotides were obtained from Amersham Life Sciences Inc. (Oakville, Ontario). Thrombin, clostripain, p-tosyl-gly-pro-arg-nitroanilide, and phytohormones were purchased from Sigma (St. Louis, MO., USA). Factor Xa and Vent™ DNA polymerase was purchased from New England Biolabs Inc. (Mississauga, Ontario).

2.1.3 Antibodies used for immunodetection

Rabbit antibodies raised against 19kDa *B. napus* oleosins were kindly provided by Jo Ross at John Innes Centre, U.K., and Jill Saponja and Dr. Joe Boothe of the University of Calgary. Mouse monoclonal antibodies raised against hirudin were generously donated by C. Koch and K. Gerner-Smidt at Statenserum Institute, Copenhagen, Denmark (Koch

et al., 1993). Alkaline phosphatase (AP)-conjugated secondary antibodies recognizing rabbit and mouse IgG were purchased from Bio-Rad and Boehringer-Mannheim, respectively. FITC-conjugated anti-mouse secondary antibodies were purchased from Sigma.

2.1.4 Plant material

Brassica napus cv Westar seeds were obtained from Agriculture Canada, Saskatoon. *Nicotiana tabacum* cv Wisconsin seeds were obtained from Dr. M. van Montagu, University of Ghent, Belgium.

2.1.5 Oligonucleotides

Oligonucleotides were synthesized using phosphoramidite chemistry on either the Millipore/Waters cyclone Plus or Beckman's oligo 1000M DNA synthesizer according to manufacturer's protocol. The oligonucleotides used in this work are summarized on Table 2.1.

2.1.6 *B. napus* cDNA

A cDNA library of polyA$^+$ RNA from 21-24 dpa *Brassica napus* cv Golden zygotic embryos in λgt11 expression vectors was kindly provided by Dr. Ljerka Kunst (Dept. of Botany, University of British Columbia).

Table 2.1 Oligonucleotides used for sequence and/or PCR analysis. 1) "+" refers to a leading strand (sense) oligonucleotide which is used to prime the lagging strand. 2) "-" refers to a lagging strand (antisense) oligonucleotide which is used to prime the leading strand.

Name	-mer	template	sense	sequence
Universal	17	pUC/M13/ pBluescript	+	5' GTA AAA CGA CGG CCA GT 3'
Reverse	17	pUC/M13/ pBluescript	-	5' AAA CAG CTA TGA CCA TG 3'
GVR06L	115	GVR06R/ synthetic hirudin variant 2	+	5' TGC GAT CGA AGG GAG AAT CAC TTA CAC TGA CTG TAC TGA ATC TGG ACA GAA CCT CTG TCT CTG TGA AGG ATC TAA CGT TTG TGG AAA GGG AAA CAA GTG TAT CCT CGG ATC TAA C 3'
GVR06R	146	GVR06L/ synthetic hirudin variant 2	+	5' GTT TGT GGA AAG GGA AAC AAG TGT ATC CTC GGA TCT AAC GGA AAG GGA AAC CAG TGT GTT ACT GGA GAA GGA ACT CCA AAC CCA GAA TCT CAC AAC AAC GGA GAC TTC GAA GAA ATC CCT GAA GAA TAC CTC CAG TAA GTC GAC GG 3'
GVR07	20	synthetic hirudin variant 2	+	5' TGC GAT CGA AGG GAG AAT CA 3'
GVR08	18	synthetic hirudin variant 2	-	5' CCG TCG ACT TAC TGG AGG 3'
HirV1L	114	HirV1R/ synthetic hirudin variant 1	+	5' GGG GTC GTC TAT ACC GAC TGT ACC GAG TCC GGT CAG AAC CTC TGT CTC TGT GAG GGT TCC AAC GTC TGT GGT CAG GGT AAC AAG TGT ATC CTCGGT TCC GAC GGT GAG AAG AAC 3'
HirV1R	111	HirV1L/ synthetic hirudin variant 1	-	5' CCC CTG GAG ATA CTC CTC TGG GAT CTC CTC AAA GTC ACC GTC GTT GTG GGA CTG TGG CTT TGG GGT TCC CTC ACC GGT GAC ACA CTG GTT CTT CTC ACC GTC GGA ACC GAG 3'
HV1-5'	66	synthetic hirudin variant 1	+	5' GCA GCA GCA GCA GTC GAC GAG CAG GAG CAA GAG GAT ATC GAG GGT AGA GTC GTC TAT ACC GAC TGT 3'
HV1-3'a	42	synthetic hirudin variant 1	-	5' AGC AGC AGC AAG CTT AGA TCT TTA CTG GAG ATA CTC CTC TGG 3'

Name	-mer	template	sense	sequence
HV1-3'b	39	synthetic hirudin variant 1	-	5' TGC TGC TGC AAG CTT TCG CGA CTG GAG ATA CTC CTC TGG 3'
HV1-3'c	52	synthetic hirudin variant 1	-	5' TGC TGC TGC AAG CTT AGA TCT TTA TCT ACC CTC GAT GGA CTG GAG ATA CTC CTC TGG 3'
Bam-AMV-col	66	oleosin coding region	+	5' CGC CGC GGA TCC GTT TTT ATT TTT AAT TTT CTT TCA AAT ACT TCC ACC ATG GCG GAT ACA GCT AGA 3'
oleo(+)133	21	oleosin coding region	+	5' AAC ATG TAT GGT CGA GAC TAC 3'
oleo(+)364	18	oleosin coding region	+	5' GCA GCT ATA ACC GTC TTC 3'
cOL3'Sal	36	oleosin coding region	-	5' GTC GTC GTC GTC GAC GGT AGT GTG CTG GGT TCC ACG 3'
GVR25	18	oleosin promoter	+	5' AAC ACT CCT ACC TCT TCC 3'
GVR26	18	oleosin coding region	+	5' AGA CAG AGC TCA GTA CTA 3'
GVR10	24	oleosin promoter	+	5' CAC TGC AGG AAC TCT CTG GTA AGC 3'
GVR11	71	oleosin coding region	-	5' CCG TCG ACT TAC TTG TCG TTA GAT TCT TCT CCC TGA ACT CTC CCT TCG ATC GCA GTA GTG TGC TGG CCA CC 3'
nos 1780R	26	NOS terminator	-	5' CGC AAG ACC GGC AAC AGG ATT CAA TC 3'

2.2 METHODS

2.2.1 Recombinant DNA techniques

All standard recombinant DNA techniques (DNA digestion by restriction endonucleases, plasmid preparations, growth of bacterial cultures, preparation of media and buffers) were performed according to Sambrook *et al.* (1989).

2.2.2 Polymerase chain reaction (PCR)

Unless otherwise stated, amplification of DNA was performed in 100 µl volume using 2-100 ng of DNA template, 5µl each of appropriate primers (20 µM), 5µl total dNTPs (2mM), 1-2 units of *Thermus aquaticus* (*Taq*) or Vent™ DNA polymerase, and 10µl 10X buffer (100mM Tris-HCl (pH 8.3), 500mM KCl, 15mM $MgCl_2$). Amplification was achieved by repeating the 3 cycles: denaturation (92^0C for 1 min.) followed by annealing of primers (1 min), and extension of primers (72^0C for 1 min) for 30 cycles. Annealing temperature was determined according to the size and G-C content of the primers, as described by Erlich (1989).

2.2.3 DNA ligation

Restricted DNA fragments were ligated using DNA isolated with glass powder ("genecleaning") (Vogelstein and Gillespie, 1979) or by ligating in Low Melting Point (LMP) agarose gel.

Ligation of "Genecleaned" DNA: The preparation of glassmilk (glass powder suspension) and the elution of DNA from agarose was performed as described in Davis *et*

al. (1986). Briefly, DNA restricted by endonucleases was electrophoresed through 0.8-1.5% agarose. The band of interest was cut out with a razor blade and melted at 45-55°C, 5 minutes, in 3 volumes sodium iodide solution (final volume ≤ 1.5 ml). To this mixture, 5µl of glassmilk was added, incubated on ice for 5 minutes, then centrifuged for 5 seconds at 14,000g. The pellet was resuspended 3 times in an ethanol wash, recentrifuged, and dried. The pellet was resuspended in a small volume of TE, heated to 45-55°C, 3 minutes, centrifuged 0.5 minutes, and the supernatant (containing the DNA of interest) removed. This purified, concentrated DNA was used in ligations according to Sambrook *et al.* (1989).

Ligation of DNA fragments in Low Melting Point (LMP) Agarose Gel: Restricted DNA was electrophoresed through 0.8-1.5% LMP agarose gel and the bands of interest cut out with a razor blade. The gel slices were melted at 65°C, 2-3 minutes, 13-15µl of gel was added to enough water for a final volume of 53µl. This mixture was heated at 65°C, 7 minutes, then cooled to 37°C. Plasmid DNA was added, if necessary. Six microliters of 10X ligase buffer (containing ATP) and 1µl of T4 DNA ligase (5U/µl) was added and the mixture incubated at room temperature, overnight. This mixture was reheated to 65°C and cooled to 37°C before transforming into 200µl of competent bacteria.

2.2.4 Bacterial transformations

Competent *E. coli* strains, DH5α DH10B and JM109 (Gibco/BRL, Burlington, Ontario) were transformed with plasmid DNA using the heat shock method as described by Hanahan (1983).

2.2.5 DNA sequencing

DNA sequencing reactions were performed using the dideoxynucleotide chain termination method (Sanger *et al.*, 1977). Eight microliters of template DNA (1.5-2μg) was denatured by addition of 2μl of 2M NaOH. After incubating for 10 minutes at room temperature, the denatured DNA was precipitated in 95% ethanol. The DNA pellet washed in 70% ethanol, dried, and resuspended in 10μl water. Two microliters each of annealing buffer (280 mM Tris-HCl, pH 7.5; 100mM $MgCl_2$; 350 mM NaCl) and the appropriate oligonucleotide (0.5-2.5 pmol/μl) was added to the 10μl of denatured template DNA. This mixture was incubated in a 65^{0}C water bath, 10 minutes, and allowed to cool slowly to room temperature. Two microliters of labeling mix (2 μM each of dGTP, dTTP, and dCTP), 1μl of DTT, 1μl of α-^{35}S dATP and 3 units of T7 DNA polymerase (2μl) were added to the 14μl DNA/primer mixture, and incubated at room temperature. After five minutes, 4.3μl aliquots of the above reaction mixtures were added to each of 2.5μl of ddGTP, ddCTP, ddTTP and ddATP chain extension/termination mixtures (15μM ddGTP, ddCTP, ddTTP **or** ddATP, respectively, plus: 150μM each dGTP, dCTP, dTTP, dATP; 10 mM $MgCl_2$; 40mM Tris-HCl, pH 7.5; 50 mM NaCl) and incubated at 37^{0}C, 5 minutes. After 5 minutes, the reactions were stopped with 5μl stop buffer (95% deionized formamide containing 20mM EDTA, pH 7.5; 0.05% (w/v) xylene cyanol FF; 0.05% (w/v)

bromophenol blue). 3-4 µl of these sequencing reactions were run on a 6% acrylamide urea gel as outlined in Sambrook *et al.* (1989).

2.2.6 Transformation of *Agrobacterium tumefaciens* with binary plasmids

pCGN-1559 plasmids (McBride and Summerfelt, 1990) containing the gene of interest were electroporated into *A. tumefaciens* EHA-101 cells (Hood *et al.*, 1986) by pulsing the cell-DNA suspension at a field strength of 12.5 kV/ cm and a time constant of 4.5 to 5 msec (Dower *et al.*, 1988). Transformed cells were plated on Luria Broth (LB) containing 100µg/ml kanamycin and 100µg/ml gentamycin sulphate and grown at 28°C for 48 hours. Plates were stored at room temperature.

2.2.7 *Agrobacterium*-mediated transformation of *B. napus*

Single colonies of *A. tumefaciens* transformed with binary plasmids of interest were inoculated into 5 ml of LB containing 100µg/ml kanamycin and 100µg/ml gentamycin sulphate and shaken at 28°C. Five hundred microliters of 48-hour cultures of *A. tumefaciens* transformants were spun, resuspended in 500µl of AB broth, respun, and resuspended in 4.5ml of AB broth. AB broth is minimal media consisting of: 0.5% glucose; 1mM $MgSO_4$; 100µM $CaCl_2$; 9µM $FeSO_4$-$7H_2O$; 34.5 mM K_2HPO_4; 16.67 mM NaH_2PO_4-H_2O; 37 mM NH_4Cl; 4.02 mM KCl. The bacterial suspension was used to infect sterile, 5-day-old *B. napus* cotyledonary petioles. Transformation of *B. napus* petioles was as outlined in Moloney *et al.* (1989). Briefly, *B. napus* cv. Westar seeds were surface sterilized for 30 minutes (with shaking) in 20% commercial bleach diluted in distilled water. These seeds were washed 3-4 times with sterile water, and germinated, 15-20 seeds

per plate, on BN1 (Murashige and Skoog (MS) (1962) media (pH 5.8) containing 0.7% phytagar and 3% sucrose). This, as with all tissue and plants grown/ cultured/germinated, was performed at room temperature with 16h light/ 8h dark photoperiod. Five days after germination, the petioles were cut 1-2 mm below the cotyledons, dipped into the *A. tumefaciens* EHA 101 suspension (described above) and stood, cotyledon side up, onto BN2 (BN1 containing 4.5mg/L benzyladenine (BA), pH 5.8). Three days after inoculation, the tissue was transferred to BN4 (BN 2 containing 20mg/L kanamycin plus 300mg/L timentin or 500mg/L carbenicillin, pH 5.8, to eliminate *Agrobacterium).* After about 2-3 weeks, transformed (green) callus formed at the base of the cut petiole, and, later, discernible shoots arose. Shoots were excised from the callus and placed in shoot elongation media (BN4 containing no BA, pH 5.8) in Magenta jars. Once the shoots had reached approximately 2 cm in height, they were transferred onto rooting media (elongation media containing 2mg/L indole butyric acid, pH 5.8) in Magenta jars. Once roots had formed, the plantlets were transferred to potting soil (2 parts peat moss, 1 part vermiculite, 1 part Terragreen or Perlite), placed in a misting chamber for 1-3 weeks (relative humidity 75%). Once fully established, plants were transferred to a greenhouse.

2.2.8 Transformation of *Nicotiana tabacum* cv Wisconsin

Tobacco seeds were surface sterilized as described above (section 2.2.7) and germinated in sterile Magenta jars containing MS media, pH 5.8 containing 0.7% phytagar and 3% sucrose (pH 5.8). Young tobacco leaf disks (1 cm in diameter) were incubated in a suspension of *A. tumefaciens* EHA-101 harbouring pCGN-OBHIRT for 5-10 min. The leaf disks were blotted onto sterile Whatman paper and placed, abaxial side down, on sterile MS media containing 3% sucrose, 2mg/L 2,4-D (pH 5.8) in bacterial culture plates (100mm x 15mm). After 2 days, the disks were transferred to selection regeneration media

(MS media (pH 5.8) containing 3% sucrose, 2mg/L BA, 0.1 mg/L NAA, 0.7% phytagar, 500 mg/L carbenicillin, 50 mg/L kanamycin) in plant tissue culture dishes. Once shoots formed, they were excised at their base and transferred to rooting media (MS media (pH 5.8) containing 3% sucrose, 0.1 mg/L NAA, 0.7% phytagar, 500 mg/L carbenicillin, 50 mg/L kanamycin) in Magenta jars. Rooted plantlets were transferred to soil and grown as described above (section 2.2.7).

2.2.9 Neomycin phosphotransferase enzyme assays

In addition to the inserted gene of interest (ie. oleosin-hirudin), a gene encoding neomycin phosphotransferase (NPTII), which confers kanamycin-resistance lies between the right and left borders of the pCGN 1559 binary vector. Integration of this latter gene (and therefore the gene of interest) into the plant genome was determined by NPT II assays. This was performed using a modified version of the method described by Radke *et al.* (1988). Approximately 100 mg of young leaf tissue was ground in 100µl of extraction buffer (50mM Na_2HPO_4, (pH 7.0) 70% β-mercaptoethanol, 10mM Na_2EDTA (pH 8.0), 0.1% Sarcosyl, 0.1% Triton X-100). The resulting extract was centrifuged at 4°C for 20 minutes (14,000 g). Twenty microliters of extract was added to each of two tubes containing 10µl of either reaction buffer (67 mM Tris-maleate (pH 7.1), 42mM $MgCl_2$, 400mM NH_4Cl, 1.7mM dithiothreitol) (negative control), or reaction buffer containing 0.4mg/ml kanamycin (test). Ten microliters of labeled ATP solution (500µl of reaction buffer, 7.9µl of 100mM ATP in reaction buffer, 2µl [γ-^{32}P] ATP [specific activity 7.5-12µCi/µl]) was added to the 30µl of negative control or test samples. This 40µl mixture was incubated 45-60 minutes at 37°C, then spotted onto Whatman cellulose phosphate P81

paper and allowed to dry lightly. The cellulose phosphate paper was washed once in water for 5 minutes, then twice (if necessary) in 10mM Na_2HPO_4, 1% SDS, for 15 minutes at 37°C, until no signal remained in the negative control.

2.2.10 Genomic DNA and RNA isolation

Genomic DNA was extracted from young leaf tissue using the protocol described by Dellaporta *et al.* (1983). Total RNA was extracted from leaf, stem, bud or embryonic tissue using the protocol of Verwoerd *et al.*, (1989) or Parcy *et al.*, (1994). Quantification of DNA or RNA extracts was achieved through spectrophotometric measurements at OD_{260nm}.

2.2.11 Radiolabelling of DNA probe for hybridizations

Oligolabelling buffer (OLB) was prepared as described by Feinberg and Vogelstein (1984). Between 10-50ng of template DNA plus 2µl each of the appropriate primer(s) (10µM) was added to a final volume of 31µl in H_2O. The mixture was boiled for 10 minutes to denature the DNA, then cooled on ice. Once cool, 4µl bovine serum albumin (10mg/ml), 10µl OLB, 2µl Klenow fragment of DNA polymerase I, and 3µl [α-^{32}P] dCTP (specific activity 7.5-12µCi/µl) was added to the template DNA/primer mixture. The mixture was incubated at room temperature for 2-8 hours. The resulting double-stranded probe was purified by passing through a G-50 Sephadex column to remove unincorporated dCTP. Radioactive incorporation was determined using a scintillation counter. Seventeen thousand-five hundred becquerels of DNA probe was added per milliliter of hybridization solution (section 2.2.12).

2.2.12 Southern and Northern blotting

Twenty to 50μg of genomic DNA was digested using restriction enzymes indicated in results section. Digested DNA was electrophoresed through 0.8% agarose in 1X TBE buffer overnight at 30V. Prior to electrophoresis, 20-50μg of total RNA was prepared as detailed in the handbook "Membrane transfer and detection methods" (Amersham, 1987). RNA was subjected to electrophoresis through 1% agarose in 1X MOPS buffer overnight at 30V. After electrophoresis, DNA or RNA was transferred from gels onto Hybond N or Hybond N^+ via capillary action according to manufacturer's (Amersham) instructions. RNA and DNA blot hybridization was performed as described in the handbook "Membrane transfer and detection methods" (Amersham, 1987). Blot hybridizations were performed overnight in 5-50 ml of buffer in either sealed bags or in a hybridization oven (VWR Scientific Mini Hybridization Oven 2700) at 42°C (northern blots) or 65°C (southern blots). Buffer used for DNA blot hybridization consisted of (final concentration): 0.5% SDS; 6X SSC (0.9 M NaCl, 90mM sodium citrate); 5X Denhardt's solution (0.1% (w/v) BSA, 0.1% (w/v) Ficoll, 0.1% (w/v) polyvinyl pyrollidone). Buffer used for RNA blot hydridization consisted of (final concentration): 5X SSPE (0.9 M NaCl, 50mM NaH_2PO_4-H_2O (pH 7.7), 0.5mM Na_2EDTA); 35% (v/v) formamide; 5X Denhardt's solution. Kodak Scientific Imaging Film X-OMAT™ AR was used in autoradiography.

2.2.13 Isolation and extraction of seed protein fractions

When only total seed protein extracts were required, dry seeds were ground with a pestle and mortar in 50mM Tris-HCl (pH 7.5). When total seed protein was to be fractionated into sedimenting, water-soluble, and the oil body phases (Fig. 1.3), extraction

buffer containing sucrose was used. For those experiments, dry (mature) seeds were ground with a pestle and mortar in extraction buffer (0.6 M sucrose, 10 mM KCl, 1 mM $MgCl_2$, 0.15 M tricine, pH 7.5). Buffer volumes used varied according to the amount of seed ground and the purpose of the experiment (see results section). Generally, 0.5-1.5g of seed were ground in 10ml of extraction buffer. This total seed protein was fractionated into sedimenting, water-soluble, and the oil body phases by overlaying with approximately 0.5 volumes of flotation buffer (extraction buffer containing 0.4 M sucrose instead of 0.6 M) and centrifuging at 15,000g in Corex tubes for 20 minutes to 1 hour. Insoluble seed proteins and oil bodies were isolated, resuspended and centrifuged (washed), as above, three times. The soluble seed proteins were centrifuged three more times to remove remaining insoluble fractions. For the fractionation studies, the volumes of total seed protein and fractionated seed proteins were equalized.

For all experiments, protein concentrations were estimated by the Bio-Rad protein assay using bovine serum albumin (BSA) as a standard.

2.2.14 Purification and analysis of recombinant hirudin

Oil bodies from plants transformed with pCGN-OBHIRT (oleosin-hirudin variant 2 fusion gene) were isolated as described in section 2.2.13. After cleavage with Factor Xa, the unternatant fraction was loaded onto an FPLC Mono Q HR 5/5 anion exchange column (Pharmacia) pre-equilibrated with buffer A (20 mM N-methyl piperazine, pH 4.7) and the column was washed with this buffer until the A_{280} returned to baseline. Proteins were eluted with 20 ml of a 0-35% gradient of buffer B (buffer A + 1.0 M NaCl) and collected in 1 ml fractions. Fractions exhibiting anti-thrombin activity were pooled and desalted on PD-10 columns (Pharmacia) equilibrated with water.

Pooled fractions from anion exchange chromatography were loaded onto a Vydac C_{18} TP1022 semipreparative reverse-phase column (Separations Group) equilibrated with 0.1% trifluoroacetic acid. The column was washed with 7% acetonitrile, 0.1% trifluoroacetic acid and developed with a 7-30% gradient of acetonitrile, 0.1% trifluoroacetic acid. Fractions exhibiting anti-thrombin activity were dried on a Speed-Vac and separated on 16.5% tricine-SDS polyacrylamide gels as described below (section 2.2.15).

2.2.15 Protein gel electrophoresis

SDS-polyacrylamide gel electrophoresis (SDS PAGE) of protein extracts were performed based on the methods of Laemmli (1970) or Schagger and Von Jagow (1987). Extracts containing "free" hirudin (hirudin as a non-fusion protein) were separated by SDS-PAGE using a modified version of Schagger and Von Jagow (1987) in order to maximize retention of hirudin in the gel during subsequent staining or electroblotting. This gel system consists of a 16.5% separating gel, a 10% spacer gel, and a 4% stacking gel all made using a 29:1 acrylamide:bis ratio. The contents of each gel layer are as follows:

separating gel: 16.5% acrylamide:bis (29:1); 1M Tris-HCl, pH 8.45; 0.1% SDS; 13.3% glycerol; 0.05% ammonium persulfate; 0.05% TEMED

spacer gel: 10% acrylamide:bis (29:1); 1M Tris-HCl, pH 8.45; 0.1% SDS; 0.05% ammonium persulfate; 0.05% TEMED

stacking gel: 4% acrylamide:bis (29:1); 0.74 M Tris-HCl, pH 8.45; 0.074% SDS; 0.08% ammonium persulfate; 0.08% TEMED

Proteins separated in Schagger and Von Jagow gels were fixed in 50% trichloroacetic acid (TCA) prior to staining or electroblotting. The electrophoretic separation of proteins containing no "free" hirudin was performed using the Laemli method according

to the Bio-Rad manufacturers protocol. Minigels (6x8cm) were run at 20mA/ gel (constant current). Larger gels (14x14 cm) were run at 30-45 mA/ gel (constant current). Low molecular weight coloured markers and broad-range prestained markers used for protein gel electrophoresis were obtained from Sigma (St. Louis, MO., USA) and Bio-Rad Laboratories (Hercules, Ca), respectively.

2.2.16 Western blotting

Protein samples containing no "free" hirudin were transferred from polyacrylamide gels onto PVDF membranes in transfer buffer [25 mM Tris-HCl (pH 8.3), 192 mM glycine, 20% methanol] at 50-70V (constant voltage) for 2 hours- overnight. Immobilon-P PVDF membranes (Millipore, Mississauga, Ontario) have a pore size of ≅ 0.45μm and are suitable for blotting of proteins >10 kDa. After fixation in TCA (2.2.15), Schagger and Von Jagow gels were equilibrated in transfer buffer containing 0.1% SDS and the proteins electroblotted onto Immobilon PVDF-P^{SQ} (Millipore). This membrane has a pore size of 0.1μm and an enhanced protein binding capacity for smaller peptides. Transfer onto Immobilon-P^{SQ} was performed in transfer buffer (above) at 300 mA (constant current) for 2-3 hours. All immunoblots were blocked with 10% skimmed milk in TBS buffer (20mM Tris-HCl, pH 7.5, 0.9% NaCl) for 1h to overnight. Blots were rinsed with TBS containing 0.3% Tween, 15 minutes, then incubated with either rabbit anti-19kDa *B. napus* oleosins antibodies or mouse anti-hirudin monoclonal (specified in results) for 3 hours at room temperature. The membranes were rinsed with TBS containing 0.3% Tween, 15 minutes, then incubated with alkaline phosphatase (AP)-conjugated secondary antibodies recognizing rabbit or mouse IgG for 1 hour. Colour development of AP was achieved using Immun-Blot® alkaline phosphatase assay kit (Bio-Rad).

2.2.17 Immunofluorescent Localization of Oleosin-Hirudin Fusion Protein

Anti-hirudin monoclonal antibodies were used to confirm the presence of hirudin on the surface of lipid bodies. Oil body proteins from transformed *B. napus* seeds were washed three times in buffer (section 2.2.13). After the third centrifugation step, the oil bodies were isolated and resuspended in thrombin assay buffer (50 mM Tris-HCl (pH 8.0), 100 mM NaCl, 5 mM $CaCl_2$) and separated into 200µl (27mg/ml) aliquots. In order to reduce non-specific staining, the oil bodies were washed three times in 300µl PBS containing 1% BSA. The preparations were incubated with anti-hirudin monoclonal antibodies (1:20 dilution) for 6 hours at 4°C. The antibodies were diluted in PBS containing 0.1% BSA. After incubation, the oil bodies were washed three times in PBS containing 0.1% BSA. FITC-conjugated anti-mouse secondary antibodies (1:60 dilution) were added and incubated at room temperature for 1 hour. Following this, the oil bodies were washed three times in PBS. Twenty microliters of oil body and 20µl of mounting media were placed on a slide. The mounting media contained a saturated solution of n-propyl gallate, in PBS. Slides were examined using a Leica Aristoplan fluorescence microscope with the standard FITC exciter and barrier filters (Filter system I3, Leica Canada, Toronto). The image was captured on Ilford FP4 film (ASA 400). A fixed exposure time was used to photograph all treatments.

2.2.18 Proteolytic digestion of oleosin-hirudin fusion proteins

Prior to proteolytic cleavage, isolated oil bodies were washed two to three times in thrombin assay buffer (50 mM Tris-HCl (pH 8.0), 100 mM NaCl, and 5 mM $CaCl_2$). Factor Xa (FXa) or clostripain (amounts specified in Results) was added to these oil body preparations, and incubated overnight at room temperature. Digestion of oil body proteins

with clostripain was performed under reducing conditions, with the addition of DTT (2.5 mM DTT, final concentration) to preparations in thrombin assay buffer. After clostripain treatment, 500µl of oil body preparations (14mg/ml) were precipitated overnight at -20°C with 11 volumes of acetone. Precipitated proteins were washed once with cold acetone, air-dried slightly, and resuspended in 500µl of thrombin assay buffer. Renaturation of hirudin occurs spontaneously and quickly when the DTT is removed and the hirudin returned to this slightly alkaline environment (pH 8.0) (Chatrenet and Chang, 1993).

2.2.19 Anti-thrombin assays

Hirudin activity was quantitated through use of chromogenic assay (Chang, 1983). This assay measures the rate of cleavage of a chromogenic substrate (p-tosyl-glycyl-prolyl-arginine-p-nitroanilide) by thrombin. Hirudin occludes the active cleft of thrombin, preventing the cleavage of this substrate. The presence of cleavage products was measured spectrophotometrically at OD_{405nm}. Unless stated otherwise, anti-thrombin assays were performed as follows: thrombin from human plasma was added to assay buffer (50 mM Tris-HCl (pH 8.0), 100 mM NaCl, 5 mM $CaCl_2$) to a concentration of 0.05 to 0.14 units/ml. This reaction buffer was stored on ice until needed. One ml of reaction buffer was added to 1-30µl of buffer (control) or test extracts and allowed to incubate at room temperature (approximately 25°C) for 1-10 minutes as indicated in results. After incubation, 50 pmol of the chromogenic substrate, p-tosyl-glycyl-prolyl-arginine-p-nitroanilide (Sigma, St. Louis, MO., USA) was added to the above mixture. The change in absorbance at OD_{405nm} was monitored over 2 minutes. The average change of absorbance per minute and the standard error of the mean was calculated and converted into units of

thrombin by reference to a thrombin standard curve. Dose-dependent assays were analyzed by a least significant difference (LSD) test at the 1% level of significance.

2.3 PLASMID CONSTRUCTION

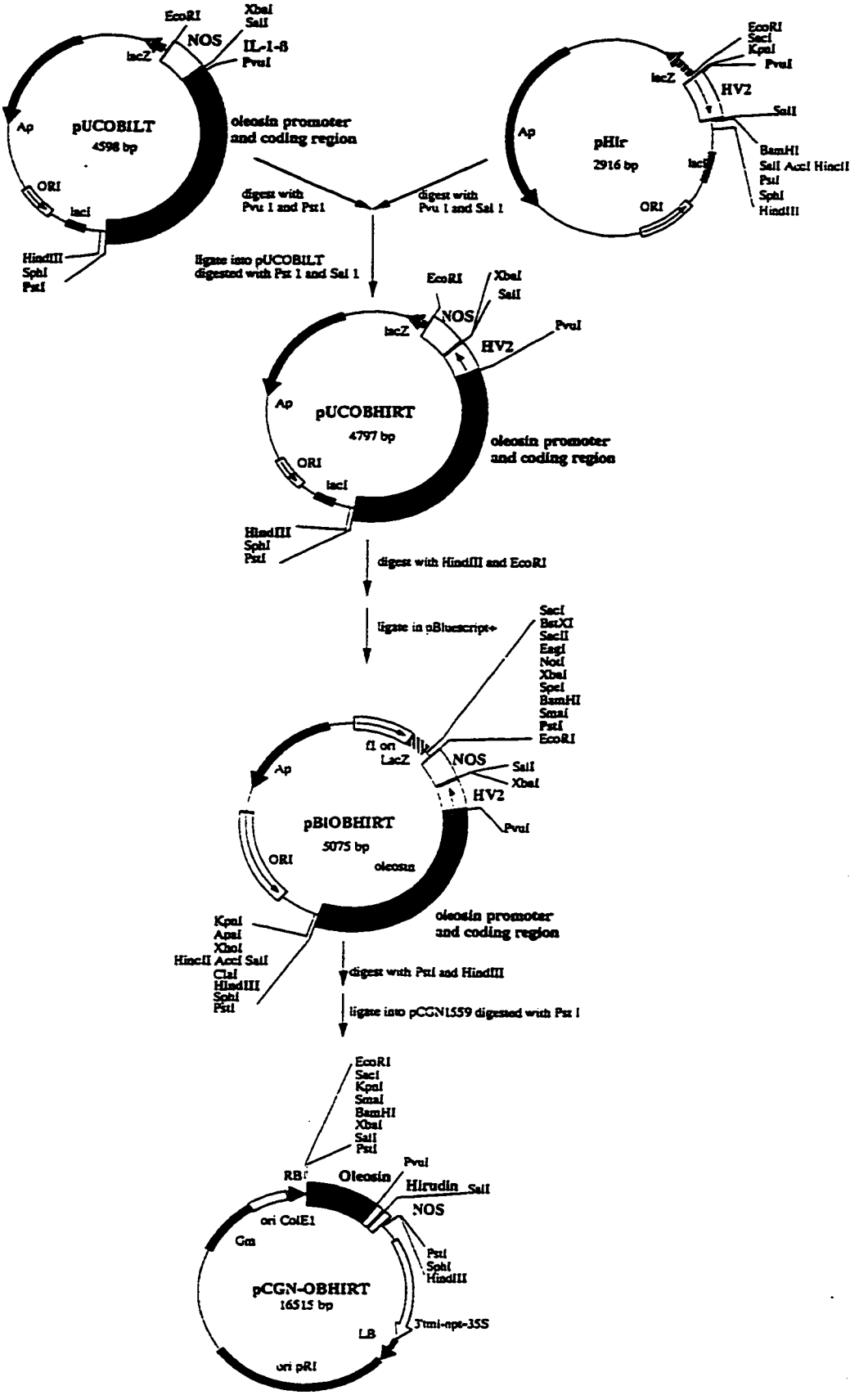
2.3.1 Oleosin-Hirudin Fusions

2.3.1.1 Oleosin-hirudin variant 2 (pCGN-OBHIRT)

The construction of pCGN-OBHIRT is summarized in Fig. 2.1. A synthetic cDNA encoding hirudin variant 2 (HV2) was designed by Gijs van Rooijen (van Rooijen, 1993) from the amino acid sequence (Harvey *et al.*, 1986) but employing *Brassica napus* (Lee and Huang, 1991) and *Arabidopsis* (van Rooijen *et al.*, 1992) codon usage. Four overlapping oligonucleotides were used in PCR amplification to generate a 222bp fragment. This fragment contained a region corresponding to the hirudin coding sequence, a region encoding a proteolytic cleavage site at the 5' end, and restriction sequences at both ends. DNA sequence analysis of the fragment was performed after end-filling and subcloning into the *Sma* 1 site of a pUC 19 plasmid vector. This plasmid was called pHIR.

The lambda genomic clone containing an *Arabidopsis* oleosin promoter and coding sequence (van Rooijen *et al.*, 1992) was used as a template in polymerase chain reactions. The oligonucleotide GVR10 (5' CACTGCAG**GAACTCTCTGGTAAGC** 3'), which is homologous to sequences -838 to -814 (bold type) of the upstream *Arabidopsis* oleosin sequence, but modified to contain a *Pst* 1 site (underlined), and the oligonucleotide GVR11 (5'CCGTCGACTTACTTGTCGTTAGATTCTTCTCCCTGAACTCTCCCTTcgatcgCAG **TAGTGTGCTGGCCACC** 3') which is homologous to the 3' coding region of the *Arabidopsis* oleosin coding region (bold type), but modified to contain a *Sal* 1 (underlined) and a *Pvu* 1 (lowercase) site, were generated. These primers were used in a 30-cycle PCR

Figure 2.1 Construction of pCGN-OBHIRT. The coding region of hirudin variant 2 was fused to the 3' end of a *B. napus* oleosin cDNA. Between these two protein-encoding genes was added a site recognized by the protease, Factor Xa. This construct is terminated by a NOS transcriptional terminator and is driven by a 2.5 kbp *Arabidopsis* oleosin promoter.

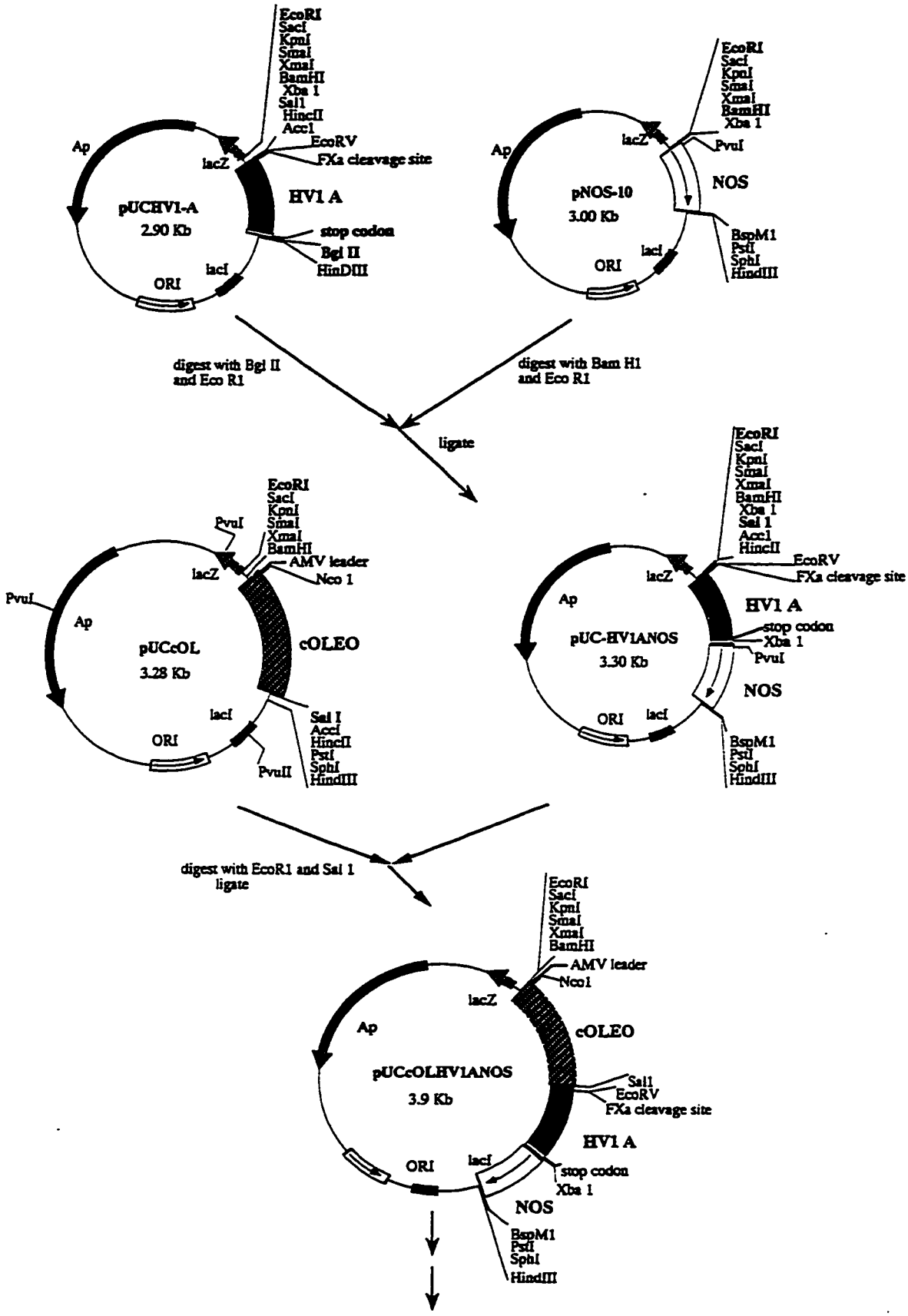

EcoRI
XbaI
SalI
NOS
IL-1-β
PvuI
lacZ
Ap
pUCOBILT
4598 bp
oleosin promoter
and coding region
ORI
lacI
HindIII
SphI
PstI
digest with
Pvu I and Pst I
digest with
Pvu I and Sal I
EcoRI
SacI
KpnI
PvuI
lacZ
HV2
SalI
Ap
pHir
2916 bp
lacI
ORI
BamHI
SalI AccI HincII
PstI
SphI
HindIII
ligate into pUCOBILT
digested with Pst I and Sal I
EcoRI
XbaI
SalI
NOS
lacZ
HV2
PvuI
Ap
pUCOBHIRT
4797 bp
oleosin promoter
and coding region
ORI
lacI
HindIII
SphI
PstI
digest with HindIII and EcoRI
ligate in pBluescript+
SacI
BstXI
SacII
EagI
NotI
XbaI
SpeI
BamHI
SmaI
PstI
EcoRI
f1 ori
LacZ
NOS
SalI
XbaI
Ap
HV2
PvuI
pBIOBHIRT
5075 bp
ORI
oleosin
oleosin promoter
and coding region
KpnI
ApaI
XhoI
HincII AccI SalI
ClaI
HindIII
SphI
PstI
digest with PstI and HindIII
ligate into pCGN1559 digested with Pst I
EcoRI
SacI
KpnI
SmaI
BamHI
XbaI
SalI
PstI
RBS
Oleosin
PvuI
Hirudin
SalI
NOS
ori ColE1
Gm
PstI
SphI
HindIII
pCGN-OBHIRT
16515 bp
LB
3'nnl-npt-35S
ori pRI

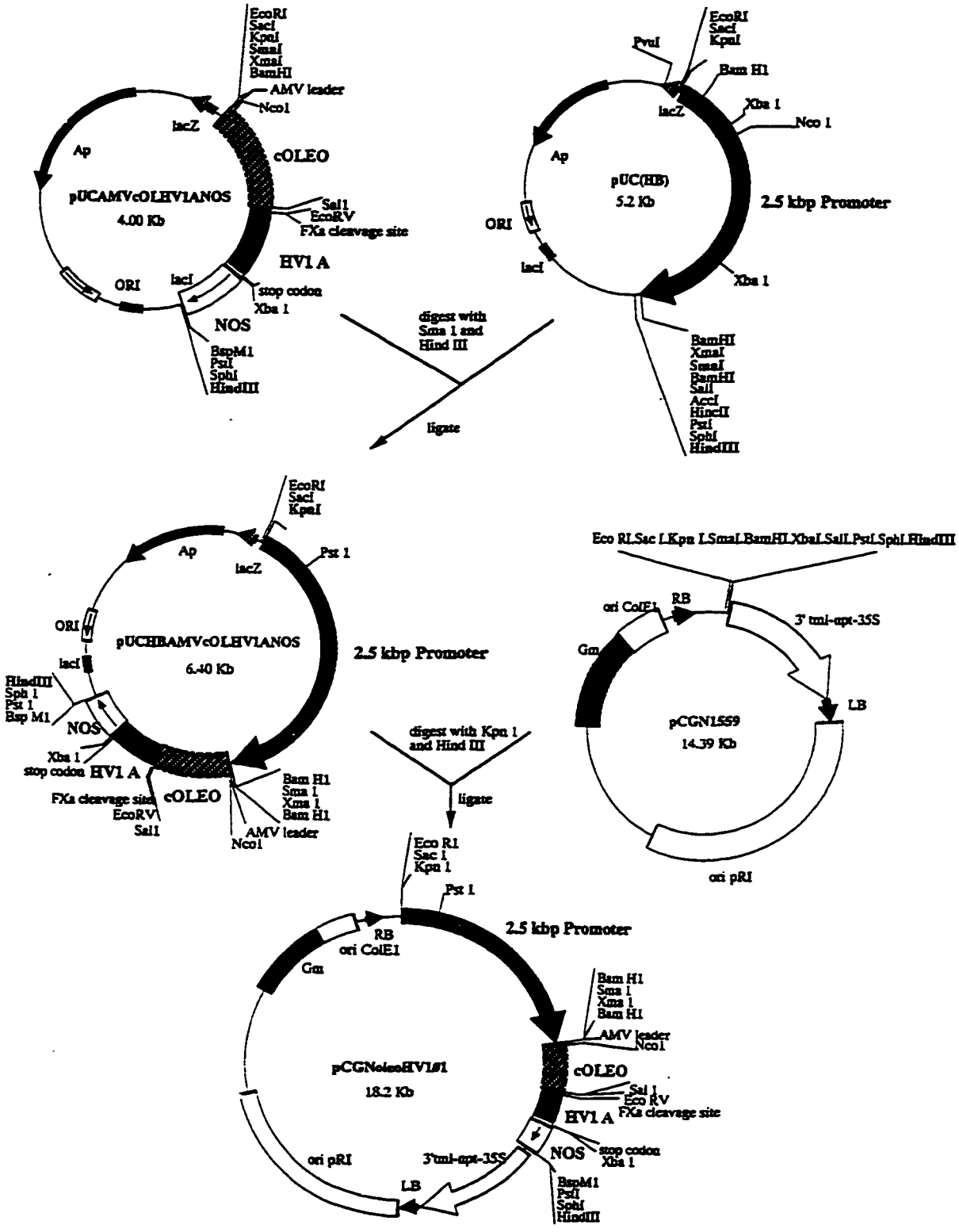
amplification of the lambda clone (annealing temperature of 45°C). The PCR reaction mixture consisted of 16 µl dNTPs (1.25 mM), 10 µl 10X PCR buffer [100 mM Tris-HCl (pH 8.3), 500 mM KCl, 15 mM $MgCl_2$, 0.1% (w/v) gelatin], 5 µl (20 mM) of each primer, 1 µl Taq DNA polymerase (1 unit/µl), 1 µl (1ng/µl template), and 62 µl H_2O. A 1652bp fragment containing a 0.84kb *Arabidopsis* oleosin promoter and 0.8kb coding region was isolated, end-filled, kinased, and cloned into the *Sma* 1 site of pUC19. This plasmid was called pOBIL. Next, pBI121 (Clontech laboratories), containing a nopaline synthetase (nos) terminator sequence flanked by 5' *Sal* 1, 3'*Eco R*1 restriction sites was digested with these enzymes and subcloned into the *Sal* 1/*Eco R*1 sites of pUC19. This plasmid was called pTERM. The nos terminator was isolated from this plasmid by cleavage with *Sal* 1 and *Eco R*1 restriction enzymes and ligated to the 3' end of pOBIL plasmid at the *Sal* 1/*Eco R*1 sites. This plasmid was called pUCOBILT. The hirudin fragment of pHIR was excised with *Pvu* 1 and *Sal* 1. The oleosin promoter and coding region of pUCOBILT was excised with *Pvu* 1 and *Pst* 1. Both of these DNA fragments were simultaneously ligated into pUCOBILT plasmid digested with *Pst* 1 and *Sal* 1. The resulting plasmid was called pUCOBHIRT. This oleosin-hirudin fusion gene of pUCOBHIRT was subcloned into the *Hind* III/ *Eco R*1 site of pBluescript KS+ to generate appropriate restriction sites at the 5' and 3' ends, then subcloned into pCGN 1559 (McBride and Summerfelt, 1990) at the *Pst* 1 site. This plasmid was designated pCGN-OBHIRT.

2.3.1.2 AMV leader-oleosin-hirudin variant 1 #1 (pCGN-oleoHV1#1)

The construction of pCGN-oleoHV1#1 is summarized in Fig. 2.2. Synthetic hirudin variant 1-A (HV1-A) was created by PCR. The primers used (20µM each) were

Figure 2.2 Construction of pCGNoleoHV1#1. The coding region of hirudin variant 1 was fused to the 3' end of a *B. napus* oleosin cDNA. Between these two protein-encoding genes was added a site recognized by the protease, Factor Xa. Immediately upstream to the oleosin cDNA is an untranslated alfalfa mosaic virus (AMV) leader sequence. This construct is terminated by a NOS transcriptional terminator. Transcription is driven by a 2.5 kbp *Arabidopsis* oleosin promoter.


EcoRI
SacI
KpnI
SmaI
XmaI
BamHI
Xba 1
Sal1
HincII
Acc1
EcoRV
FXa cleavage site
Ap
lacZ
pUCHV1-A
2.90 Kb
HV1 A
stop codon
Bgl II
HinDIII
lacI
ORI
pNOS-10
3.00 Kb
NOS
PvuI
BspM1
PstI
SphI
HindIII
digest with Bgl II
and Eco R1
digest with Bam H1
and Eco R1
ligate
AMV leader
Nco 1
pUCcOL
3.28 Kb
cOLEO
Sal I
PvuII
pUC-HV1ANOS
3.30 Kb
digest with EcoR1 and Sal 1
ligate
pUCcOLHV1ANOS
3.9 Kb

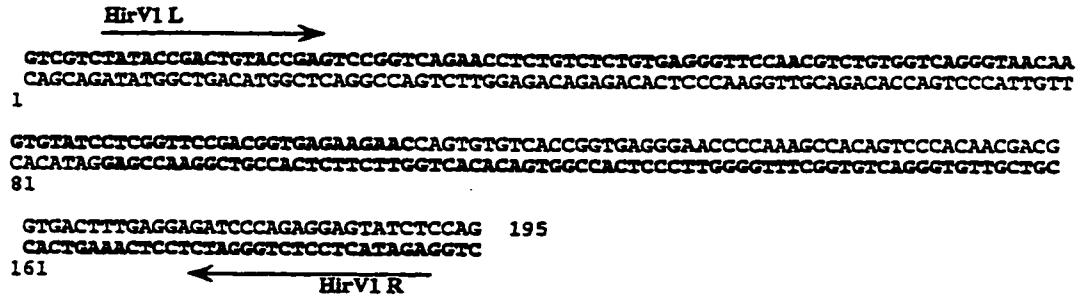

EcoRI
SacI
KpnI
SmaI
XmaI
BamHI
AMV leader
NcoI
lacZ
cOLEO
Ap
pUCAMVcOLHV1ANOS
4.00 Kb
SalI
EcoRV
FXa cleavage site
HV1 A
ORI
lacI
stop codon
Xba I
NOS
BspM1
PstI
SphI
HindIII
digest with
Sma I and
Hind III
ligate
PvuI
EcoRI
SacI
KpnI
Bam H1
Xba I
lacZ
Nco I
Ap
pUC(HB)
5.2 Kb
2.5 kbp Promoter
ORI
lacI
Xba I
BamHI
XmaI
SmaI
BamHI
SalI
AccI
HincII
PstI
SphI
HindIII
EcoRI
SacI
KpnI
Ap
Pst 1
lacZ
ORI
pUCHBAMVcOLHV1ANOS
6.40 Kb
2.5 kbp Promoter
lacI
HindIII
Sph 1
Pst 1
Bsp M1
NOS
Xba 1
stop codon
HV1 A
FXa cleavage site
EcoRV
SalI
cOLEO
Bam H1
Sma 1
Xma 1
Bam H1
AMV leader
NcoI
digest with Kpn 1
and Hind III
ligate
Eco RI Sac I Kpn I SmaI BamHI XbaI SalI PstI SphI HindIII
ori ColE1
RB
3' tml-apt-35S
Gm
LB
pCGN1559
14.39 Kb
ori pRI
Eco RI
Sac I
Kpn I
Pst 1
RB
ori ColE1
2.5 kbp Promoter
Gm
Bam H1
Sma 1
Xma 1
Bam H1
AMV leader
Nco1
pCGNoleoHV1#1
18.2 Kb
cOLEO
Sal 1
Eco RV
FXa cleavage site
HV1 A
NOS
stop codon
Xba 1
ori pRI
3'tml-apt-35S
LB
BspM1
PstI
SphI
HindIII

HirV1L (1µl), HirV1R (1µl), HV1-5' (5µl), and HV1-3'a (5µl). The sequence of these primers is listed in Table 2.1. HirV1L and HirV1R have a 24 bp overlap at their 3' ends, as shown in Fig. 2.3. The complete synthetic HV1 sequence is identical at the amino acid level, to published HV1 (Dodt *et al.*, 1984) but has oleosin codon usage. HirV1L and HirV1R act as a template for the other hirudin primers, HV1-5', HV1-3'a, HV1-3'b, and HV1-3'c, as shown in Figure 2.4. HV1-5' (5' GCAGCAGCAGCAGTCGACGAGCA GGAGCAAGAGGATatcgagggtagaGTCGTCTATACCGACTGT 3') is identical to the 5' end of the top strand of HV1 (bold), includes a Factor Xa recognition site (lowercase) and sites for *Eco* RV (underlined) and *Sal* I (double underlined). HV1-3'a (5'AGCAGCAGCAAGCTTAGATCTttaCTGGAGATACT CCTCTGG 3') has sequences identical to the bottom strand, 3' end of HV1 (bold). Included was a site for translational termination (stop codon; taa) (lowercase), and for restriction endonucleases *Bgl* II (underlined) and *Hind* III (double underlined). The annealing temperature for the polymerase chain reaction involving these four oligonucleotides was 50°C using Vent™ polymerase. The PCR product was genecleaned and digested with *Sal* 1, then *Hind* III, and ligated into pUC 19 also digested with these endonucleases. The resulting plasmid was called pUCHV1-A.

To facilitate expression in *E. coli* (if necessary), all oleosin-hirudin variant 1 fusion genes were constructed using a *B. napus* oleosin cDNA, rather than the genomic clone, which is interrupted by an intron. *B. napus* oleosin cDNA was used as a template in a PCR reaction using the primers BamAMVcol and col 3'Sal. BamAMVcol (5'CGCCGCGGATCCgttttttatttttaattttctttcaaatacttccaccATGGCGGATACAGCTAGA 3') is complementary to the bottom strand of the 5' end of an oleosin cDNA (bold). It contains an AMV leader sequence (lower case) which ends in a start codon, and begins

Figure 2.3 a) Generation of synthetic hirudin variant 1 DNA sequence using the overlapping, top and bottom strand primers, HirV1 L and HirV1 R (respectively). The sequence of the primers (shown in bold) overlap by 24 bases. b) Protein sequence of hirudin variant 1 generated by HirV1 L and HirV1 R via PCR. The amino acid composition is identical to that described in *Hirudo medicinalis*.

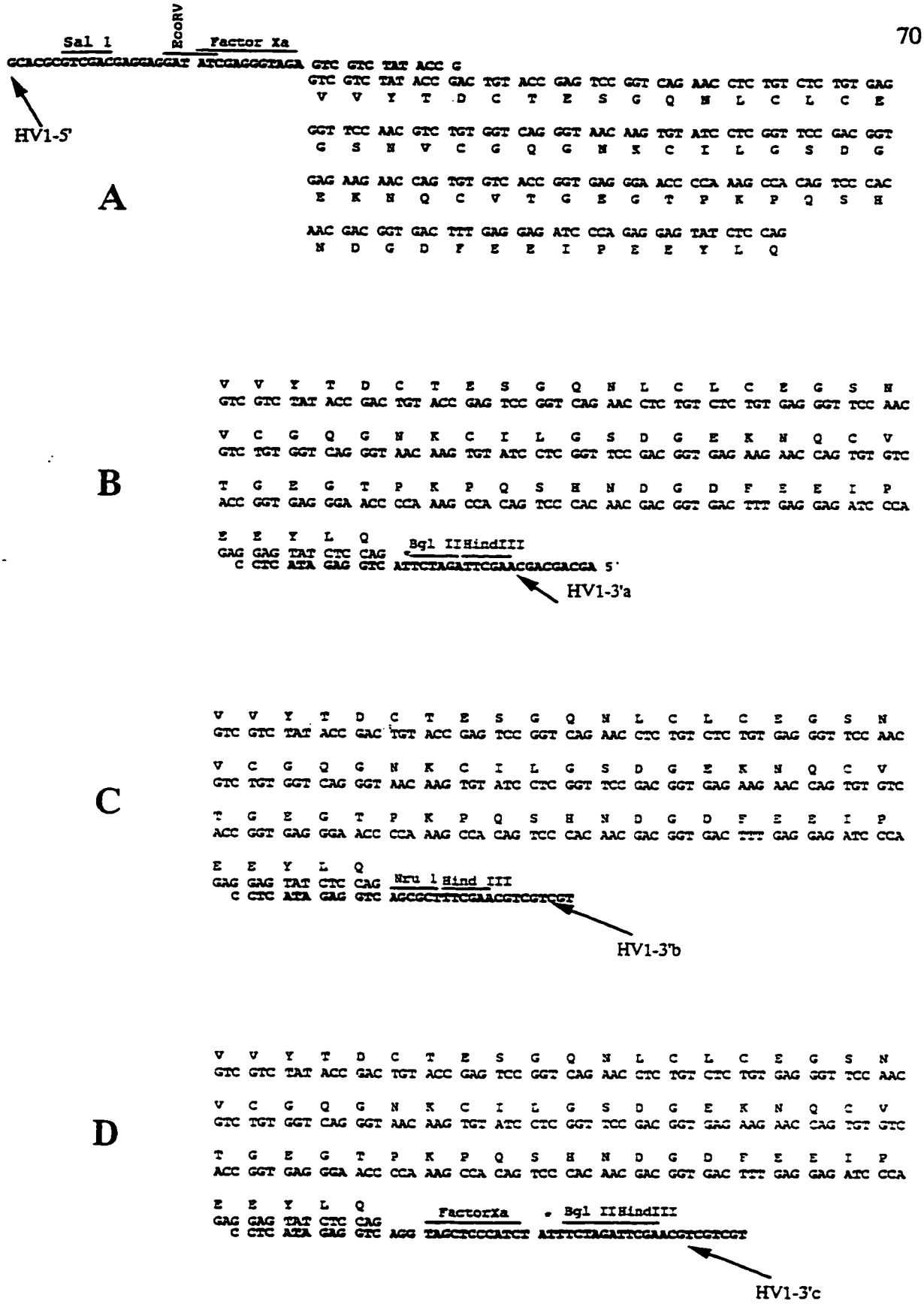
A. Hirudin variant 1 DNA sequence (oleosin codon usage)



B. Hirudin variant 1 protein sequence

VVYTDCTESGQNLCLCEGSNVCGQGNKCILGSDGEKNQCVTGEGTPKPQSHNDGDFEEIPEEYLQ

Figure 2.4 Primers (shown in bold) which were used to generate synthetic hirudin variant 1 DNA having different 5' and 3' ends. a) The primer, HV1-5' includes sequences encoding a Factor Xa and restriction endonuclease recognition sites. b) HV1-3'a contains a stop codon (*) and restriction endonuclease recognition sites. c) HV1-3'b contains restriction endonuclease recognition sites but no stop codon. d) HV1-3'c possesses sequences encoding Factor Xa proteolytic cleavage site, a stop codon (*) and restriction endonuclease recognition sites.


Sal 1
EcoRV
Factor Xa
HV1-5'
A
B
Bgl II HindIII
HV1-3'a
C
Nru 1 Hind III
HV1-3'b
D
FactorXa
Bgl II HindIII
HV1-3'c

with two restriction sites, *Bam* HI (underlined) and *Nco* I (double underlined). col 3'Sal (5' GTCGTCGTCGTCGACGGTAGTGTGCTGGGTTCCACG 3') is complementary to the 3' end of the top strand of a *B. napus* oleosin cDNA (bold) and contains a *Sal* I recognition sequence (underlined). The annealing temperature was 45°C. Twenty microliters of the PCR reaction was digested with *Sal* I then *Bam* HI then ligated into pUC19 which was also digested with these enzymes. The resulting plasmid was called pUCcOL.

After digestion with *Bgl* II and *Eco* RI, the HV1 insert of pUCHV1-A was isolated in LMP agarose and ligated into pNOS-10 digested with *Bam* HI and *Eco* RI. The resulting plasmid, pUC-HV1ANOS, was digested with *Eco* RI and *Sal* I, and into that site, inserted a *B. napus* oleosin cDNA derived from digestion, and insert isolation, of pUCcOL digested with *Eco* RI and *Sal* I. The resulting plasmid, pUCAMVcOLHV1ANOS was digested with *Sma* I and *Hind* III. The insert was isolated in LMP agarose and inserted behind the 2.5 kbp *Arabidopsis* oleosin promoter (pUC(HB)), which had also been digested with *Sma* I and *Hind* III. pUC(HB) was constructed by Dr. A. Plant (Plant *et al.*, 1994). The resulting plasmid was called pUCHBAMVcOLHV1ANOS. pUCHBAMVcOLHV1ANOS was digested with *Kpn* I and *Hind*III, the enzymes heat denatured at 85°C, 30 minutes, and the entire digest used in a ligation with pCGN 1559 digested with *Kpn* I and *Hind*III. The resulting plasmid, pCGN-oleoHV1#1, was sequence verified using the primers GVR 25, GVR 26, oleo(+) 133, oleo(+) 364, reverse, and BamAMVcOL, and analyzed by restriction digest before being mobilized into *Agrobacterium tumefaciens* EHA 101.

2.3.1.3 Oleosin-hirudin variant 1 #2 (pCGN-oleoHV1#2)

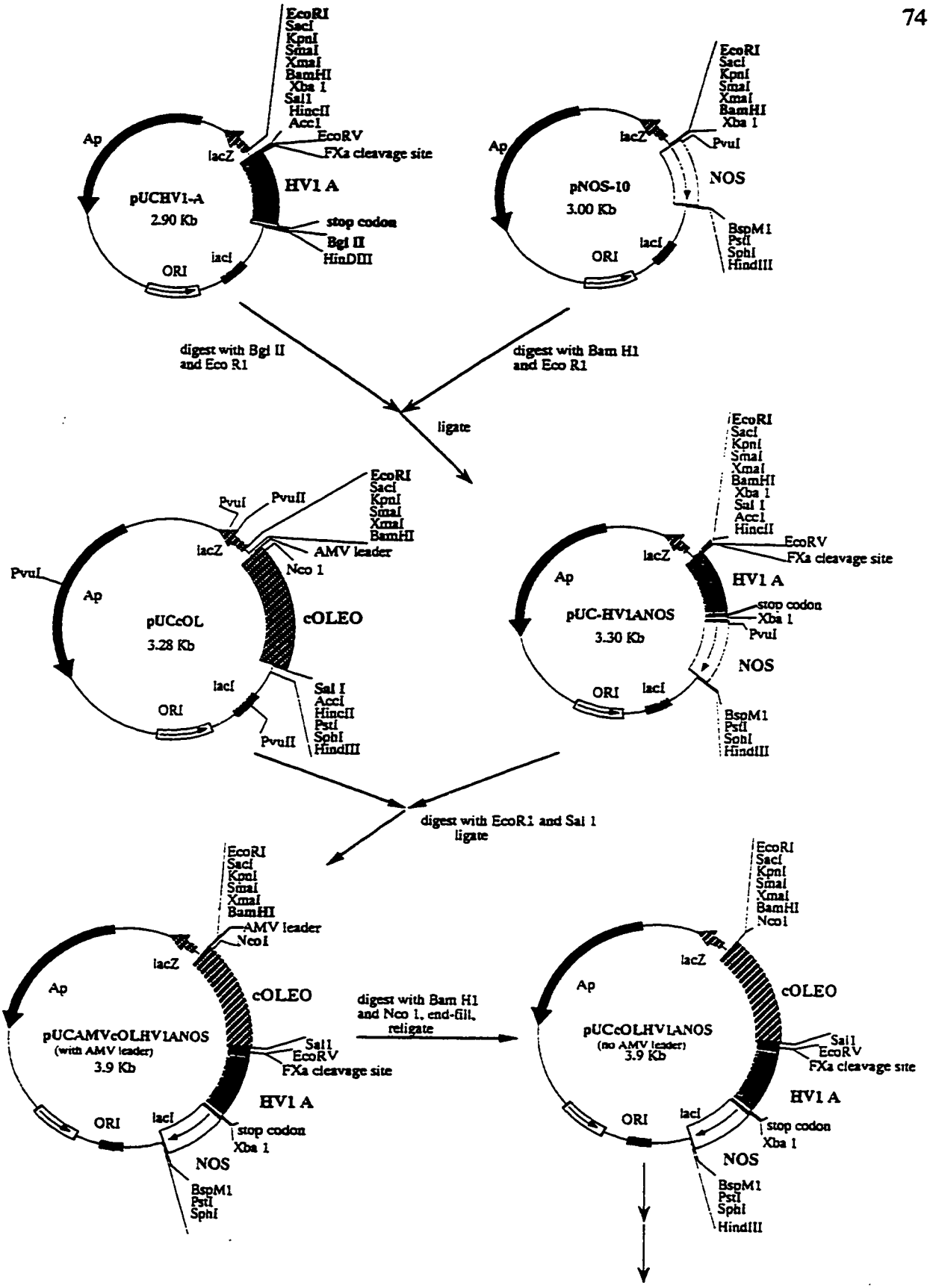
In order to determine the effect of the AMV translational enhancer on overall protein accumulation, the construct pCGN-oleoHV1#2 was created. This plasmid is exactly the

same as pCGN-oleoHV1#1, except the AMV leader sequence was removed. The construction of this plasmid is summarized in Fig. 2.5. The AMV sequence of pUCAMVcOLHV1ANOS (constructed as described above) lies between the restriction sites *Bam* HI and *Nco* I. This plasmid was digested with these enzymes, end-filled with Klenow fragment of DNA polymerase I, and religated. The removal of the AMV leader sequence was verified by sequence analysis using universal primer. The resulting insert of this plasmid was inserted behind the 2.5kbp oleosin promoter (pUC(HB)) and then into pCGN 1559 exactly as described for the construction of pCGN-oleoHV1#1 (see sections 2.3.1.2 and Fig. 2.2). The sequence of the this final construct, pCGN-oleoHV1#2 was verified using the primers used to verify pCGN-oleoHV1#1 (section 2.3.1.2).

2.3.2 Oleosin-hirudin variant 1 concatamers

Constructs containing oleosin-hirudin fusion genes were constructed in which HV1 was present in repeating, concatameric sequences behind the oleosin coding region. Three oleosin-hirudin concatamer constructs were created, having either 1, 2, or 4 hirudin-encoding sequences following oleosin. The final constructs are summarized in Fig. 2.6. After each hirudin-encoding sequence, and between the oleosin-hirudin sequence, a sequence specifying a FXa cleavage site was added. Two types of synthetic hirudin-encoding DNA were constructed for the production of oleosin-hirudin concatamers: HV1-B and HV1-C. HV1-C has sequences encoding a FXa cleavage site at its 5' and 3' ends, plus a stop codon at its 3' end. HV1-B has sequences encoding a FXa cleavage site at its 5' but not 3' end, and contains no stop codon, allowing for translational read-through of HV1 sequences which follow it.

Figure 2.5 Construction of pCGNoleoHV1#2. This gene fusion construct is identical to pCGNoleoHV1#1, except that the 5' untranslated alfalfa mosaic virus (AMV) leader sequence is not present. This was achieved by digestion of pUCcOLHV1A NOS with the endonucleases Bam H1 and Nco 1, followed by end-filling and religation (pUCcOLHV1A NOS).


EcoRI
SacI
KpnI
SmaI
XmaI
BamHI
Xba 1
Sal1
HincII
Acc1
EcoRV
FXa cleavage site
Ap
lacZ
HV1 A
pUCHV1-A
2.90 Kb
stop codon
Bgl II
HinDIII
lacI
ORI
pNOS-10
3.00 Kb
NOS
PvuI
BspM1
PstI
SphI
HindIII
digest with Bgl II
and Eco R1
digest with Bam H1
and Eco R1
ligate
PvuII
AMV leader
Nco 1
pUCcOL
3.28 Kb
cOLEO
Sal I
pUC-HV1ANOS
3.30 Kb
digest with EcoR1 and Sal 1
ligate
pUCAMVcOLHV1ANOS
(with AMV leader)
3.9 Kb
pUCcOLHV1ANOS
(no AMV leader)
3.9 Kb
digest with Bam H1
and Nco 1, end-fill,
religate

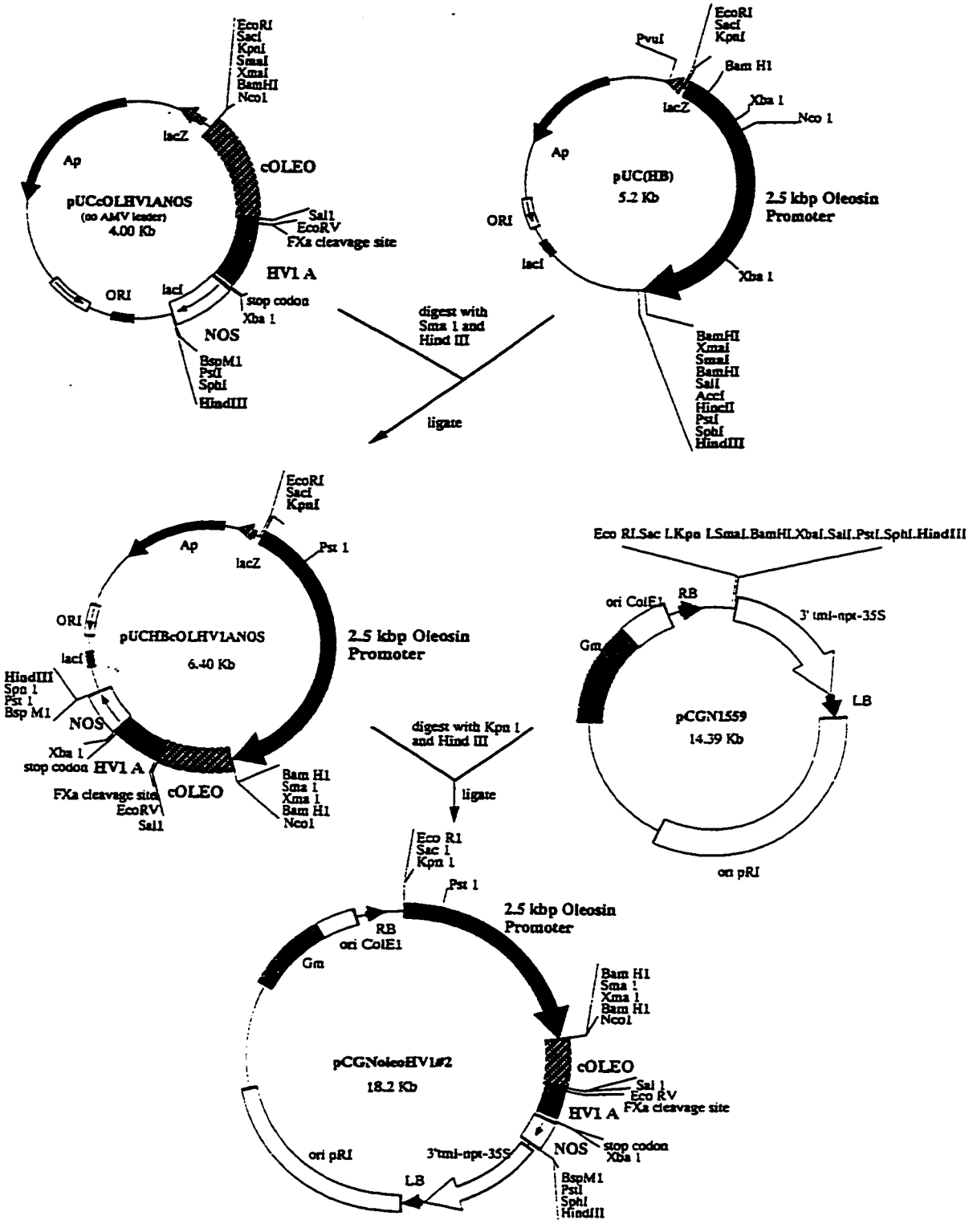
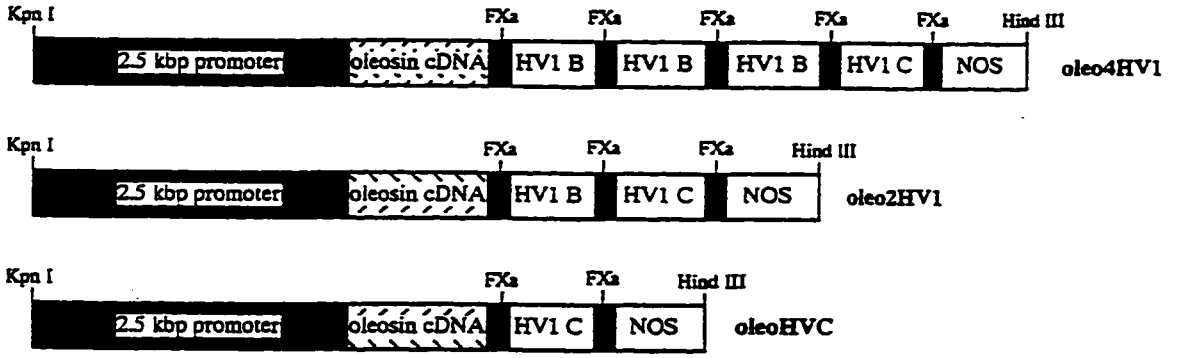
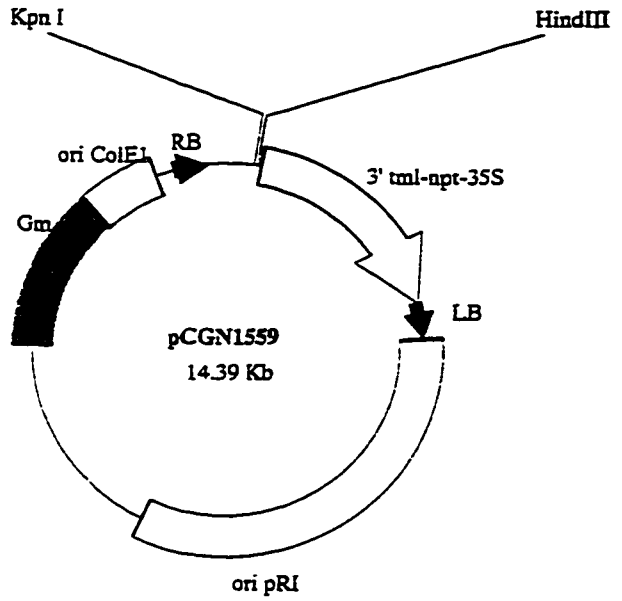

EcoRI
SacI
KpnI
SmaI
XmaI
BamHI
NcoI
lacZ
Ap
cOLEO
pUCcOLHV1ANOS
(no AMV leader)
4.00 Kb
SalI
EcoRV
FXa cleavage site
HV1 A
ORI
lacI
stop codon
Xba I
NOS
BspM1
PstI
SphI
HindIII
digest with
Sma I and
Hind III
ligate
PvuI
Bam H1
Xba I
Nco I
pUC(HB)
5.2 Kb
2.5 kbp Oleosin
Promoter
BamHI
XmaI
SmaI
BamHI
SalI
AccI
HincII
PstI
SphI
HindIII
Psx 1
pUCHBcOLHV1ANOS
6.40 Kb
HindIII
Spn 1
Pst 1
Bsp M1
Xba I
stop codon
FXa cleavage site
EcoRV
SalI
Bam H1
Sma 1
Xma 1
Bam H1
Nco1
digest with Kpn I
and Hind III
ligate
Eco RI Sac I Kpn I SmaI BamHI XbaI SalI PstI SphI HindIII
ori ColE1
RB
3' tml-npt-35S
Gm
pCGN1559
14.39 Kb
LB
ori pRI
Eco R1
Sac I
Kpn I
Pst 1
2.5 kbp Oleosin
Promoter
RB
ori ColE1
Gm
Bam H1
Sma I
Xma 1
Bam H1
Nco1
pCGNoleoHV1#2
18.2 Kb
cOLEO
Sal 1
Eco RV
FXa cleavage site
HV1 A
NOS
stop codon
Xba 1
BspM1
PstI
SphI
HindIII
ori pRI
3'tml-npt-35S
LB

Figure 2.6 Overview of constructs containing oleosin-hirudin variant 1 concatamer fusion genes in pCGN 1559. Construction of these gene fusions is diagrammed in Figs. 2.7-2.10.

Abbreviations: HV1 (x), hirudin variant 1; NOS, nopaline synthase terminator; FXa, Factor Xa


Kpn I
FXa
FXa
FXa
FXa
FXa
Hind III
2.5 kbp promoter
oleosin cDNA
HV1 B
HV1 B
HV1 B
HV1 C
NOS
oleo4HV1
Kpn I
FXa
FXa
FXa
Hind III
2.5 kbp promoter
oleosin cDNA
HV1 B
HV1 C
NOS
oleo2HV1
Kpn I
FXa
FXa
Hind III
2.5 kbp promoter
oleosin cDNA
HV1 C
NOS
oleoHVC


Kpn I
HindIII
ori ColE1
RB
3' tml-npt-35S
Gm
LB
pCGN1559
14.39 Kb
ori pRI

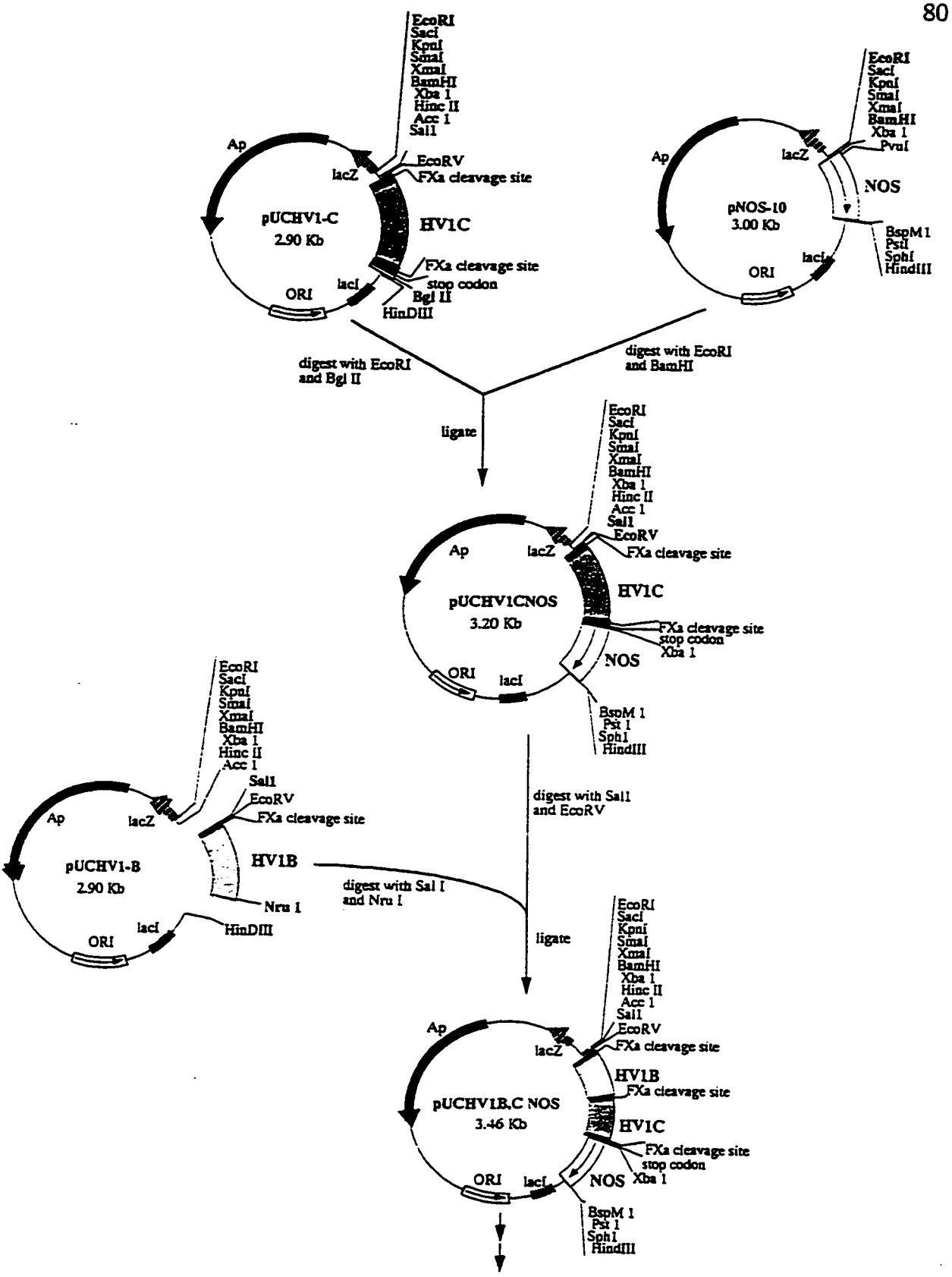
2.3.2.1 pUC-HV1 concatamers

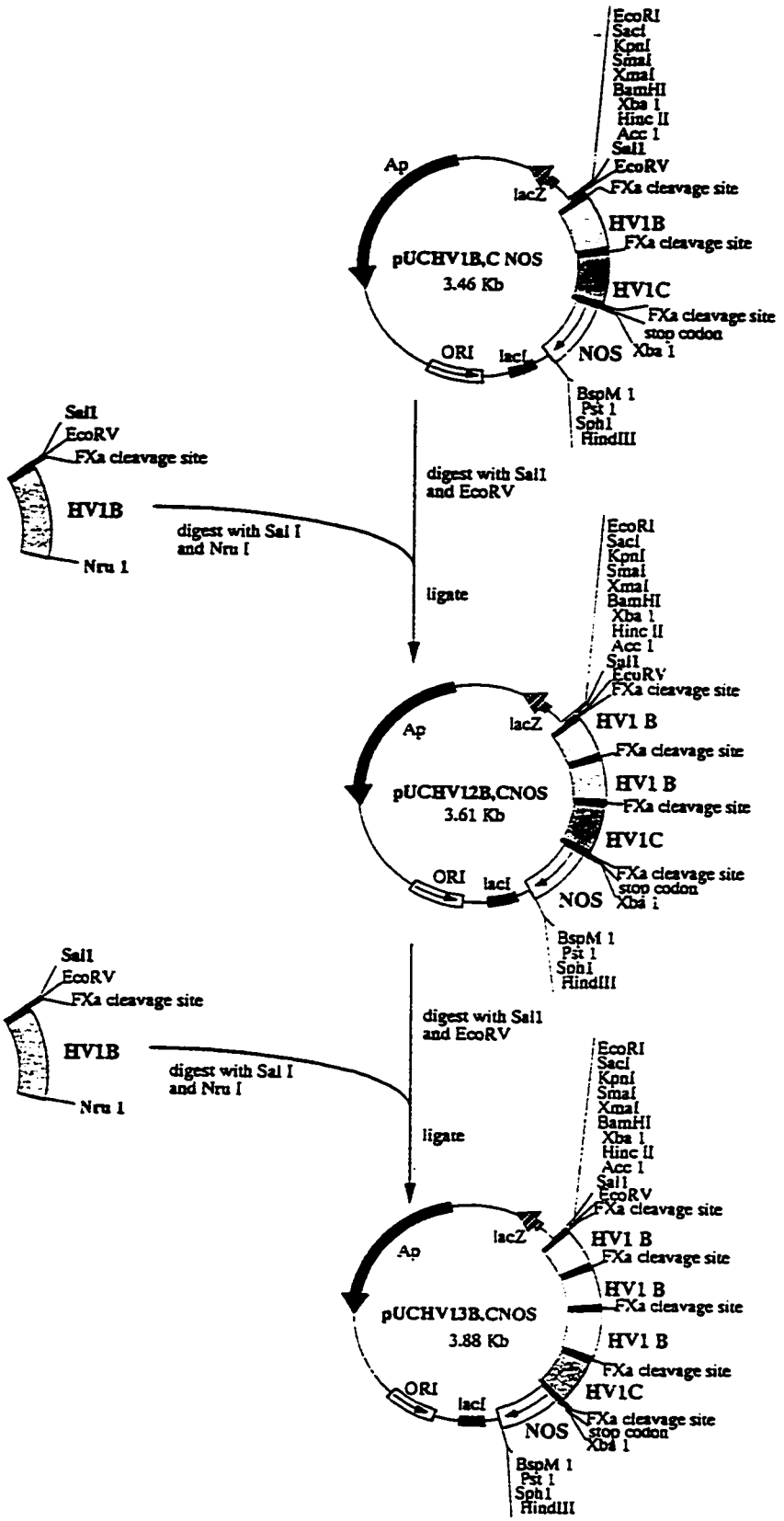
Synthetic HV1-C was created by PCR using the same conditions and primers as that for the production of HV1-A (above) except 5µl of the primer HV1-3'c was used instead of HV1-3'a. HV1-3'c (5'TGCTGCTGCAAGCTTAGATCTTAtctaccctcgat GGACTGGAGATACTCCTCTGG 3'), a reverse primer, has sequences identical to the bottom strand, 3' end of HV1 (bold). Included was a site for translational termination (stop codon; taa)(outlined) a FXa cleavage site (lowercase), and for restriction endonucleases *Bgl* II (underlined) and *Hind* III (double underlined). The PCR product was genecleaned and digested with *Sal* 1, then *Hind* III, and ligated into pUC 19 also digested with these endonucleases. The resulting plasmid was called pUCHV1-C.

Synthetic HV1-B was created by PCR using the same conditions and primers as that for the production of HV1-A (above) except 5µl of the primer HV1-3'b was used instead of HV1-3'a. HV1-3'b (5' TGC TGC TGC AAG CTT TCG CGA **CTG GAG ATA CTC CTC TGG** 3') has sequences identical to the bottom strand, 3' end of HV1 (bold). Two sites for the restriction endonucleases *Nru* I (underlined) and *Hind* III (double underlined) were included. The PCR product was genecleaned and digested with *Sal* 1, then *Hind* III, and ligated into pUC 19 also digested with these endonucleases. The resulting plasmid was called pUCHV1-B.

The production of hirudin concatamers was done by stepwise addition of HV1-B infront of an HV1-C-NOS fusion gene. This construction is illustrated in Fig. 2.7. After digestion with *Bgl* II and *Eco* RI, the HV1 insert of pUCHV1-C was isolated in LMP agarose and ligated into pNOS-10 digested with *Bam* HI and *Eco* RI. The resulting plasmid, pUC-HV1CNOS, was digested with *Eco* RV and *Sal* I. Into that site was inserted HV1-B derived from digestion of pUCHV1-B with *Nru* I and *Sal* I. The resulting plasmid,

Figure 2.7 Construction of pUCHV1 concatamers. HV1C consists of the coding region of hirudin variant 1, in addition to 3' sequences encoding a Factor Xa cleavage site, followed by a stop codon. To the 5' end of HV1C, which was terminated by a NOS transcriptional terminator (HV1CNOS), sequential, in-frame additions of HV1B were ligated. HV1B consists of hirudin variant 1 coding sequence having no stop codons, and a Factor Xa recognition site at its 5' end. Constructs having a total of 1,2,3, and 4 hirudin concatamers were created.


EcoRI
SacI
KpnI
SmaI
XmaI
BamHI
Xba 1
Hinc II
Acc 1
SalI
Ap
lacZ
EcoRV
FXa cleavage site
pUCHV1-C
2.90 Kb
HV1C
FXa cleavage site
stop codon
Bgl II
HinDIII
ORI
lacI
pNOS-10
3.00 Kb
PvuI
NOS
BspM1
PstI
SphI
HindIII
digest with EcoRI
and Bgl II
digest with EcoRI
and BamHI
ligate
pUCHV1CNOS
3.20 Kb
FXa cleavage site
stop codon
Xba 1
BspM 1
Pst 1
Sph1
HindIII
pUCHV1-B
2.90 Kb
HV1B
Nru 1
digest with Sal I
and Nru I
digest with SalI
and EcoRV
pUCHV1B,C NOS
3.46 Kb


EcoRI
SacI
KpnI
SmaI
XmaI
BamHI
Xba 1
Hinc II
Acc 1
SalI
EcoRV
FXa cleavage site
HV1B
FXa cleavage site
HV1C
FXa cleavage site
stop codon
Xba 1
BspM 1
Pst 1
SphI
HindIII
Ap
lacZ
pUCHV1B,C NOS
3.46 Kb
ORI
lacI
NOS
digest with SalI
and EcoRV
ligate
digest with Sal I
and Nru I
Nru 1
pUCHV12B,CNOS
3.61 Kb
HV1 B
pUCHV13B,CNOS
3.88 Kb

pUCHV1B,CNOS was digested with *Eco* RV and *Sal* I. Again, HV1-B, derived from digestion of pUCHV1-B with *Nru* I and *Sal* I was inserted into that site. The sequential addition of HV1-B into the growing fusion gene continued until there were 8 hirudin concatamers in front of NOS (ie. pUC-7HV1B-HV1C-NOS). With each addition of HV1-B, the sequence of the HV1-B inserted was verified by a universal primer. However, only 1, 2, and 4 hirudin concatamers were ultimately chosen for fusion behind oleosin (see results).

2.3.2.2 pCGN-oleosin-HV1 concatamers

The production of oleosin-HV1 gene fusions containing 1, 2, or 4 hirudin-encoding sequences in pCGN-oleosin-HV1 was identical at the cloning level. For each, the plasmid pUCHVAMVcOLHV1ANOS (see above) was used. As shown in Figs. 2.8, 2.9, and 2.10, the HV1A-NOS sequences of pUCHBAMVcOLHV1ANOS were excised with restriction nucleases (*Sal* I and *Hind* III) and replaced with either one (HV1-C), 2 (HV1B,C), or 4 (HV13B,C) hirudin-encoding sequences (each followed by a NOS terminator). The resulting plasmids consisted of the 2.5kbp oleosin promoter (HB) followed by an oleosin cDNA (cOLEO) and either one (HV1-C), 2 (HV1B.C), or 4 (HV13B,C) hirudin-encoding sequences- followed by a nopaline transcriptional terminator (NOS). These plasmids were called, respectively, "pUCHVAMVcOLHV1CNOS", "pUCHVAMVcOLHV1B,C NOS", and "pUCHVAMVcOLHV13B, C NOS". Each of these plasmids was digested with *Kpn*I and *Hind* III, and the insert ligated into pCGN-1559 (also digested with these enzymes) via low melting point (LMP) agarose gel ligation. The resulting plasmids, called pCGN-oleoHV1C, pCGN-oleo2HV1, and pCGN-oleo4HV1, respectively, were analyzed by restriction digest before being transformed into *Agrobacterium tumefaciens* EHA 101.

Figure 2.8 Construction of pCGNoleoHV1C. The coding region of hirudin variant 1 (HV1), which contained sequences encoding a Factor Xa cleavage site, followed by a stop codon at its 3' end, was fused, in-frame, to the 3' end of a *Brassica napus* oleosin cDNA. This construct was terminated by a NOS transcriptional terminator and preceded by a 2.5 kpb *Arabidopsis* oleosin promoter .

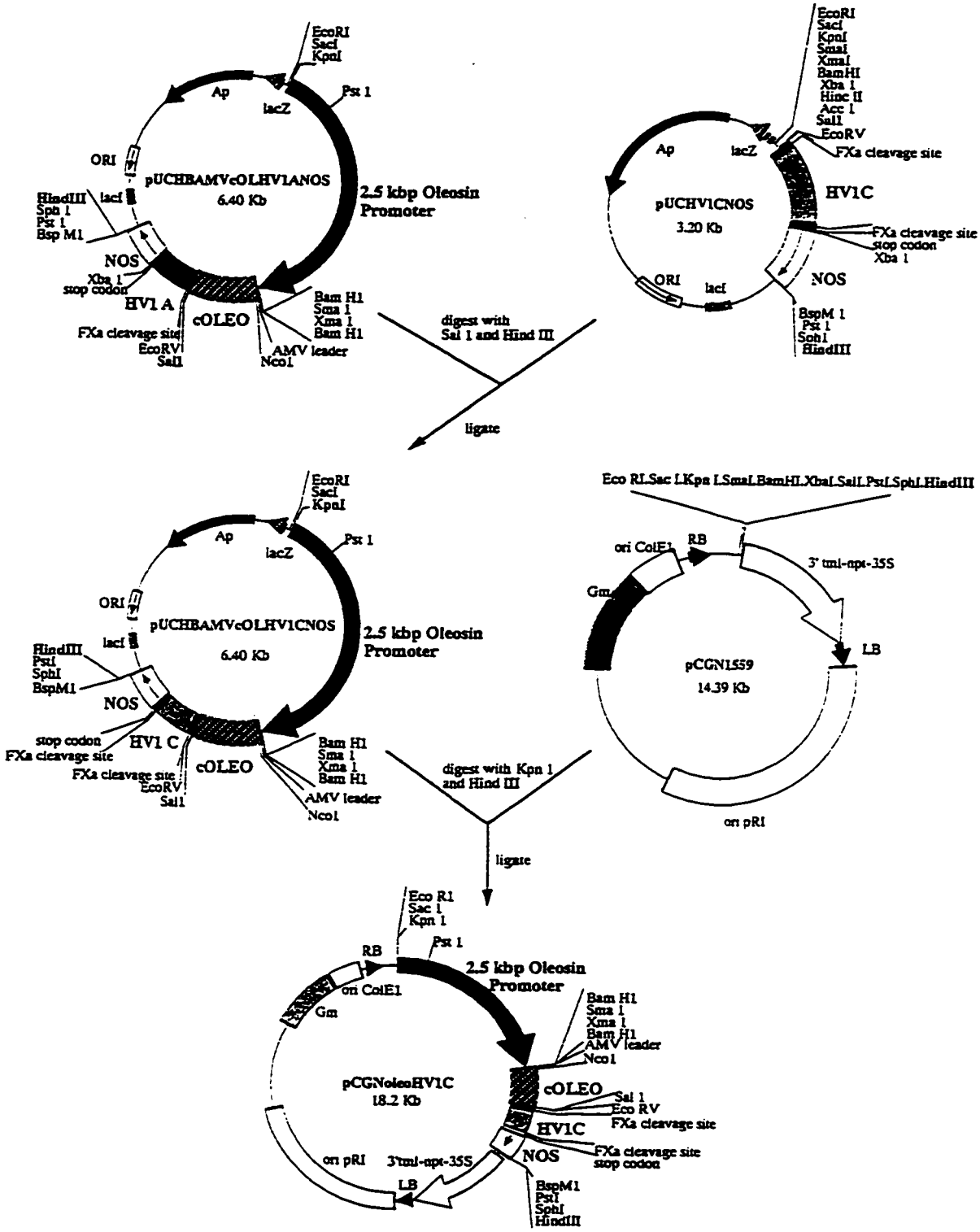

EcoRI
SacI
KpnI
Ap
lacZ
Pst 1
ORI
pUCHBAMVcOLHV1ANOS
6.40 Kb
2.5 kbp Oleosin Promoter
HindIII
Spn 1
Pst 1
Bsp M1
lacI
NOS
Xba 1
stop codon
HV1 A
cOLEO
Bam H1
Sma 1
Xma 1
Bam H1
FXa cleavage site
EcoRV
SalI
AMV leader
NcoI
digest with
Sal 1 and Hind III
ligate
EcoRI
SacI
KpnI
SmaI
XmaI
BamHI
Xba 1
HincII
Acc 1
SalI
EcoRV
FXa cleavage site
pUCHV1CNOS
3.20 Kb
HV1C
FXa cleavage site
stop codon
Xba 1
NOS
BspM 1
Pst 1
SphI
HindIII
ORI
lacI
pUCHBAMVcOLHV1CNOS
6.40 Kb
HV1 C
FXa cleavage site
Eco RI Sac I Kpn I SmaI BamHI XbaI SalI PstI SphI HindIII
ori ColE1
RB
3' tml-npt-35S
Gm
pCGN1559
14.39 Kb
LB
ori pRI
digest with Kpn 1 and Hind III
Eco R1
Sac 1
Kpn 1
Pst 1
pCGNoleoHV1C
18.2 Kb
3'tml-npt-35S

Figure 2.9 Construction of pCGNoleo2HV1. The coding region of hirudin variant 1 containing no stop codon (HV1B), was fused to HV1C, which contains sequences encoding a Factor Xa cleavage site, followed by a stop codon at its 3' end. This hirudin concatamer construct was fused, in-frame, to the 3' end of a *Brassica napus* oleosin cDNA. This construct was terminated by a NOS transcriptional terminator and preceded by a 2.5kpb *Arabidopsis* oleosin promoter .

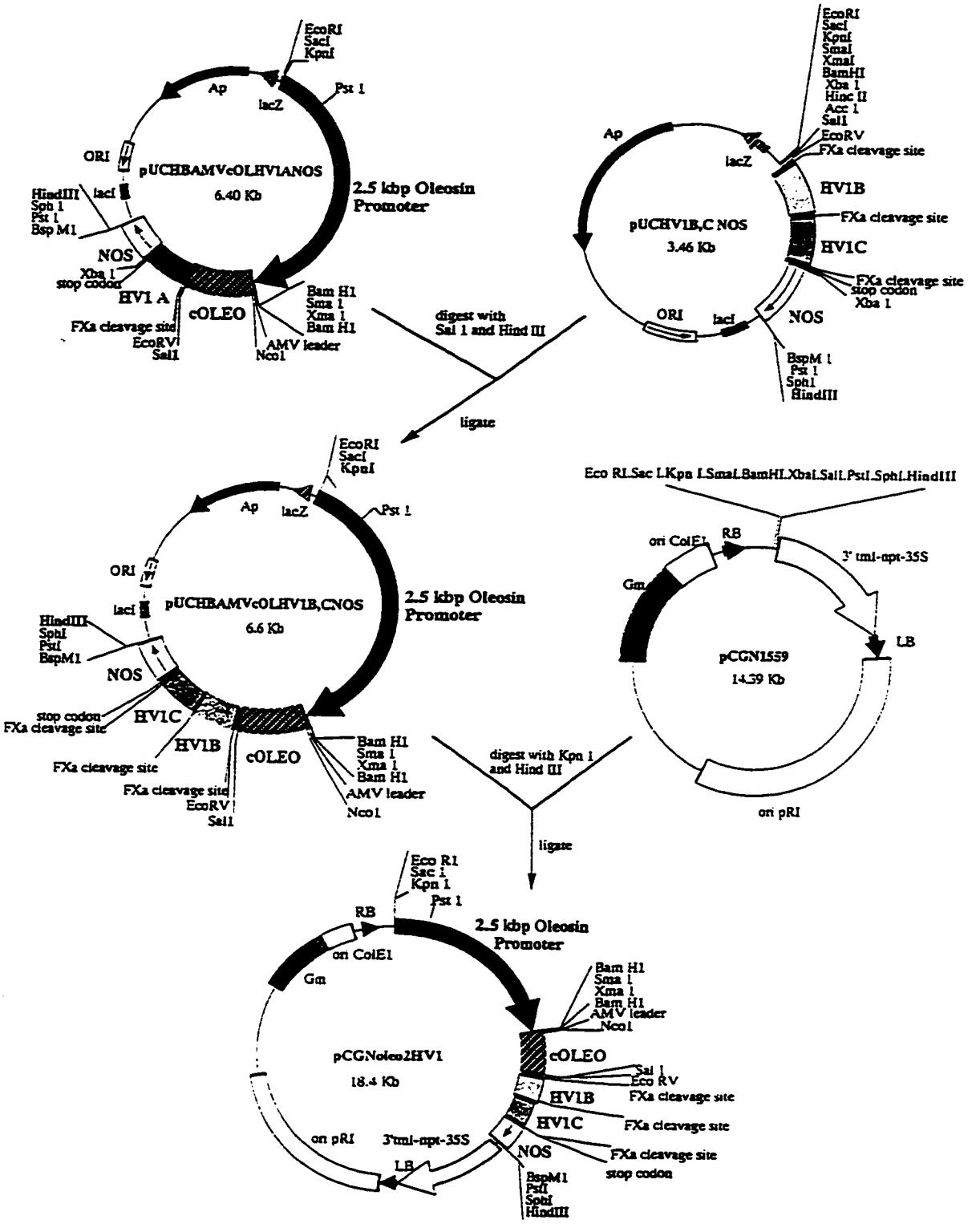
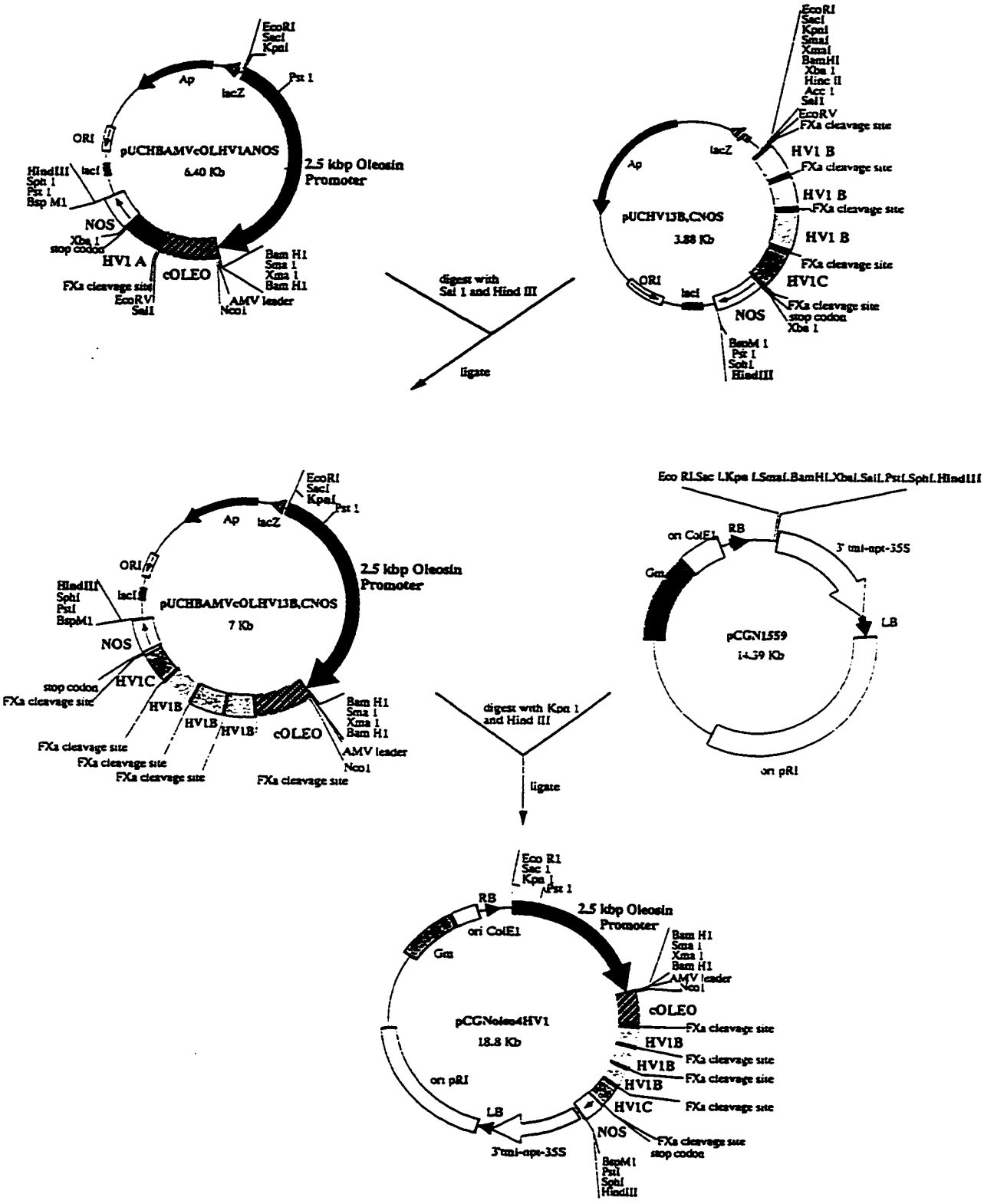

EcoRI
SacI
KpnI
Ap
lacZ
Pst 1
ORI
pUCHBAMVcOLHV1ANOS
6.40 Kb
2.5 kbp Oleosin Promoter
HindIII
Sph 1
Pst 1
Bsp M1
lacI
NOS
Xba 1
stop codon
HV1 A
cOLEO
Bam H1
Sma 1
Xma 1
Bam H1
FXa cleavage site
EcoRV
SalI
AMV leader
Nco1
digest with Sal 1 and Hind III
ligate
EcoRI
SacI
KpnI
SmaI
XmaI
BamHI
Xba 1
Hinc II
Acc 1
SalI
EcoRV
FXa cleavage site
HV1B
pUCHV1B,C NOS
3.46 Kb
HV1C
FXa cleavage site
stop codon
Xba 1
ORI
lacI
NOS
BspM 1
Pst 1
SphI
HindIII
EcoRI
SacI
KpnI
Ap
lacZ
Pst 1
ORI
lacI
pUCHBAMVcOLHV1B,CNOS
6.6 Kb
2.5 kbp Oleosin Promoter
HindIII
SphI
PstI
BspM1
NOS
stop codon
FXa cleavage site
HV1C
HV1B
cOLEO
FXa cleavage site
FXa cleavage site
EcoRV
SalI
Bam H1
Sma 1
Xma 1
Bam H1
AMV leader
Nco1
Eco RI Sac I Kpn I SmaI BamHI XbaI SalI PstI SphI HindIII
ori ColE1
RB
3' tml-npt-35S
Gm
LB
pCGN1559
14.39 Kb
ori pRI
digest with Kpn 1 and Hind III
ligate
Eco R1
Sac 1
Kpn 1
Pst 1
RB
2.5 kbp Oleosin Promoter
ori ColE1
Gm
Bam H1
Sma 1
Xma 1
Bam H1
AMV leader
Nco1
pCGNoleo2HV1
18.4 Kb
cOLEO
Sal 1
Eco RV
FXa cleavage site
HV1B
HV1C
FXa cleavage site
ori pRI
3'tml-npt-35S
NOS
FXa cleavage site
LB
stop codon
BspM1
PstI
SphI
HindIII

Figure 2.10 Construction of pCGNoleo4HV1. Three concatameric hirudin variant 1 sequences containing no stop codons (HV1B), were fused to HV1C, which contains sequences encoding a Factor Xa cleavage site, followed by a stop codon at its 3' end. This 4-hirudin concatamer construct was fused, in-frame, to the 3' end of a *Brassica napus* oleosin cDNA. This construct was terminated by a NOS transcriptional terminator and preceded by a 2.5kpb *Arabidopsis* oleosin promoter .


pUCHBAMVcOLHV1ANOS
6.40 Kb
2.5 kbp Oleosin Promoter
HV1 A
cOLEO
NOS
digest with Sal I and Hind III
ligate
pUCHV13B,CNOS
3.88 Kb
HV1 B
HV1C
pUCHBAMVcOLHV13B,CNOS
7 Kb
pCGN1559
14.39 Kb
pCGNoleo4HV1
18.8 Kb
digest with Kpn I and Hind III
ligate

Chapter 3:

Expression of oleosin-hirudin variant 2 fusion gene in tobacco and *Brassica napus*

3.1 Introduction

The economic importance of many biologically active peptides makes their large-scale synthesis highly desirable. Biologically active peptides or proteins may be produced through chemical synthesis, or via biological expression systems such as bacterial or yeast fermentation, or mammalian cell cultures. As described in Chapter 1, the various problems described in these and other expression systems have been the driving force in the search for alternative expression systems for the production of heterologous proteins. Plant-based expression systems have received attention as alternative hosts for the production of recombinant proteins and peptides (Krebbers and Vandekerckhove, 1990). The ease by which DNA sequences may be transferred and stably integrated into the genome of many plant species (Horsch *et al.*, 1985; Radke *et al.* 1988), the ability of plants to be regenerated from single cells (totipotency), and the efficient selection of regenerated plant transformants is conducive to this kind of experimentation.

In this chapter, I describe investigations into the feasibility of using oleosins as a fusional carrier for the production of the blood anticoagulant, hirudin (variant 2) (Harvey *et al.*, 1986). As stated in Chapter 1 (section 1.5), *Brassica napus* oleosins possess a number of important qualities which make them suitable "carriers" of heterologous proteins of interest. Hirudin variant 2 is an attractive model for this type of evaluation because it has a low molecular weight (approximately 7 kDa), has minimal processing requirements, is relatively hydrophilic, and its anti-thrombin activity is easily assayed colorimetrically. At the time that these experiments were initiated, a genomic clone of an *Arabidopsis* oleosin had been isolated (van Rooijen *et al*,. 1992). The deduced amino acid sequence of this clone is highly homologous (91% identical) to that of *Brassica napus*. In addition, the 867 nucleotides of the 5' untranslated upstream region of this *Arabidopsis* genomic clone is 79% identical in sequence to that of *B. napus*. Given this similarity, it was expected that

this clone would be expressed in a seed-specific manner and accurately targeted to oil bodies when introduced into *B. napus*. This genomic clone was modified to include an in-frame translational fusion encoding hirudin at its 3' end. To facilitate purification of hirudin, a sequence encoding a FXa/ clostripain cleavage site was included between the protein-encoding genes (Fig. 3.1). The following is an investigation of the expression, accumulation, stability, targeting and biological activity of the oleosin-hirudin variant 2 fusion protein in *B. napus* and tobacco seeds.

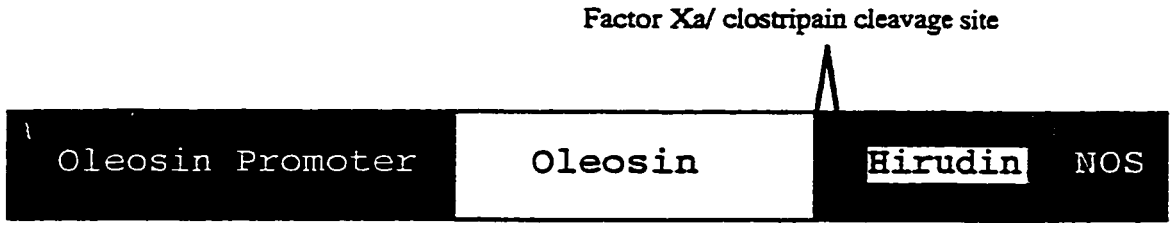
3.2 Results

3.2.1 Accumulation of oleosin-hirudin variant 2 fusion protein in tobacco

3.2.1.1 Western blot analysis of seed protein from tobacco transformants

Tobacco leaf disks were transformed (section 2.2.8) with the oleosin-hirudin variant 2 fusion gene (pCGN-OBHIRT) (section 2.3.1.1, Fig. 2.1) as described in Chapter 2. Dry seed (0.5g each) from wild-type (WT) tobacco plants or three tobacco plant lines transformed with pCGN-OBHIRT were ground in 3 ml of 50mM Tris-HCl (pH 7.5). Seventy-five micrograms each of total seed protein from wild type and the three transformed tobacco lines was loaded onto a 15% SDS-polyacrylamide gel and electrophoresed. The proteins were transferred onto PVDF membranes and incubated with mouse monoclonal antibodies raised against hirudin. As shown in Fig. 3.2, an approximately 26 kDa band (arrow), corresponding to the expected molecular weight of the oleosin-hirudin fusion protein, was detected in the three transformed plant seed extracts (OBHIRT-A, B, and C). This band was not apparent in the wild-type (WT) extract. The hirudin antibody cross-reacts with a series of lower molecular weight proteins in all of the samples. However, this cross-reactivity was apparent only when total seed protein extracts were used. When preparations of oil body protein from wild-type and transformed tobacco

Figure 3.1 A) Diagrammatic representation of oleosin-hirudin variant 2 fusion gene construct introduced into *B. napus* and tobacco via *Agrobacterium*-mediated transformation. the oleosin hirudin fusion gene is under the control of an *Arabidopsis* oleosin prompter (838 bp), and is terminated by a 267 bp nopaline synthase transcriptional terminator (NOS). B) DNA and deduced amino acid sequence of the translated region of the oleosin-hirudin fusion gene. Bases 1-519 (amino acids 1-173) represents the nucleotide and deduced amino acid sequence of the *Arabidopsis* oleosin. The 195 bp/ 65 amino acid sequence of hirudin follows the Factor Xa/clostripain recognition site (boxed).

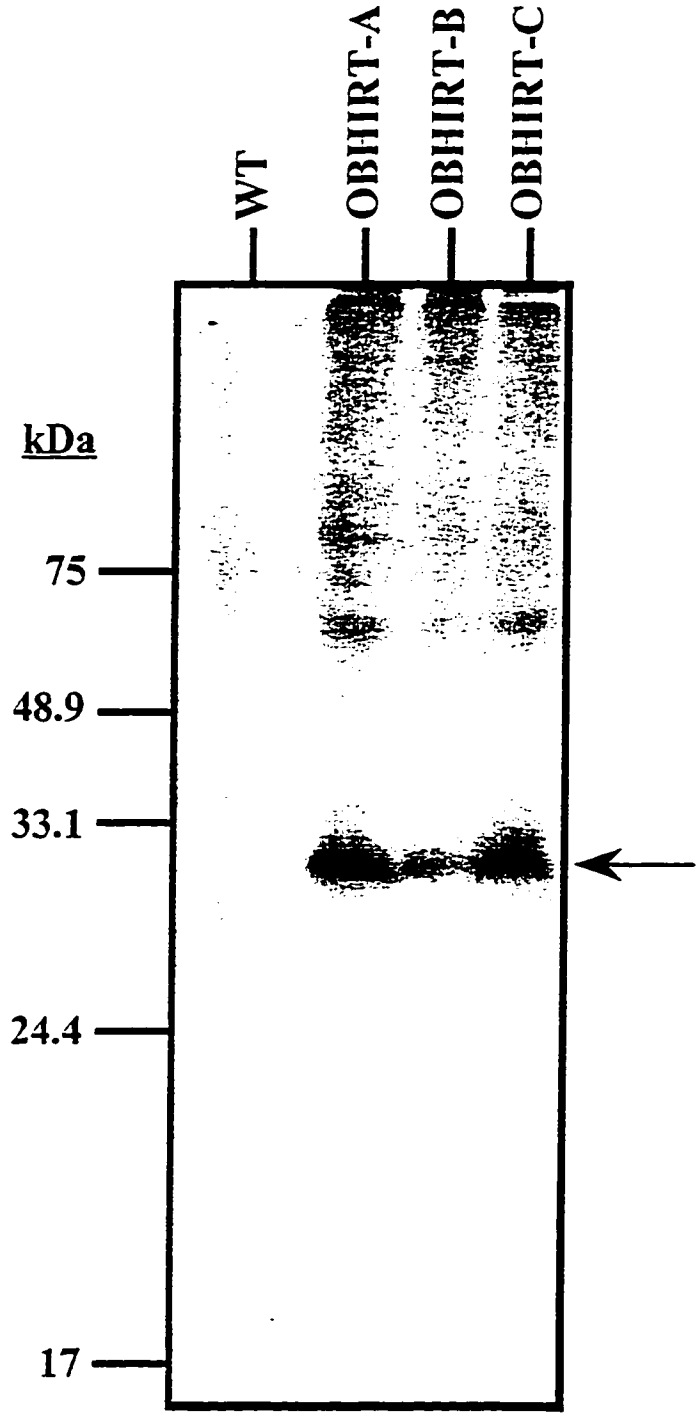


A

```
1/1                                                 31/11
ATG GCG GAT ACA GCT AGA GGA ACC CAT CAC GAT ATC ATC GGC AGA GAC CAG TAC CCG ATG
M   A   D   T   A   R   G   T   H   H   D   I   I   G   R   D   Q   Y   P   M
61/21                                               91/31
ATG GGC CGA GAC CGA GAC CAG TAC CAG ATG TCC GGA CGA GGA TCT GAC TAC TCC AAG TCT
M   G   R   D   R   D   Q   Y   Q   M   S   G   R   G   S   D   Y   S   K   S
121/41                                              151/51
AGG CAG ATT GCT AAA GCT GCA ACT GCT GTC ACA GCT GGT GGT TCC CTC CTT GTT CTC TCC
R   Q   I   A   K   A   A   T   A   V   T   A   G   G   S   L   L   V   L   S
181/61                                              211/71
AGC CTT ACC CTT GTT GGA ACT GTC ATA GCT TTG ACT GTT GCA ACA CCT CTG CTC GTT ATC
S   L   T   L   V   G   T   V   I   A   L   T   V   A   T   P   L   L   V   I
241/81                                              271/91
TTC AGC CCA ATC CTT GTC CCG GCT CTC ATC ACA GTT GCA CTC CTC ATC ACC GGT TTT CTT
F   S   P   I   L   V   P   A   L   I   T   V   A   L   L   I   T   G   F   L
301/101                                             331/111
TCC TCT GGA GGG TTT GGC ATT GCC GCT ATA ACC GTT TTC TCT TGG ATT TAC AAG TAC GCA
S   S   G   G   F   G   I   A   A   I   T   V   F   S   W   I   Y   K   Y   A
361/121                                             391/131
ACG GGA GAG CAC CCA CAG GGA TCA GAC AAG TTG GAC AGT GCA AGG ATG AAG TTG GGA AGC
T   G   E   H   P   Q   G   S   D   K   L   D   S   A   R   M   K   L   G   S
421/141                                             451/151
AAA GCT CAG GAT CTG AAA GAC AGA GCT CAG TAC TAC GGA CAG CAA CAT ACT GGT TGG GAA
K   A   Q   D   L   K   D   R   A   Q   Y   Y   G   Q   Q   H   T   G   W   E
481/161                                             511/171
CAT GAC CGT GAC CGT ACT CGT GGT GGC CAG CAC ACT ACT gcg atc gaa ggg aga atc act
H   D   R   D   R   T   R   G   G   Q   H   T   T   A   I   E   G   R   I   T
541/181                                             571/191
tac acc gac tgt act gaa tct gga cag aac ctc tgt ctc tgt gaa gga tct aac gtt tgt
Y   T   D   C   T   E   S   G   Q   N   L   C   L   C   E   G   S   N   V   C
601/201                                             631/211
gga aag gga aac aag tgt atc ctc gga tct aac gga aag gga aac cag tgt gtt act gga
G   K   G   N   K   C   I   L   G   S   N   G   K   G   N   Q   C   V   T   G
661/221                                             691/231
gaa gga act cca aac cca gaa tct cac aac aac gga gac ttc gaa gaa atc cct gaa gaa
E   G   T   P   N   P   E   S   H   N   N   G   D   F   E   E   I   P   E   E
721/241
tac ctc cag
Y   L   Q
```

B

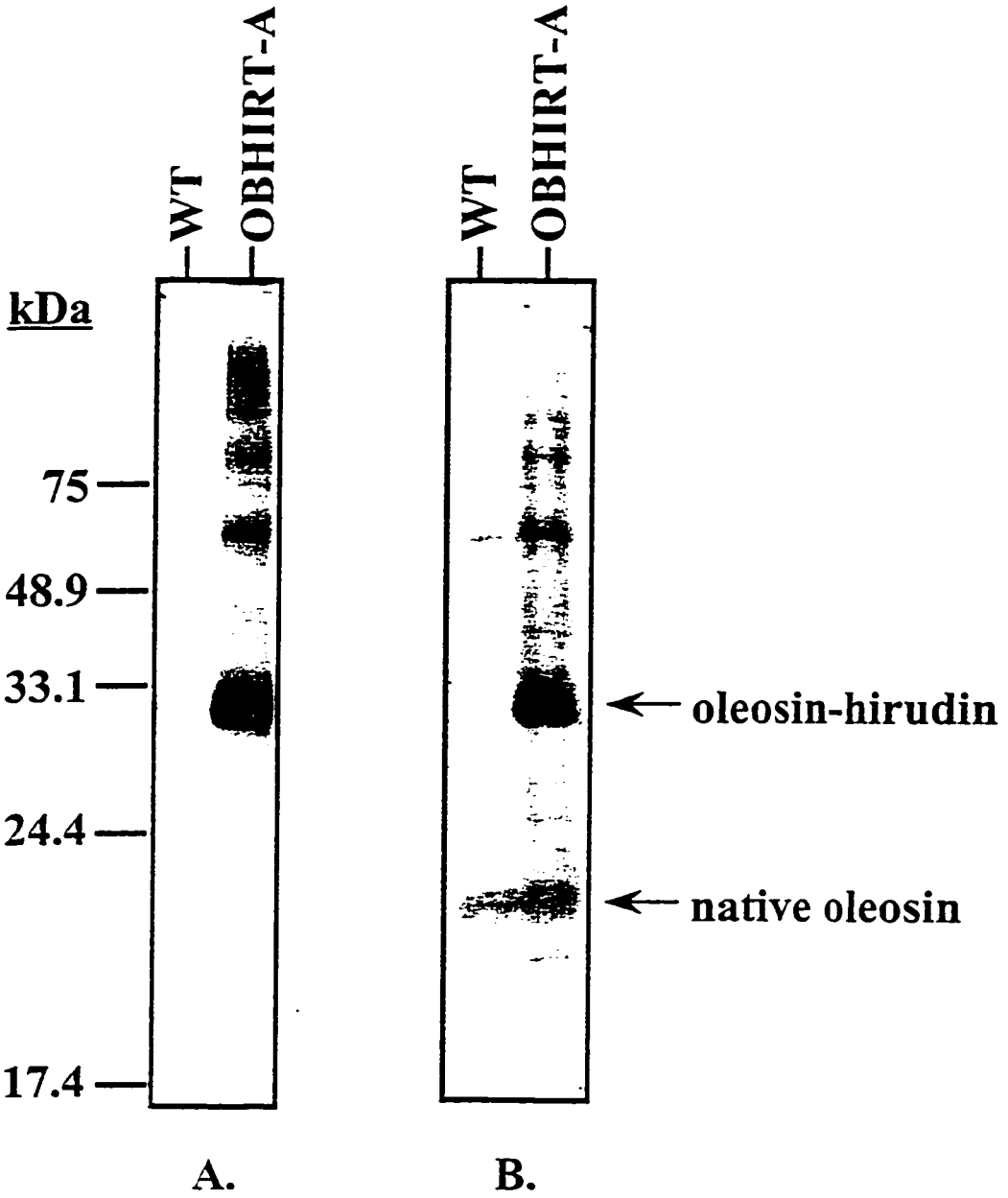
Figure 3.2 Immunological detection of oleosin-hirudin variant 2 fusion protein in transformed tobacco. Total seed protein extracts (75μg total protein extract/lane) derived from wild type (WT) tobacco and three tobacco lines transformed with the pCGN-OBHIRT construct (OBHIRT-A, OBHIRT-A, OBHIRT-C) were subjected to SDS-PAGE (15% polyacrylamide gel), transferred onto a PVDF membrane, and incubated in the presence of mouse monoclonal antibodies raised against hirudin. Antibodies recognized a band (arrow) in transformed, but not wild type extracts.


WT
OBHIRT-A
OBHIRT-B
OBHIRT-C
kDa
75
48.9
33.1
24.4
17

were subjected to Western blotting, these lower molecular weight bands were not apparent (Fig. 3.3). Seed from one of these plants, OBHIRT-A, was further analyzed in order to determine if the ca. 26kDa band shown in Fig. 3.2 is, indeed, an oleosin-hirudin fusion protein, and, further, to determine if this protein targets to the oil body fraction of tobacco extracts.

As shown in Fig. 3.3, oil body proteins prepared from wild-type (WT) tobacco seed and seed from tobacco transformed with the oleosin-hirudin variant 2 fusion gene (OBHIRT-A) were electrophoresed through a 15% polyacrylamide gel, as above. Oil body samples (10μg/lane) were electrophoresed in duplicate, and transferred onto two PVDF membranes. One of these membranes was incubated with antibodies raised against oleosin and the other incubated with antibodies raised against hirudin. As shown in Fig. 3.3, panel A, a band corresponding to a molecular mass of ca. 26 kDa was recognized by anti-hirudin antibodies in oil body proteins of transformed tobacco (OBHIRT-A). These antibodies did not detect this band, nor any other band, in wild-type (WT) oil body preparations. This result clearly demonstrates that the 26kDa band is unique to transformed *vs.* wild type plant seed oil bodies. As shown in Fig. 3.3, panel B, a band having the same electrophoretic mobility as that recognized by the anti-hirudin antibodies in panel A (upper arrow), was also detected by antibodies recognizing 19 kDa *B. napus* oleosins in transformed (OBHIRT-A) plant oil body preparations. This 26kDa band was not detected in wild type oil body extracts by either antibody. In addition, tobacco oleosins, having a molecular weight of approximately 20 kDa, were recognized at approximately equal levels by the anti-*B. napus* antibodies in both plant types. This demonstrates that the amount of oil body protein loaded from each plant type was essentially the same. In addition, the ability of anti-*B. napus* oleosin antibodies to recognize tobacco oleosins demonstrates that they share common epitope(s).

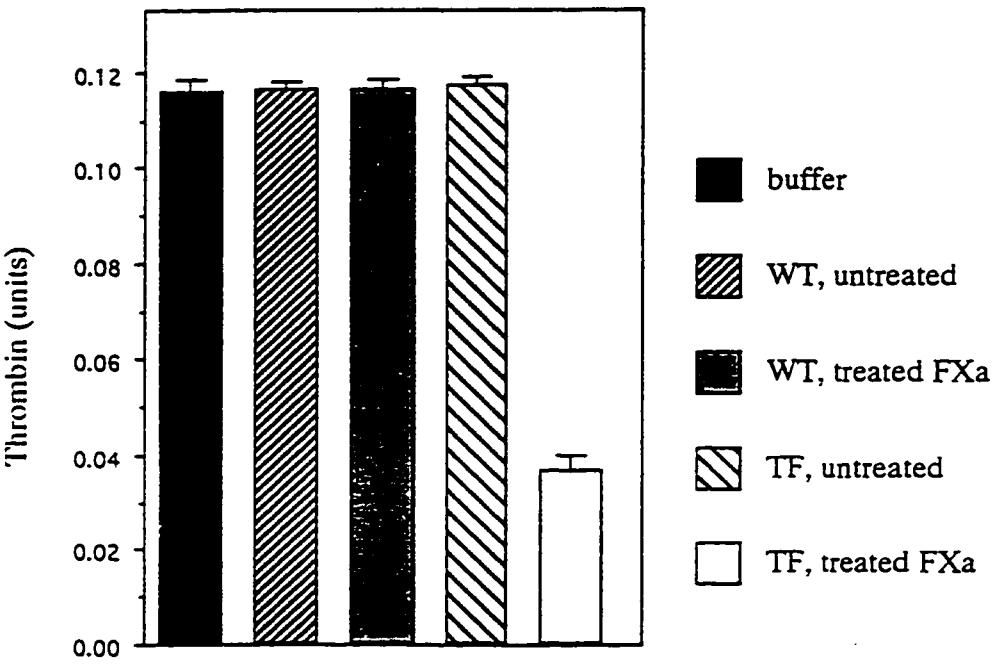
Figure 3.3 Western blot analysis of oil body protein extracts (10μg protein/lane) derived from wild type (WT) tobacco or a tobacco line transformed with the pCGN-OBHIRT construct (OBHIRT-A). Samples were run in duplicate via SDS-PAGE (15% polyacrylamide gel) and transferred onto two PVDF membranes. One membrane was incubated with mouse monoclonal antibodies raised against hirudin (panel A). The other blot was incubated in the presence of rabbit polyclonal antibodies raised against the 19kDa *B. napus* oleosin. As indicated by the upper arrow, a band of equal molecular weight was recognized by both antibodies in transformed, but not wild type oil body extracts (panels A and B). Native tobacco oleosins (lower arrow) were recognized by antibodies raised against 19kDa *B. napus* oleosins in both wild type and transformed oil body extracts (panel B).


WT
OBHIRT-A
kDa
75
48.9
33.1
24.4
17.4
A.
WT
OBHIRT-A
oleosin-hirudin
native oleosin
B.

3.2.1.2 Anti-thrombin assays of seed protein from tobacco transformants

In order to determine if hirudin expressed as an oleosin fusion in transgenic tobacco can be recovered in a biologically active form, oil body proteins from the transgenic tobacco line, OBHIRT-A, were prepared and digested with the endoprotease, Factor Xa. Six hundred milligrams of dry tobacco seed from transformed and wild-type seed were ground in 10ml each of extraction buffer as described in Chapter 2 (section 2.2.13). This extraction was performed in triplicate for both seed types (i.e. three oil body preparations were performed for both wild type and transformed seed) (referred to as prep 1, 2, and 3 in Appendix 1) in order to determine if there was sampling error. After the oil body "fat pad" was recovered for each preparation, it was resuspended and recentrifuged (washed) three times in thrombin assay buffer (50mM Tris-HCl, 100mM NaCl, 5mM $CaCl_2$). The final samples were quantitated and diluted to 6.3μg oil body protein/μl. Three hundred microliters from each preparation was incubated with 6.6μl of buffer (untreated) or 6.6μl of Factor Xa (1mg/ml) (treated) overnight at room temperature with gentle shaking. The preps incubated in the presence or absence of Factor Xa were centrifuged and the unternatant was isolated. Fifteen microliters of unternatant from each prep was assayed in triplicate for anti-thrombin activity as described in Chapter 2 (section 2.2.19). The results of this assay are summarized in Fig. 3.4. Fifteen microliters of buffer (control) or unternatant from wild-type (WT) or transformed (TF) oil body proteins treated or untreated with Factor Xa was added to 1ml of reaction buffer containing thrombin and a chromogenic substrate. Thrombin activity present after addition of these extracts was determined by measuring the rate of absorption at 405nm. Each assay was performed in triplicate, the mean calculated, and the standard error of the mean was derived therefrom. As shown in Fig. 3.4, thrombin activity is approximately the same in reaction buffer containing

Figure 3.4 Measurement of anti-thrombin activity in oil body extracts from wild type and transformed tobacco seeds. Oil bodies were prepared in triplicate (prep 1, 2, and 3, Appendix 1) from wild type (WT) tobacco seed and seed from tobacco transformed (TF) with pCGN-OBHIRT. Each preparation was treated, or untreated, in the presence of Factor Xa (FXa). The unternatant from these samples, or buffer alone, was added to thrombin in buffer. Thrombin activity remaining in each sample after addition of unternatant or buffer was measured colorimetrically (OD_{405nm}), in triplicate, and converted to units of thrombin using a thrombin standard curve. No significant difference in anti-thrombin activity was measured between oil body preparations of WT treated or untreated, or TF untreated with Factor Xa (Appendix 1). The mean thrombin activity and the standard error of the mean measured for each sample type is given. As shown, thrombin activity is only reduced when unternatant from FXa-treated, transformed oil bodies (TF, treated) is added.


Thrombin (units)
0.12
0.10
0.08
0.06
0.04
0.02
0.00
buffer
WT, untreated
WT, treated FXa
TF, untreated
TF, treated FXa

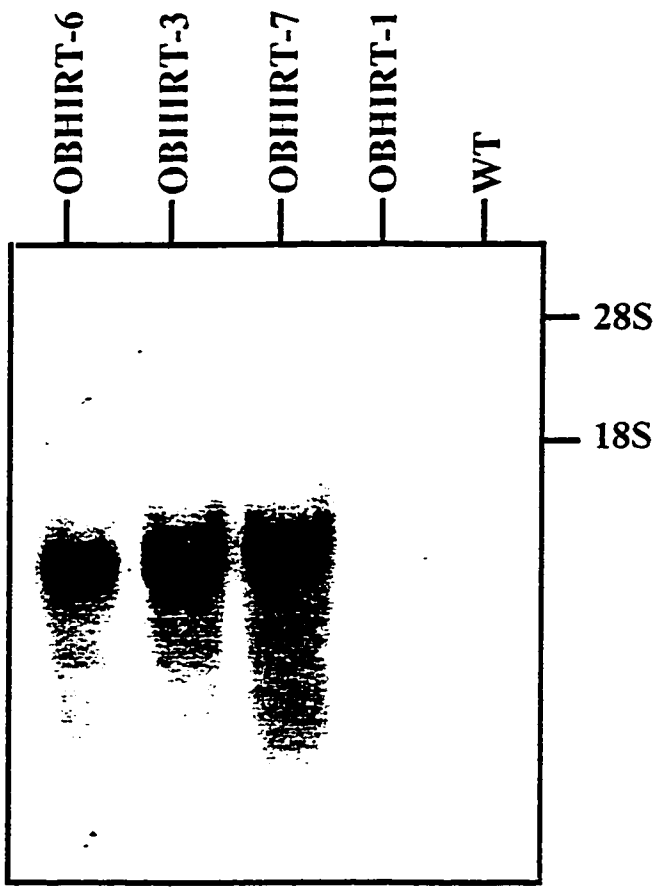
15µl of buffer, alone, or wild-type extracts (either treated or untreated with FXa). This demonstrates that addition of seed unternatant has no effect on the assay, itself. Interestingly, no anti-thrombin activity is apparent in transformed oil body unternatant untreated with FXa. However, when 15µl of unternatant from transformed oil bodies treated with FXa was added to the reaction buffer, a significant decrease in thrombin activity was observed. As illustrated by Appendix 1, no significant difference between each of the three preps, based on anti-thrombin activity, was observed. This demonstrates the reproducibility of these results (i.e. low systematic error), and the reliability of the assay itself (low sampling error). The overall anti-thrombin activity in transformed tobacco oil body proteins digested with FXa is 0.84 anti-thrombin units (ATU)/mg oil body protein.

3.2.2 Expression of an oleosin-hirudin variant 2 fusion gene in *B. napus*

3.2.2.1 Northern blot analysis

The oleosin-hirudin variant 2 fusion gene construct, pCGN-OBHIRT, (section 2.3.1.1, Fig. 2.1) was introduced into the *Brassica napus* genome as described in Chapter 2 (section 2.2.7). Southern blot analysis of kanamycin resistant plants identified a number of transformants which had incorporated an average of 2-3 copies of the fusion gene into their genome. These transformants were screened for the expression of the oleosin-hirudin mRNA. RNA was extracted from developing seeds (mid-cotyledonary stage) of transformed and wild type *B. napus* plants. Expression of the oleosin-hirudin mRNA was observed on northern blots probed with radiolabelled hirudin-encoding cDNA (sections 2.2.11 and 2.2.12). As shown in Fig. 3.5, hirudin gene expression occurred in three of the four transformants tested. The hirudin DNA probe did not hybridize to seed RNA from wild-type plants. Northern blot analysis of RNA from different tissues of transformed and

Figure 3.5 Northern blot demonstrating expression of the oleosin-hirudin variant 2 fusion gene in oilseed rape embryos. Thirty-five micrograms of total RNA from mid-cotyledonary embryos derived from wild type (WT) and four different *B. napus* transformants (OBHIRT-6, -3, -7. and -1) were loaded onto a 1.2% agarose gel containing 6% formaldehyde. The gel was run overnight at 20V and the RNA blotted onto Hybond N. The membrane was probed with ^{32}P-labeled hirudin cDNA (1.75 X 10^4 becquerels/ml hybridization buffer).


OBHIRT-6
OBHIRT-3
OBHIRT-7
OBHIRT-1
WT
28S
18S

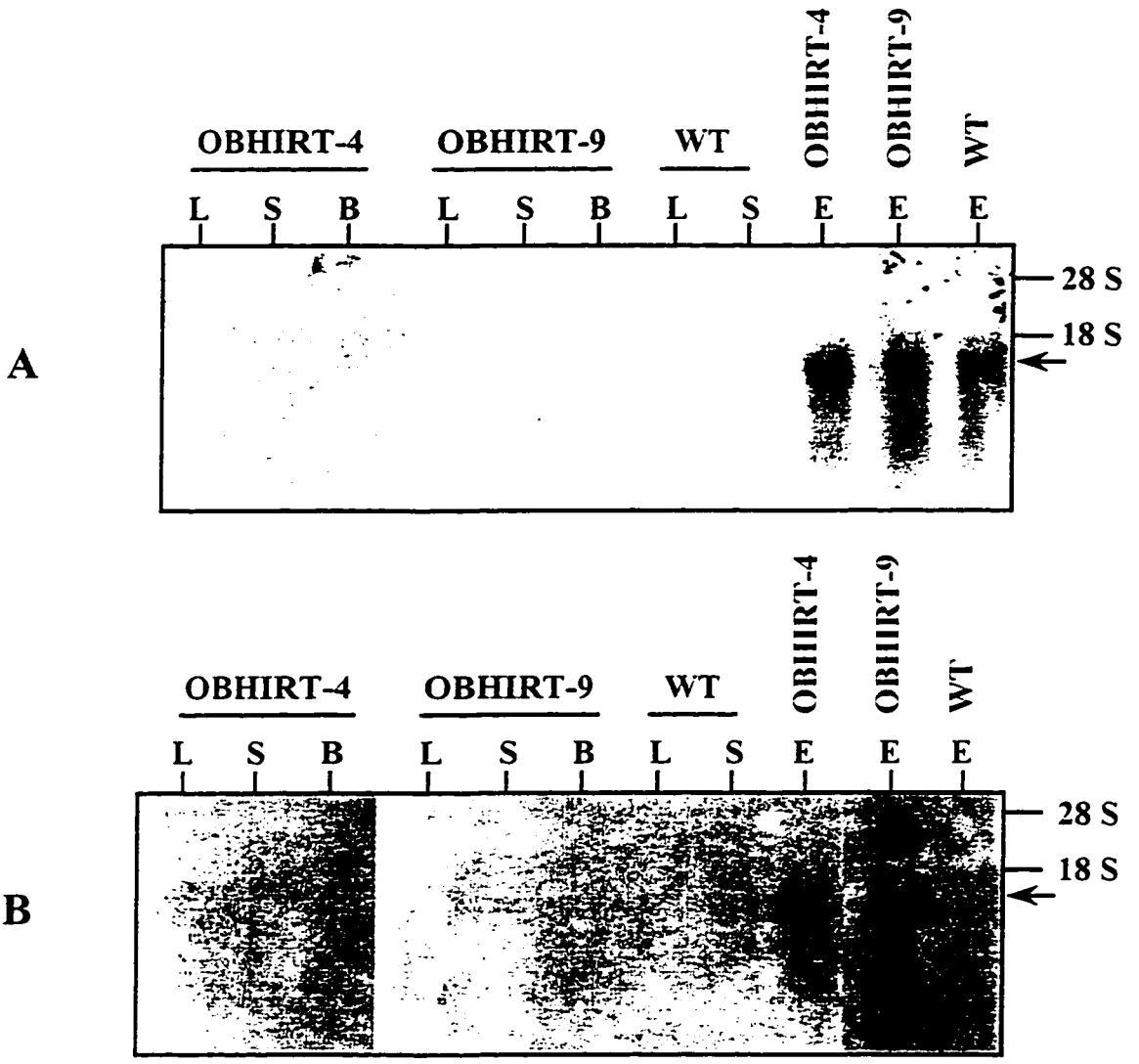
wild-type *B. napus* was also performed in order to determine if the expression of this heterologous gene is embryo-specific. Fig. 3.6 shows a northern blot of leaf (L), stem (S), bud (B), and embryonic (mid-cotyledonary) (E) total RNA from wild-type (WT) *B. napus* and two transformed plant lines (OBHIRT-4 and OBHIRT-9). This blot was probed with radiolabelled DNA encoding oleosin (panel A) then stripped and reprobed with radiolabelled hirudin-encoding cDNA (panel B). As shown in Fig. 3.6, panels A and B, neither probe hybridized to RNA from leaf, stem or bud tissue of WT or transformed plants. As expected, the oleosin probe hybridized to embryonic RNA from all three plant types (panel A). However, radiolabelled hirudin-encoding cDNA hybridized to the embryonic RNA of transformed, but not wild-type, embryonic RNA (panel B). These results demonstrate that the 0.8kbp upstream region of the *Arabidopsis* oleosin gene drives seed-specific expression of the heterologous fusion gene in *Brassica napus*.

3.2.2.2 Localization of oleosin-hirudin variant 2 fusion protein in B. napus seeds

From the above experiments (section 3.2.2.1), it is clear that the oleosin-hirudin variant 2 fusion gene is expressed in a seed-specific manner. Seeds from these plants were used for the determination of accumulation and targeting of the oleosin-hirudin fusion protein in transgenic *B. napus*.

Dry seed (0.5g) from the *B. napus* transformant, OBHIRT-9, were ground in 8 ml extraction buffer, overlaid with 5 ml of flotation buffer, and centrifuged (section 2.2.13). The insoluble, soluble, and oil body fractions were isolated and washed, and their volumes equalized to 13 ml each (section 2.2.13). The volumes of each of these fractions were normalized in order to maintain the relative proportion of seed protein found in each fraction. Total seed protein from wild type and OBHIRT-9 was isolated by grinding 0.5g

Figure 3.6 Northern blot demonstrating seed-specific expression of the oleosin-hirudin variant 2 fusion gene in oilseed rape embryos. Forty micrograms of total RNA from leaf (L), stem (S), bud (B) or mid-cotyledonary embryos (E) was electrophoresed through a 1.2% agarose gel containing 6% formaldehyde. Tissue was derived from wild type (WT) plants or two *B. napus* lines transformed with pCGN-OBHIRT (OBHIRT-4, -9). RNA was blotted onto Hybond N and the membrane was probed with ^{32}P-labeled oleosin-encoding DNA (panel A). The blot was stripped of radiolabel and reprobed with ^{32}P-labeled hirudin cDNA (panel B). To each milliliter of hybridization solution was added 1.75×10^4 becquerels of radiolabelled DNA. As expected, radiolabelled oleosin-encoding DNA hybridized to embryonic RNA (E) from transformed and wild type *B. napus* (panel A, arrow). This band was also apparent in embryonic RNA from transformed, but not wild type plants (panel B, arrow) probed with radiolabelled hirudin cDNA. As shown in both panels, neither probe hybridized to RNA derived from wild type leaf or stem tissue, nor from transformed leaf, stem or bud tissue.


A
OBHIRT-4
OBHIRT-9
WT
OBHIRT-4
OBHIRT-9
WT
L S B L S B L S E E E
28 S
18 S
B
OBHIRT-4
OBHIRT-9
WT
OBHIRT-4
OBHIRT-9
WT
L S B L S B L S E E E
28 S
18 S

dry seed in 13 ml of extraction buffer. Once extracted, the total seed protein and oil body protein from OBHIRT-9 seed was loaded in equal volumes (and therefore proportional amounts) onto a 15% polyacrylamide gel, electrophoresed, and stained with Coomassie blue. As shown in Fig. 3.7, total seed protein from *B. napus* is comprised primarily of the storage proteins napin (4 and 9 kDa) and cruciferin (22-33kDa) and oleosin (19kDa). Isolation and repeated "washing" of the oil body fat pad via flotation centrifugation derived from this total seed protein results in the removal of most of the storage proteins with minimal loss of oleosins (Fig. 3.7).

The localization of the oleosin-hirudin fusion protein in seed extracts was determined by western blot analysis. Aliquots of equalized fractions from total, soluble, resuspended insoluble, and oil body protein from transformed (TF) plant seeds, and total seed protein from wild type (WT) plants, were loaded onto an SDS-polyacrylamide gel (Fig. 3.8). After electrophoresis, the proteins were transferred onto a PVDF membrane and incubated with anti-hirudin monoclonal antibodies. As shown in Fig. 3.8, the oleosin-hirudin fusion protein, which has an expected molecular weight of 26 kDa, was preferentially localized to the oil body fraction (TF, OBP). The oleosin-hirudin fusion protein was also detected by these antibodies, although to a much lesser extent, in the soluble protein fraction. This may be due either to the difficulty associated with complete removal of oil bodies from the soluble seed protein fraction, or the disruption of oil bodies and subsequent release of oleosins during preparation. Whenever oleosin-hirudin fusion proteins were detected immunologically in seed fractions other than the fat pad, native oleosin proteins were also detected. The anti-hirudin antibodies did not recognize any protein from wild-type total seed protein (WT, total).

It was expected that the oleosin-hirudin fusion protein would be recognized by both anti-hirudin and anti-oleosin antibodies. To prove this, total seed protein from transformed

Figure 3.7 Oleosin enrichment from total seed protein via floatation centrifugation. Total seed protein and oil body protein was obtained from an equal weight of seed (0.5 g) from plants transformed with the oleosin-hirudin variant 2 fusion gene (pCGN-OBHIRT). The final volume of total seed protein and oil body protein obtained were equalized (13 ml). Equal volumes from both seed preparations (5.4µl each) were loaded and electrophoresed via SDS-PAGE (15% polyacrylamide gel). After electrophoresis, the gel was stained with Coomassie blue.

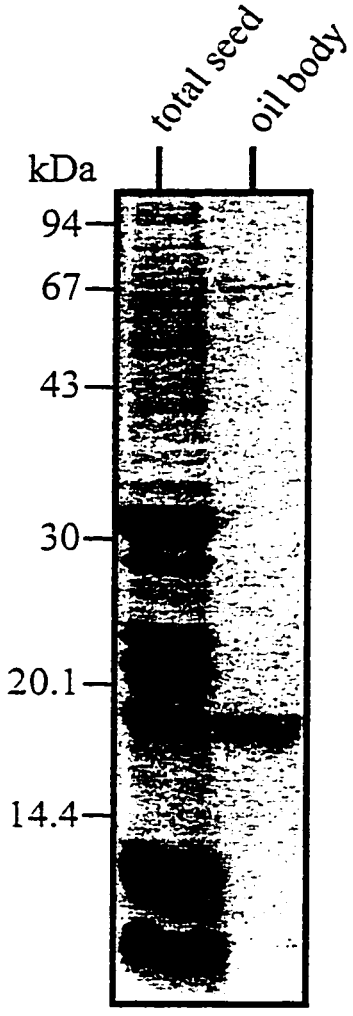
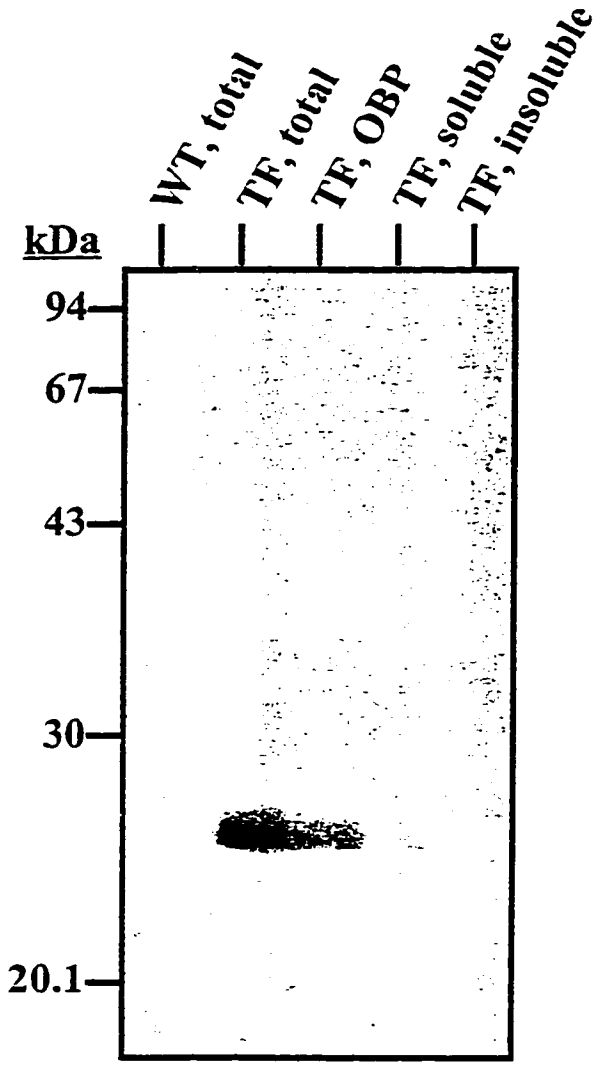

total seed
oil body
kDa
94
67
43
30
20.1
14.4

Figure 3.8 Immunological determination of localization of oleosin-hirudin fusion protein by fractionation of seed proteins. Total seed protein (total), oil body protein (OBP), soluble, or insoluble (pellet) seed protein was extracted from equal amounts of transformed (TF) *Brassica napus* seeds as described in section 2.2.13. Total seed protein from wild-type (WT) seeds was also extracted. After normalizing the volumes of all fractions, equal volumes of protein were electrophoresed via SDS-PAGE (12% polyacrylamide gel), blotted onto PVDF membrane, and incubated with anti-hirudin monoclonal antibodies. No the antibody binding is observed to proteins derived from wild type seed or from the transfomed seed pellet fraction.


WT, total
TF, total
TF, OBP
TF, soluble
TF, insoluble
kDa
94
67
43
30
20.1

and wild-type seeds was loaded in duplicate and transferred onto two membranes. One membrane was probed with anti-oleosin antibodies, the other with anti-hirudin antibodies. As seen in Fig. 3.9, panel A, a band corresponding to a molecular weight of approximately 26kDa was recognized by the anti-hirudin antibody in transformed (TF), but not wild type (WT) oil body proteins. When a membrane with these identical proteins was incubated with anti-oleosin antibodies (Fig. 3.9, panel B), native oleosins, having a molecular weight of 19-24 kDa, were recognized in both plant types. However, an additional band in transformed, but not wild type oil bodies (indicated by the arrow), having an identical molecular weight to the band recognized by the anti-hirudin antibodies (panel A), was recognized by anti-oleosin antibodies.

3.2.2.3 Immunofluorescence localization of hirudin variant 2 on oil bodies

In order to determine if hirudin fused to oleosin is localized on the cytosolic side of the oil body, the following experiment was performed. Oil bodies from OBHIRT-9 were isolated and incubated with anti-hirudin primary antibodies. Antibody binding to oil bodies was detected by addition of FITC-conjugated secondary antibodies and visualization with fluorescence microscopy. As shown in Fig. 3.10, transformed oil bodies (Fig. 3.10C) showed significant fluorescence as a result of this treatment (Fig. 3.10D). To verify that antibody binding was due to the presence of hirudin on the surface of oil bodies, the oil bodies (Fig. 3.10E) were treated with Factor Xa. Oil body fluorescence was significantly reduced after proteolytic treatment (Fig. 3.10F), although low levels of residual fluorescence were observed. This is probably a result of incomplete cleavage of hirudin from its oleosin "carrier." Untransformed oil bodies (Fig. 3.10A) showed no fluorescence when probed with the above primary and secondary antibodies (Fig. 3.10B). No

Figure 3.9 Immunological detection of fusion protein with hirudin and oleosin antibodies. Fifty micrograms each of total seed protein from wild-type (WT) or transformed (TF) seeds was subjected to SDS-PAGE, transferred onto PVDF membrane, and incubated with either anti-hirudin (panel A) or anti-oleosin (panel B) antibodies. The presence of an approximately 26kDa protein in the transformed seed proteins incubated with either antibody is indicated by an arrow. This band is absent in wild type seed protein.

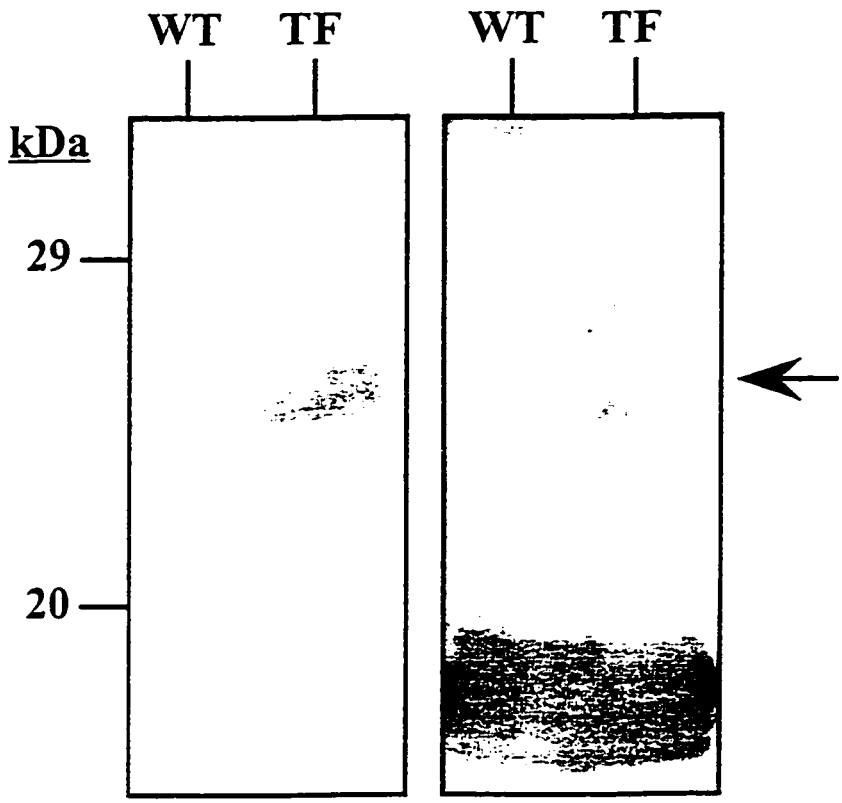
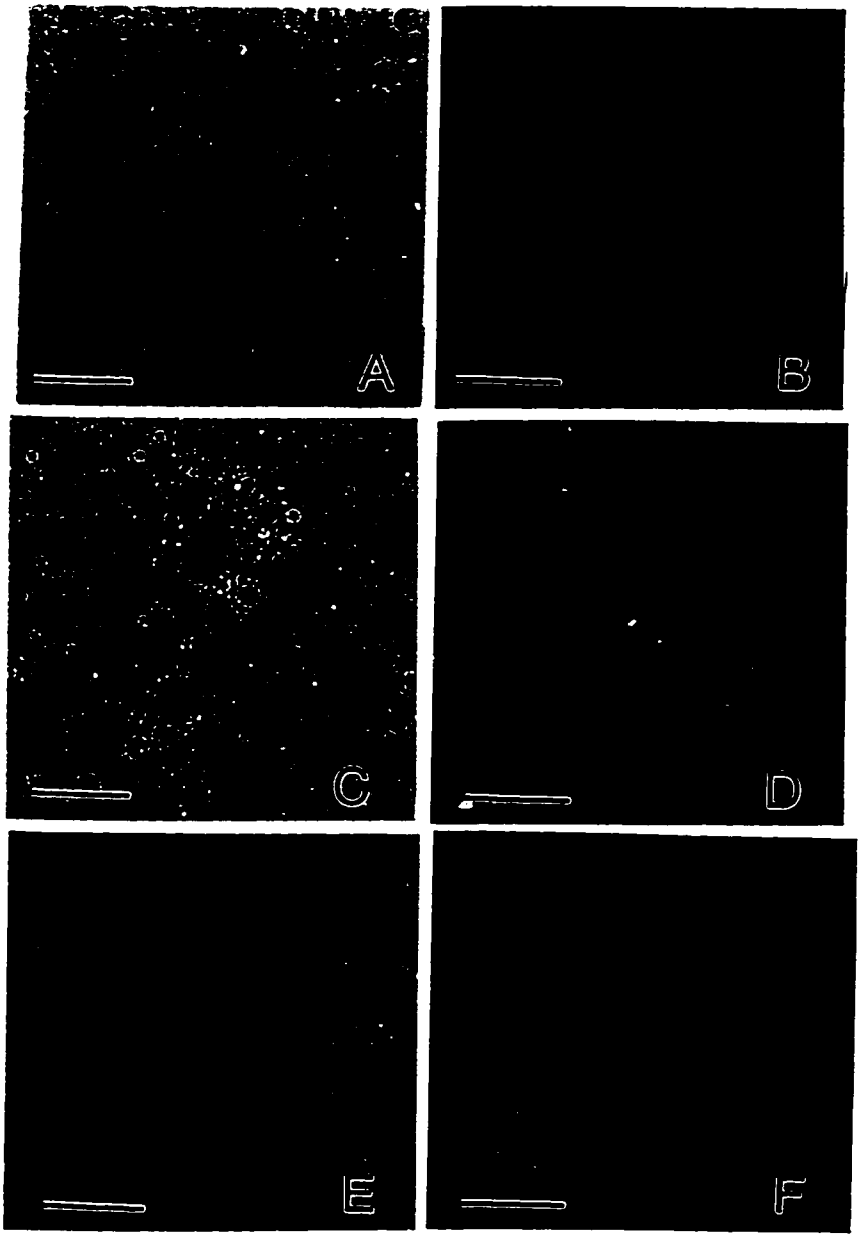

WT
TF
WT
TF
kDa
29
20
A.
B.

Figure 3.10 Immunofluorescent localization of hirudin. Oil bodies form transformed and wild-type seeds were isolated and incubated with anti-hirudin monoclonal antibodies and the appropriate FITC-conjugated secondary antibodies (see 2.2.18). Panels A, C, and E, light microscopy; Panels B, D and F, fluorescence microscopy. Transformed oil bodies were treated (E, F) or untreated (C, D) with Factor Xa. As shown, oil body fluorescence is significantly reduced after Factor Xa treatment. Wild type oil bodies (Panel A) demonstrated no fluorescence (Panel B). Scale bar= 25μm.


A
B
C
D
E
F

fluorescence was observed when transformed oil bodies were probed with mouse serum and secondary antibodies, or secondary antibodies, alone (data not shown).

3.2.2.4 Detection and quantitation of biologically active hirudin variant 2

Hirudin activity is measured using a colorimetric thrombin inhibition assay. This assay measures the rate of thrombin-mediated proteolysis of a chromogenic substrate (p-tosyl-gly-pro-arg-nitroanilide) that, upon cleavage, absorbs light at 405nm. *B. napus* oil bodies from wild type and transformed (OBHIRT-9) seeds were analyzed for anti-thrombin activity. Transformed and wild-type oil bodies were treated in the presence or absence of Factor Xa overnight at room temperature. Thrombin was added to the unternatant isolated from these samples, or to buffer, alone. The thrombin proteolytic activity present in each mixture, as measured by the rate of absorption at 405nm, was determined for each sample. As seen in Fig. 3.11, the activity of the thrombin was unchanged by addition of unternatant from wild type oil bodies (either cleaved with FXa or uncleaved) as compared to the addition of buffer, alone. Significantly, oil body proteins from seeds transformed with the oleosin-hirudin fusion gene also demonstrate no anti-thrombin activity prior to Factor Xa treatment. Anti-thrombin activity in transformed plants only occurred after the hirudin was released from the oleosin-hirudin fusion by specific proteolytic digestion. These results correspond to those obtained from assays involving oil bodies from transformed tobacco (section 3.2.1.2). The hirudin activity released corresponds to approximately 0.55 anti-thrombin units per mg transformed oil body protein cleaved with Factor Xa.

Treatment of oil body proteins with the protease clostripain (an inexpensive enzyme which cleaves after arginine residues) showed similar results as those yielded with Factor Xa cleavage. As shown in Table 3.1, addition of 20, 40, and 60 µl of unternatant from wild type oil bodies treated or untreated with clostripain, or from untreated oil bodies of

Figure 3.11 Measurement of antithrombin activity in transformed and wild-type oil body extracts. Oil bodies from transformed (TF) and wild-type (WT) seeds were treated in the presence (X) or absence of Factor Xa. The unternatant from these samples, or buffer alone, was added to thrombin in buffer. The amount of thrombin activity (in units) remaining after addition of oil body unternatant or buffer was determined as described in sections 2.2.20 and 3.2.1.2. The thrombin activity present in the samples cleaved with Factor Xa have been adjusted to account for the presence of contaminating thrombin present in commercial Factor Xa. As shown, thrombin activity is only reduced when unternatant from Factor Xa-treated, transformed oil bodies (TF, X) is added.

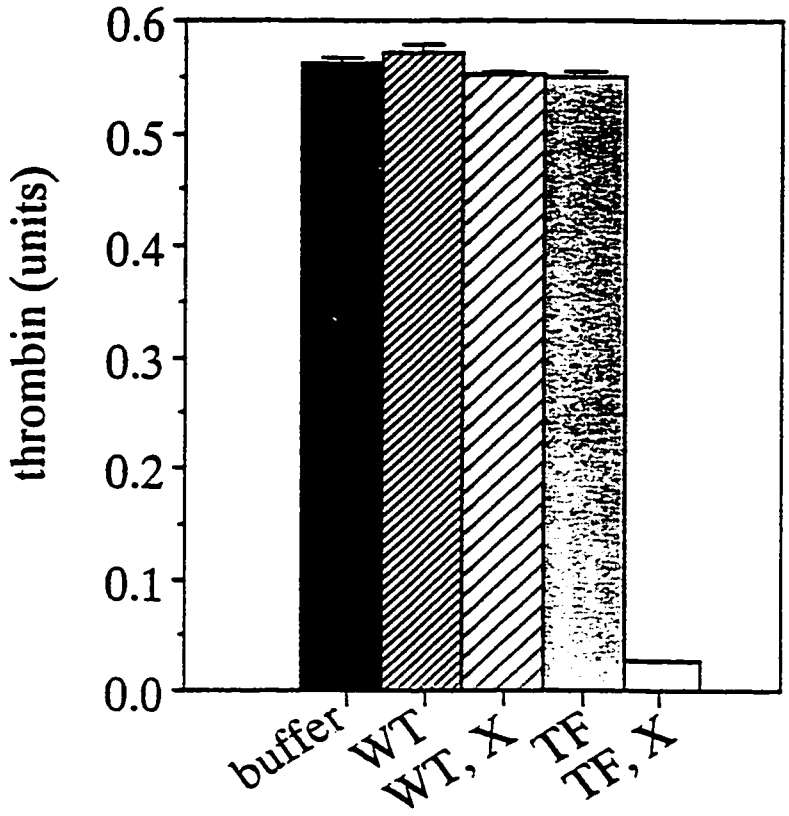

thrombin (units)
0.6
0.5
0.4
0.3
0.2
0.1
0.0
buffer
WT
WT, X
TF
TF, X

Table 3.1 Determination of dose-dependent inhibition of thrombin activity by wild-type and transformed oil body extracts. Oil body proteins from transformed (TF) and wild-type (WT) extracts were digested in the presence or absence of clostripain. Acetone precipitated proteins were resuspended in buffer and added in 20, 40, or 60µl volumes to thrombin in buffer. As a control, 20, 40, or 60µl of buffer, alone, was added to thrombin in buffer. Thrombin activity, in thrombin units, is listed. All assays were performed in triplicate. Individual pairwise treatments were found to be significantly different using analysis of variance (ANOVA) at 99% confidence (Appendix 2). Least significant difference (LSD) tests separated data into statistically similar or different groupings, shown as superscripts. A letter common to each data point indicates no significant difference. Significant anti-thrombin activity was only observed in transformed oil body proteins treated with clostripain. This activity was dose-dependent.

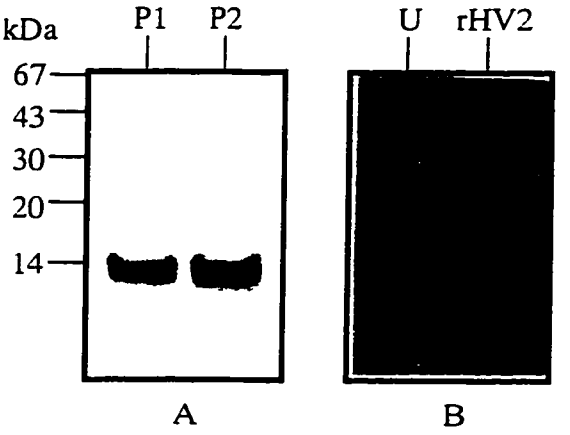
	Thrombin Activity		
	volume of sample added		
sample	20ul	40 ul	60 ul
buffer	0.1368[a]	0.1290[a]	0.1222[a]
WT	0.1348[a]	0.1339[a]	0.1262[a]
WT+ clostripain	0.1378[a]	0.1382[a]	0.1330[a]
TF	0.1402[a]	0.1372[a]	0.1368[a]
TF+ clostripain	0.0814[b]	0.0336[c]	0.0122[d]

transformed seeds had no effect on thrombin activity. However, addition of protein extracts from transformed oil bodies cleaved with clostripain (Table 3.1, TF+clostripain) resulted in a dose-dependent inhibition of thrombin activity. In this assay, approximately 0.2 anti-thrombin units per mg transformed oil body protein cleaved with clostripain was measured. The reduction in recovery of anti-thrombin activity (per mg of oil body protein) obtained from clostripain-treated oil bodies compared to FXa-treated oil bodies is likely due to the nature of protein preparation. Clostripain is only enzymatically active in the presence of 2.5 mM DTT. Optimal anti-thrombin activity of hirudin is dependent on the formation of three intramolecular disulfide bridges (Chang, 1983). As a result, care must be taken in order to remove the DTT from samples incubated with this reducing agent. In addition, complications were encountered due to non-specific cleavage by clostripain. Mass spectroscopy and N-terminal amino acid sequence analysis of hirudin purified from the unternatant of oil bodies treated with clostripain revealed that the recovered hirudin was largely cleaved at the Lys 35 position. These data were determined by Ciba Geigy laboratories, Basel, Switzerland. On account of these technical hurdles, all other cleavage experiments were performed using Factor Xa, despite its relative high cost compared to clostripain.

3.2.2.5 Electrophoretic analysis of recombinant hirudin variant 2

Recombinant hirudin was purified from the unternatant fraction following Factor Xa cleavage of oil bodies from transformed seed. Purification was accomplished through anion exchange chromatography followed by reverse-phase chromatography. After each step, hirudin-containing fractions were identified using the anti-thrombin chromogenic assay. Two overlapping peaks of hirudin activity were obtained following reverse-phase chromatography. Fig. 3.12, panel A, represents a protein gel through which proteins from

Figure 3.12 Electrophoretic and immunoblot analysis of recombinant hirudin. A). *B. napus*- derived recombinant hirudin was purified and analyzed on a 16.5% Tricine-SDS gel. Approximately 3.4μg of peak 1 (P1) and 7.6μg of peak 2 (P2) obtained following reverse-phase chromatography were loaded in individual lanes. Protein was visualized via Coomassie Blue-staining. B). An aliquot of the unternatant fraction (U) from transformed seed corresponding to approximately 80μg of original oil body protein was separated on a 16.5% Tricine-SDS gel along with 3 units of commercial, recombinant hirudin variant 2 (rHV2) (Sigma) derived from yeast. The resulting gel was electroblotted to PVDF-P^{sq} membrane and probed with anti-hirudin antibodies.


kDa
P1
P2
67
43
30
20
14
A
U
rHV2
B

each of these peaks were electrophoresed and stained with Coomassie blue. Both peaks were comprised largely of a single molecular mass species indicating a relatively high degree of purity. From this protein gel, the apparent molecular mass of plant-derived hirudin is approximately 14 kDa, which is significantly higher than its predicted mass of 6893 Da. However, as shown in Fig. 3.12, panel B, in which plant-derived hirudin from Factor Xa-treated oil body unternatant is immunoblotted together with commercially available, yeast-derived hirudin, the electrophoretic mobility for both *B. napus* (U)- and yeast-derived recombinant hirudin variant 2 (rHV2) are identical to each other, and to the mobility observed on the protein gel (Fig. 3.12, panel A). These results suggest that the difference between apparent and actual molecular mass is due to the aberrant mobility of hirudin through SDS gels. Although the two overlapping peaks recovered chromatographically have similar anti-thrombin activity (measured via colorimetric assays, data not shown), amino acid sequence analysis (performed at Plant Biotechnology Institute, Saskatoon) indicated that the difference between these two peaks was due to a C-terminal truncation after the 63 amino acid (tyrosine) (see Fig. 3.1 for amino acid sequence of HV2). It is unclear whether this truncation is due to premature termination during fusion protein translation, or if it occurs during oil body isolation or as a result of proteolytic treatment of the oil bodies.

3.2.2.6 Stability of oleosin-hirudin variant 2 fusion protein in seeds

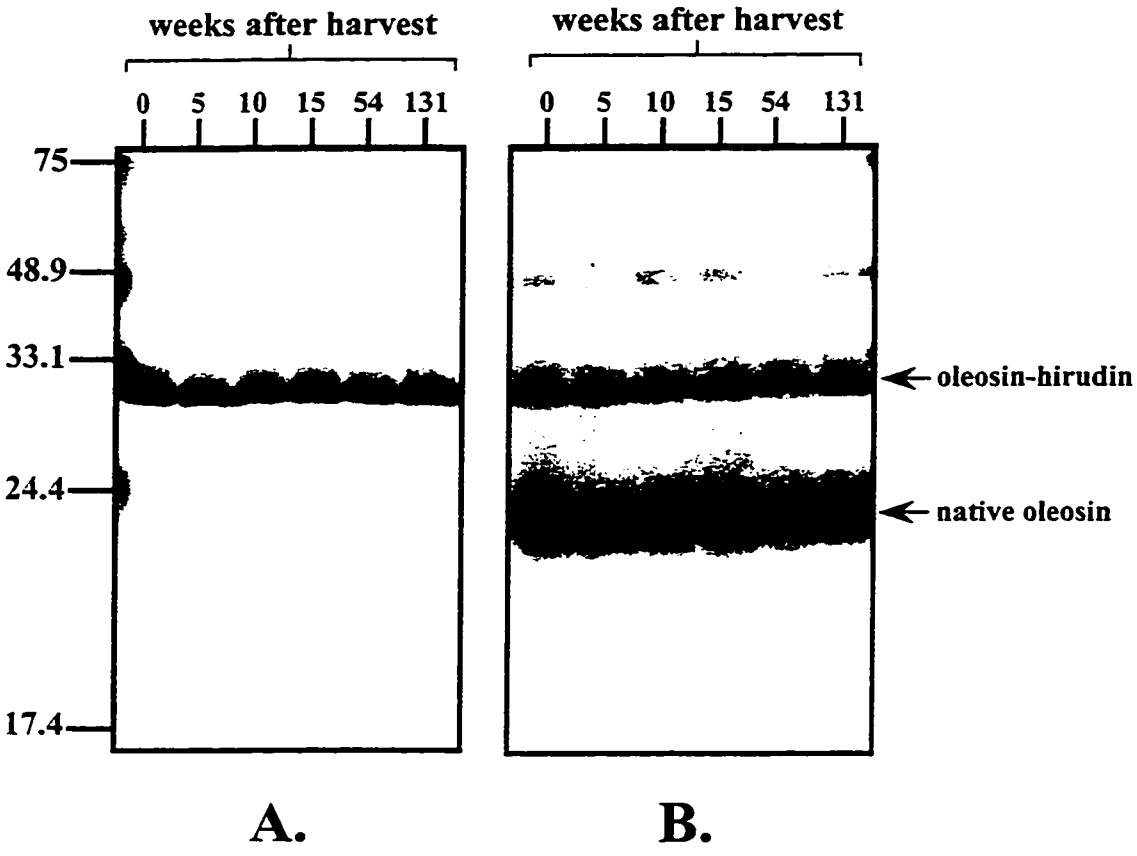
Dry seeds from second generation, self-pollinated, *B. napus* transformed with the oleosin-hirudin variant 2 fusion gene (OBHIRT-4) were harvested and stored for up to 156 weeks. At week 0 (day of harvest), 5, 10, 15, 54, and 131 weeks after harvest, 0.2 g of seed were ground with a pestle and mortar in 2 ml of 50mM Tris-HCl (pH 7.5) and frozen at -80^{0}C. Once all samples were collected, this total seed protein was loaded, 5μg/lane, in

duplicate, on a large 15% polyacrylamide gel and electrophoresed. Also elecrophoresed was 3, 4, 5, 6 and 7μg/lane of protein collected at week 0, for use as a protein standard. After electrophoresis, the proteins were transferred onto 2 PVDF membranes. As shown in Fig. 3.13, panel A (upper arrow), a band corresponding to the expected molecular weight (26kDa) of the oleosin-hirudin fusion protein is detected by the anti-hirudin antibodies at approximately equal levels in all samples tested. This indicates that the fusion protein is stable in dry seeds over long periods of storage. To verify that this is an oleosin-hirudin fusion protein, an identical blot was incubated with anti-oleosin antibodies (Fig. 3.13, panel B). As shown, the anti-19kDa *B. napus* oleosin antibodies detected this 26kDa band, in addition to native oleosins (lower arrow). The level of intensity of the oleosin-hirudin fusion protein detected by anti-hirudin antibodies fell within a linear range, as indicated by a protein standard (2-7μg/lane of week 0 protein) (data not shown). There appears to be slightly higher levels of the oleosin-hirudin fusion protein in samples taken at week 0 and week 131 (panel A). As indicated by the protein standards also immunoblotted, this level of intensity is equivalent to that found in approximately 6μg of total seed protein. Similar increases in band intensity were also noted in the blot incubated with anti-oleosin antibodies, and can thus be attributed to sampling error.

3.2.2.7 Field comparison of wild type and transgenic Brassica napus

Second and third generation *B. napus* seeds from plants homozygous for the oleosin-hirudin variant 2 fusion gene were planted under field conditions in 1994-1996. The agronomic traits of the transgenic plants were compared to wild type *B. napus* grown under identical conditions. Dr. Henning Muendel (Agriculture and Agri-Food Canada Research Centre, Lethbridge) and Joanna Pinto (University of Calgary) managed these

Figure 3.13 Stability of the oleosin-hirudin variant 2 fusion protein in dry seeds. Dry seeds from plants transformed with pCGN-OBHIRT were stored at room temperature over a period of 131 weeks. Total seed protein was isolated from these seeds at 0, 5, 10, 15, 54, and 131 weeks after harvest and frozen at -80^{0}C. Five micrograms of total seed protein from each sample was loaded, in duplicate, onto a 15% polyacrylamide gel. The proteins were electrophoresed, transferred onto two PVDF membranes, and incubated with either anti-hirudin (panel A) or anti-oleosin (panel B) antibodies. A band corresponding to the predicted molecular weight of the oleosin-hirudin fusion protein was detected by both antibodies (upper arrow). The level of intensity of this band, with the exception of minor sampling error, is approximately the same. No degredation products were detected by either antibody.


weeks after harvest
0 5 10 15 54 131
75
48.9
33.1
24.4
17.4
A.
weeks after harvest
0 5 10 15 54 131
oleosin-hirudin
native oleosin
B.

field trial studies and subsequent seed analysis. There were no significant difference between the wild type and transgenic lines, with regards to stand count (plants/m^2), seed yield (kg/hectare), percent oil content, or percent germination.

3.3 Discussion

In this chapter, the ability to use oleosins as "carriers" of heterologous proteins of interest, such as hirudin variant 2, was investigated. A gene construct, pCGN-OBHIRT (Fig. 2.1, 3.1), was introduced *B. napus*. This fusion gene construct consists of an *Arabidopsis* oleosin promoter and coding region fused, in frame, to a synthetic hirudin variant 2 cDNA. Oleosins make up a large component of the total seed protein in *B. napus* seeds (approximately 10%) (Murphy and Cummins, 1989). It was reasoned that the high level of accumulation of this native protein species in *B. napus* would also afford a high degree of expression and accumulation of the fusion protein in this plant. However, at the time of these experiments, it was unknown if high levels of expression of the heterologous protein would result in deleterious effects in the development of the transgenic plant or seed. Therefore, an easily transformed plant species, such as tobacco, which expresses a modest amount of oleosin proteins, was also transformed with this fusion gene construct. The expression, synthesis, and targeting of a *B. napus* oleosin gene in tobacco has been previously demonstrated (Batchelder *et al.*, 1994). Given the degree of similarity between *Arabidopsis* oleosin promoter and coding region used in this investigation to the gene encoding the predominant *B. napus* oleosin, it was expected that the fusion gene construct would also be expressed in transgenic tobacco.

Four transgenic tobacco plants harbouring the pCGN-OBHIRT fusion gene construct were analyzed for the accumulation of the heterologous protein. As shown in Fig. 3.2, a band corresponding to the expected molecular weight of the fusion protein

(approximately 26kDa) was detected by anti-hirudin antibodies in transformed, but not wild-type total seed extracts. Lower molecular weight bands were also observed on this western blot in both wild type and transformed seed extracts. These bands have the same electrophoretic mobility as the predominant seed proteins of tobacco, and are likely due to cross-reactivity of the antibody. These bands are not observed when these proteins are removed during oil body preparation, as shown in Fig. 3.3, panel A. As shown in this Figure, an approximately 26kDa band is recognized in transformed oil body extracts by antibodies raised against both hirudin and oleosin protein. In addition to confirming that this band is an oleosin-hirudin fusion protein, these data also demonstrate that this fusion protein is associated with the oil body fraction of transgenic tobacco.

The N-terminal portion of mature hirudin is essential for binding and blocking the active site of thrombin (Betz *et al.*, 1992). Studies in which hirudin has been expressed in heterologous systems have demonstrated that foreign amino acid sequences, such as short leader peptides, present at the N-terminus of hirudin inhibit biological activity, presumably by blocking hirudin binding to the thrombin active cleft (Fortkamp *et al.*, 1986; Loison *et al.*, 1988). In these reports, active hirudin was obtained only by specific proteolytic processing of the signal sequence. These results suggest that plant-derived hirudin would only be active after specific proteolytic cleavage from its oleosin "carrier". Thus, a sequence recognized by the protease, Factor Xa, was added between the coding regions of oleosin and hirudin in the fusion gene construct in order to facilitate not only the release of hirudin from the fusion protein, but to afford its ability to bind thrombin. Factor Xa cleaves after arginine in the recognition sequence, ile-glu-gly-arg (Fig. 3.1).

The results presented in Fig. 3.4 demonstrate that hirudin activity, as measured by its ability to inhibit thrombin, is not detected in wild-type oil body "unternatant" which has been treated or untreated with Factor Xa. The importance of accurate proteolytic release of

hirudin from its oleosin carrier is demonstrated by the fact that anti-thrombin activity is only detected in transformed unternatant which has been subject to specific endoproteolytic digestion with Factor Xa. Recovery of active hirudin following proteolytic treatment of the transformed oil body fraction indicates that proper folding of hirudin is able to occur and, furthermore, that the protease recognition site is accessible while the fusion protein remains attached to the oil body. The pathway of hirudin folding has been studied extensively (Chang, 1983; Chatrenet and Chang, 1993) and has revealed that correct formation of the three disulfide bridges present in the native protein is necessary for full activity. Furthermore, it has been shown that hirudin folding occurs spontaneously *in vitro* via a near-random mechanism of disulfide bridge formation, followed by reorganization, to produce the final, biologically active product (Chatrenet and Chang, 1992, 1993). Thus, it is not surprising that hirudin produced in plants, in addition to that secreted by yeast (Loison *et al.*, 1988) and baculovirus-infected insect cells (Benatti *et al.*, 1991) does not required any denaturation/renaturation with such compounds as reduced/oxidized glutathione or protein disulfide isomerase in order for activity to be observed. Although hirudin released from its oleosin carrier is active, it is not clear whether hirudin folds *in vivo* or if this occurs during oil body preparation. However, recent studies have shown that the C-terminus of oleosin, and therefore hirudin, does not enter the lumen of the endoplasmic reticulum (Abel *et al.*, 1997, van Rooijen and Moloney, 1995a), where protein folding is known to occur.

In summary, the oleosin-hirudin fusion gene is expressed in transgenic tobacco. The fusion protein accumulates in transgenic seed, and is associated with the oil body protein fraction. Biologically active hirudin is detected in the oil body unternatant of transformed seed extracts treated with Factor Xa, demonstrating that the proteolytic cleavage site is accessible to the protease while associated with oil bodies and that hirudin is

capable of correct disulfide bond formation necessary for thrombin inhibition. This is the first report of the successful expression, accumulation, and isolation of biologically active hirudin variant 2 in tobacco.

In order for heterologous protein expression to be economically viable, high levels of the foreign proteins must accumulate in a manner which can be easily, and cheaply purified. The successful expression and accumulation of the oleosin-hirudin fusion protein in tobacco implies that this fusion protein could also be successfully expressed in other plants which accumulate oleosins. The use of *B. napus* for the expression of this fusion protein is an obvious choice, due to the fact that oleosins represent a large component of the total seed protein (Murphy and Cummins, 1989), and *B. napus* is a commercially important crop plant for which the infrastructure required for its harvest is readily available. As shown in Fig. 3.7, isolation of the oil bodies by flotation centrifugation results in a marked enrichment of oleosins. Our current estimates suggest that greater than 99% of contaminating seed proteins are removed following the simple two stage differential density centrifugation. The processing of canola for the purpose of purifying oil body extracts in the manner shown here differs from that used for standard oilseed milling. In the latter scenario, oil is extracted using a combination of temperature, heat, and pressure (Eskin *et al.*, 1996). For our purposes, a procedure which reflects more of the type used to wet-mill corn would be more appropriate (Carter *et al.*, 1974).

As indicated by northern blot analysis (Figs. 3.5 and 3.6), the fusion gene construct is expressed in a seed-specific manner in transgenic *B. napus*. The seed-specific regulation of expression by the 0.8kbp *Arabidopsis* oleosin promoter used to drive the fusion construct is in agreement with other studies using this promoter (Plant *et al.*, 1994). As demonstrated in transgenic tobacco, the oleosin-hirudin fusion protein also accumulated in transgenic *B. napus* and correctly targeted to the oil body fraction of transgenic seeds

(Fig. 3.8 and 3.9). Thus, it is apparent that oleosins can tolerate considerable modification at its C-terminus without affecting its targeting. This observation is in agreement with a recent study, in which fusion of β-glucuronidase gene to the 3' end of an *Arabidopsis* oleosin promoter and coding region was shown to correctly target to the oil body membrane (van Rooijen and Moloney, 1995b). The ability to express a wide range of proteins as oleosin fusions has significant implications. For example, Factor Xa, is an expensive protease which would not be suitable for use in large-scale processes. However, it is possible that oleosin fusions carrying active and specific proteases may be constructed. By mixing these oil bodies with oil bodies harboring oleosins fused to proteins of interest, it may be possible to overcome the need for commercial proteases, resulting in a significant reduction in the cost associated with protein isolation. This possibility is significant, as it would limit the number of protein species present in the unternatant of proteolytically treated oil bodies. For example, Factor Xa added to oil bodies remain in the unternatant fraction and thus, represents a contaminating protein. This is particularly problematic as it is a blood protease involved in the conversion of prothombin into thrombin.

Biologically active hirudin was detected in the unternatant of transgenic oil bodies treated with Factor Xa (Fig. 3.11). Furthermore, the specific activity of reverse-phase HPLC-purified hirudin variant 2 yielded from *B. napus* transformed with the pCGN-OBHIRT construct, was approximately 14,000 anti-thrombin units/mg protein. This specific activity is in agreement with that published in previous studies (Loison *et al.*, 1988).

Field trials of transgenic *B. napus* expressing the oleosin-hirudin fusion gene showed no significant difference in agronomic qualities compared to wild-type seed grown under identical conditions. The lack of anti-thrombin activity in oil body unternatant from transgenic seed undigested with Factor Xa (Figs. 3.4 and 3.11, Table 3.1) has practical

implications for field production of these recombinant proteins, as they would remain inactive (and therefore environmentally benign) until specific proteolytic release.

The cost associated with the purification of recombinant proteins is a significant factor in determining the feasibility of commercial production. The substantial level of purification achieved through flotation-separation of oil bodies demonstrates the utility of this technique for the production of recombinant proteins in plants. This level of purity is comparable with that of bacterial (Bender *et al.*, 1990; Dodt *et al.*, 1986) and yeast (Riehl-Bellon *et al.*, 1989) secretion systems. As a result, the further purification of hirudin to near-homogeneity was accomplished in only two chromatographic steps using standard methodology. Production of recombinant proteins in seeds also offers several distinct advantages over synthesis in bacteria and yeast. The low cost of seed-based expression systems contrasts with the expensive and labour-intensive operation of large scale fermentation systems. There was no detectable decrease in fusion protein accumulation between seeds stored for over two years versus those newly harvested (Fig. 3.13), indicative of the stability of seed proteins during storage. This means that there is no need to synchronize field-production with demand for a given protein. Provided that expression levels are adequate and purification costs reasonable, seed-based expression systems should be very cost-efficient.

In conclusion, a system for the general production and recovery of recombinant peptides synthesized as fusions to seed oil body proteins has been described. The fidelity of this system has been demonstrated with the production of a functional pharmaceutical protein, hirudin. The system is flexible with respect to the different types of proteins it can accommodate and enables a rapid and simple purification of the recombinant product. Furthermore, low costs of seed production and compatibility with existing agricultural

processing procedures make it an attractive alternative to conventional bacterial and yeast fermentation systems.

Chapter 4:

Expression of oleosin-hirudin variant 1 fusion gene in *Brassica napus*

4.1 Introduction

In most examples of heterologous gene expression, it is necessary to optimize expression levels if the system is to have practical value. One method by which this may be achieved in plants is through the employment of viral translational enhancers (rev. in Turner and Foster, 1995). A growing number of enhancers constituting the 5' untranslated region (5'UTR) of plant virus genomic or subgenomic RNA have been identified. These include the 5'UTR of brome mosaic virus RNA3, potato virus X, potato virus Y, tobacco mosaic virus (omega), turnip mosaic potyvirus, tobacco etch virus, plum pox potyvirus, and alfalfa mosaic virus RNA4. There appears to be no consensus sequence between these leader sequences, although they are characterized by an absence of secondary structure (rev. in Turner and Foster, 1995). Although the molecular mechanisms by which translational enhancement is achieved remains unclear, the lack of secondary structure in many of these leader sequences is believed to enhance the efficiency of ribosomal scanning for the AUG translation initiation codon. Furthermore, it is believed that AMV-mRNAs have a higher affinity or reduced requirement for limiting components of the translational machinery, such as the cap-binding protein complex (Jobling and Gehrke, 1987).

The first reported plant viral leader sequence which functioned as a translational enhancer was that of the alfalfa mosaic virus RNA4 (AMV) (Jobling and Gehrke, 1987). These authors demonstrated that the AMV leader enhanced translational efficiency of chimeric AMV- barley α-amylase by as much as 35-fold and AMV-human interleukin 1β by 6-7 fold. The AMV leader has also been shown to enhance expression of GUS by 6-fold in *E. coli* and 8-fold in tobacco protoplasts (Gallie *et al.*, 1987). Studies in transgenic tobacco have reported an 8-fold increase in protein accumulation (Datla *et al.*, 1993). This increase in protein accumulation was shown to occur independent of transcript abundance.

In an effort to enhance the expression and accumulation of the oleosin-hirudin fusion gene/protein in transgenic *B. napus,* two new constructs were created which contained modifications to the original OBHIRT construct (section 2.3.1.1). These included the use of a stronger 2.5kbp *Arabidopsis* oleosin promoter (Plant *et al.*, 1994), and the addition of an AMV RNA4 leader sequence to the 5'end of the oleosin-hirudin fusion gene. In this chapter, the level of protein accumulation and the anti-thrombin activity detectable in Factor Xa-treated oil body extracts will be compared between seeds harbouring these new constructs, and that of the original transgenic seeds harbouring the OBHIRT construct.

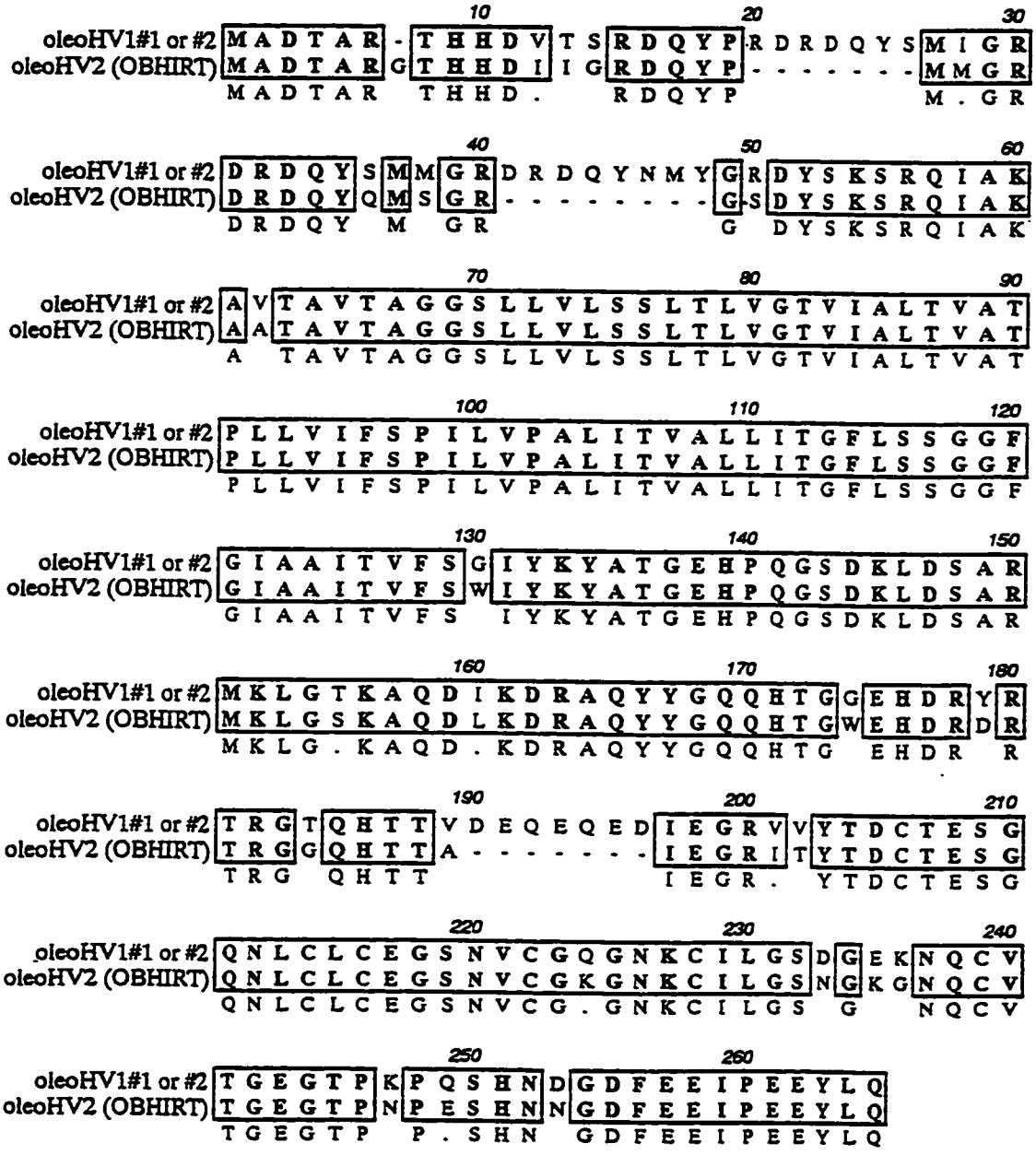
4.2 Results

4.2.1 Deduced amino acid sequence and 5' untranslated region of oleosin-hirudin variant 1 fusion gene

The construction of pCGN-oleoHV1#1 and 2, representing oleosin-hirudin variant 1 fusion proteins with, or without an AMV leader sequence, respectively, is described in sections 2.3.1.2 and 2.3.1.3, respectively. The amino acid sequence of the coding regions of the oleoHV1#1 and oleoHV1#2 fusion gene are identical. OleoHV1#1 and oleoHV1#2 are also sometimes referred to as "+AMV" and "-AMV", respectively. This sequence is shown in Fig. 4.1 (top strand) and is compared to the predicted amino acid sequence of the oleosin-hirudin variant 2 fusion gene construct, pCGN-OBHIRT (oleoHV2) (lower strand), described in section 2.3.1.1. The coding region of a 19 kDa *Brassica napus* oleosin (188 amino acids) was used in the construction of oleoHV1. At the C-terminus of this oleosin lies a hydrophilic, negatively charged stretch of 8 amino acids. This "linker" was added in an attempt to enhance electrostatic repulsion between the surface of the oil body, which is negatively charged, and the region surrounding the FXa cleavage site might thereby facilitate more efficient, site-specific proteolysis. The "linker" is followed by 4

Figure 4.1 Comparison of the deduced amino acid sequence encoded by the pCGN-OBHIRT (oleoHV2) (section 2.3.1.1) and oleosin-hirudin variant 1 (oleoHV1) fusion genes (section 2.3.1.2 and 2.3.1.3) introduced into *B. napus* via *Agrobacterium*-mediated transformation. Boxes indicate amino acids which are identical or similar. Identical amino acids are denoted in bold print. A cDNA sequence encoding a 188 amino acid *B. napus* oleosin was used in the construction of oleoHV1. A genomic *Arabidopsis* oleosin, encoding 173 amino acids, was used in the construction of oleoHV2. Four amino acids, representing the Factor Xa/clostripain recognition site (amino acids 197-200), are followed by either hirudin variant 1 or 2.

ClustalW Formatted Alignments



amino acids, (ile-glu-gly-arg), which represents a Factor Xa recognition site, and, finally, hirudin variant 1 (amino acids 201-265). The oleosin-hirudin variant 1 fusion protein (oleoHV1#1 or #2) has an expected molecular weight of 30 kDa. This is higher than the predicted molecular mass of the protein encoded by pCGN-OBHIRT (oleoHV2), which is approximately 26 kDa, due to the use of a higher molecular weight oleosin (a *B. napus* oleosin cDNA was used in the construction of oleosin-hirudin variant 1 as opposed to a genomic *Arabidopsis* oleosin used in the construction of OBHIRT). The oleoHV1 fusion gene (either possessing or lacking the AMV untranslated leader sequence) is under the control of a 2.5 kbp *Arabidopsis* oleosin promoter and is terminated by a 267bp nopaline synthase transcriptional terminator.

Figure 4.2 depicts the aligned DNA sequences of the proximal upstream region of constructs oleoHV1 #1 and #2, and pCGN-OBHIRT. The 36bp AMV leader sequence lies immediately 5' to the start codon of oleoHV1#1 (nucleotides 42-77). The 3' end of the promoter of OBHIRT (nucleotides 31-37) differs from that of oleoHV1 #1 and #2. This is because the genomic clone of the *Arabidopsis* oleosin promoter and coding region used in the construction of pCGN-OBHIRT was modified only at its 5' and the 3' ends (i.e. the sequences between the upstream and coding regions was unaltered). Alternatively, the 3' end of the 2.5kbp promoter used in the construction of oleoHV1#1 and #2 was modified to contain Sma 1 and BamH1 restriction sites (nucleotides 31-41). The start codon begins at base 78.

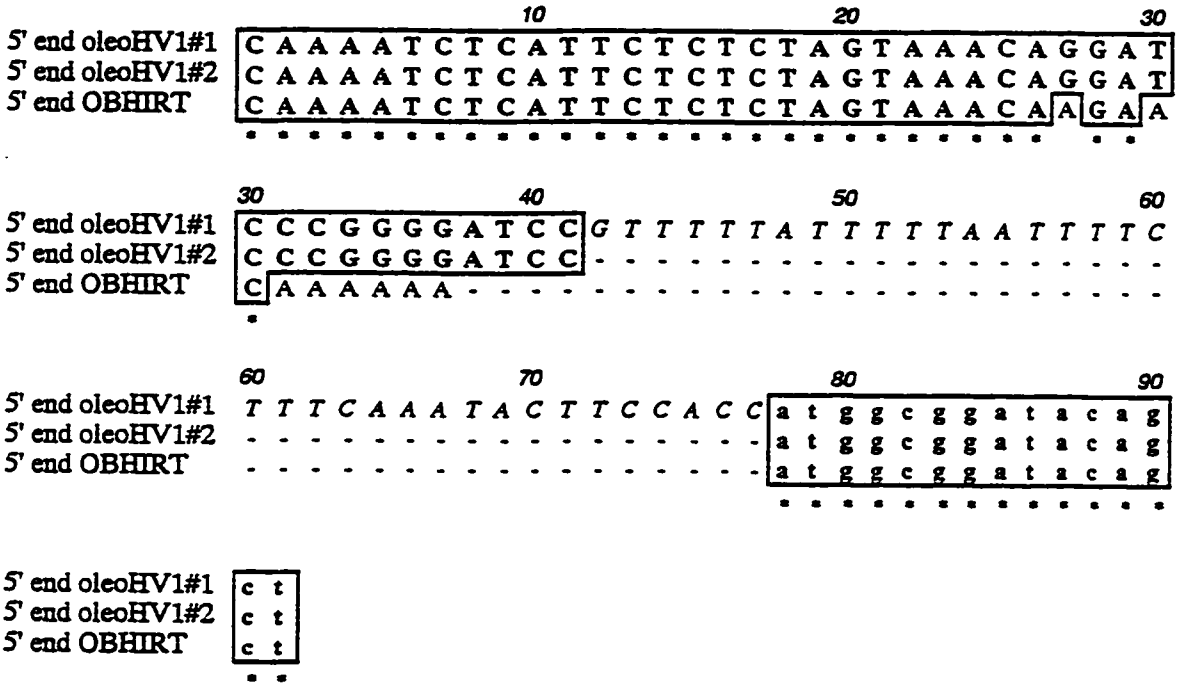
4.2.2 Accumulation of oleosin-hirudin variant 1 fusion protein in *B. napus*

4.2.2.1 Western blot analysis

The oleosin-hirudin variant 1 fusion gene constructs, oleoHV1#1 and #2 were introduced, along with a gene conferring kanamycin-resistance, into the *B. napus* genome as described in Chapter 2 (section 2.2.7). Southern blot analysis of kanamycin-resistant

Figure 4.2 Sequence alignment of the DNA sequences proximal to the start codon of the fusion gene constructs OBHIRT (section 2.3.1.1), oleoHV1#1 (section 2.3.1.2) and oleoHV1#2 (section 2.3.1.3). Uppercase letters represents the 5' upstream region proximal to the promoter, italicized type represents the AMV leader sequence, and lowercase letters represent the coding region of each construct. Boxed areas and bold print indicates bases which are identical. As shown, the only difference between the oleoHV1#1 and #2 constructs is the presence, or absence, of the AMV leader sequence, respectively.

ClustalW Formatted Alignments



plants identified a number of transformants which had incorporated between 1-3 copies of oleoHV1 #1 or #2 into their genome (Figs. 4.3, 4.4). In order to determine the ratio of seeds derived from transgenic plants which contained no insert (nulls) *vs.* those possessing the kanamycin-resistance gene (and, therefore, oleosin-hirudin fusion gene), segregation analysis was performed. Sixty to one hundred seeds from wild type and transformed plants were surface sterilized with bleach (section 2.2.7) and plated on BN1 containing 50mg/L kanamycin. The seeds were germinated and grown for approximately one month. Seedlings were scored for kanamycin-resistance using three different criteria. As shown in Fig. 4.5, panel A , wild type (WT) seeds grown in the presence of kanamycin (+ Kan) exhibit chlorotic true leaves and stems, and reduced secondary root growth. This is in contrast to WT seeds grown in the absence of kanamycin (-Kan). First generation (i.e. heterozygous), transformed (TF) seedlings containing 1 copy of the oleoHV1#1 fusion gene construct (as determined by Southern blot analysis) exhibit a 3:1 ratio of kanamycin resistance *vs.* susceptibility (Fig. 4.5, panel B). First and second generation seed from a number of different oleoHV1#1 and #2 plant lines were scored for segregation of kanamycin-resistance using this method. Comparison of the results obtained from segregation analysis (described above), with those obtained from Southern blot analysis demonstrated that, with one exception, the number of gene copies inserted into the plant genome corresponded with the ratio of kanamycin susceptibility *vs.* resistance expected for plants containing gene inserts which undergo independent assortment. For example, plants containing two copies of the oleoHV1 gene at different loci demonstrated a segregation pattern of kanamycin resistance: susceptibility of 15:1. This is the expected Mendalian ratio for a dominant gene undergoing independent assortment. The gene conferring kanamycin resistance is a dominant gene (Horsch, 1985). Plant #27, which harbours two copies of the oleoHV1#1 fusion gene (Fig. 4.3), however, demonstrated a 3:1 segregation pattern for

Figure 4.3 Southern blot analysis of 50μg of DNA from wild type (WT) *B. napus* or plants transformed with oleoHV1#1 fusion gene construct. The DNA was restricted with Hind III, electrophoresed and blotted as described in section 2.2.12, and the membrane probed with ^{32}P-labeled hirudin cDNA (1.75 X 10^4 becquerels/ml hybridization buffer).

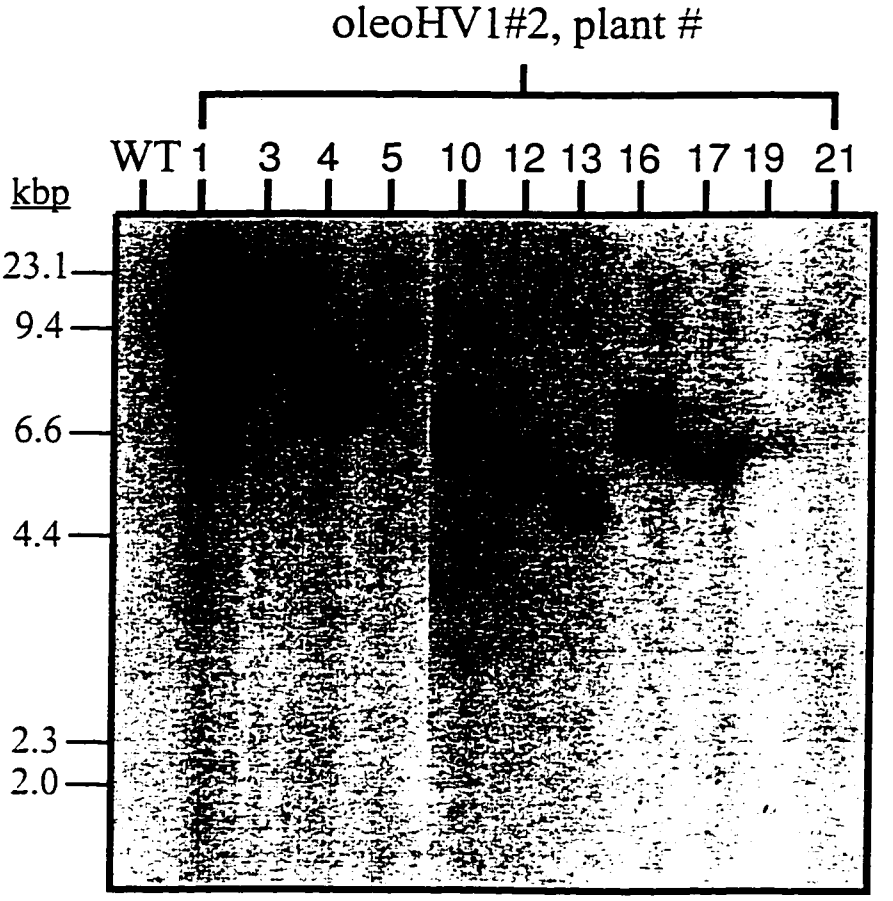

oleoHV1#2, plant #
WT 1 3 4 5 10 12 13 16 17 19 21
kbp
23.1
9.4
6.6
4.4
2.3
2.0

Figure 4.4 Southern blot analysis of 50μg of DNA from plants transformed with oleoHV1#2 fusion gene construct. The DNA was restricted with Hind III, electrophoresed and blotted as described in section 2.2.12, and the membrane probed with ^{32}P-labeled hirudin cDNA (1.75 X 10^4 becquerels/ml hybridization buffer).

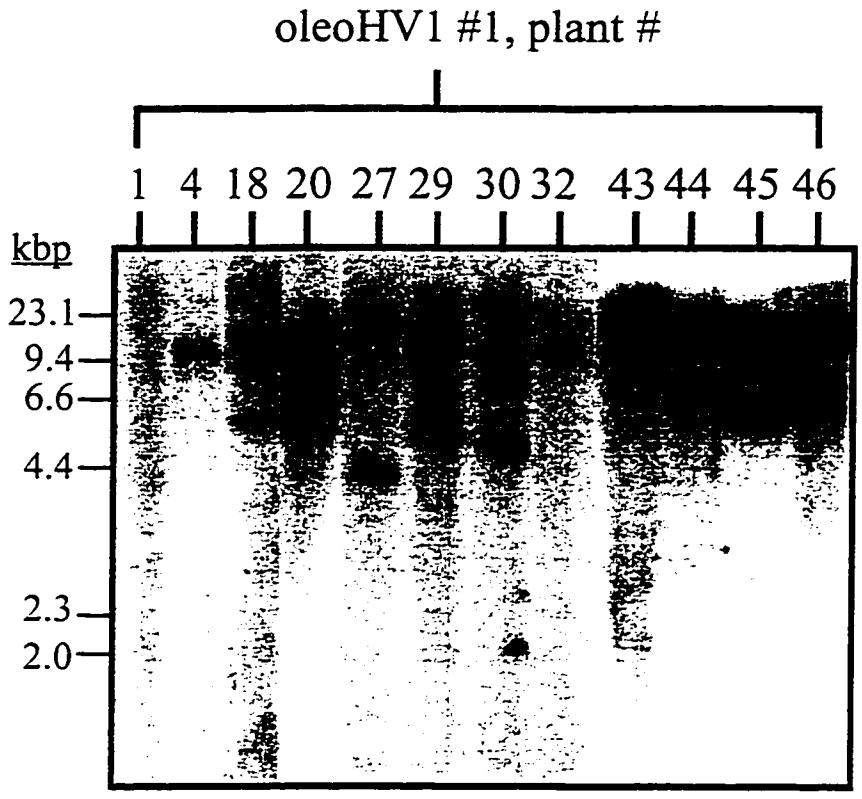

oleoHV1 #1, plant #
1 4 18 20 27 29 30 32 43 44 45 46
kbp
23.1
9.4
6.6
4.4
2.3
2.0

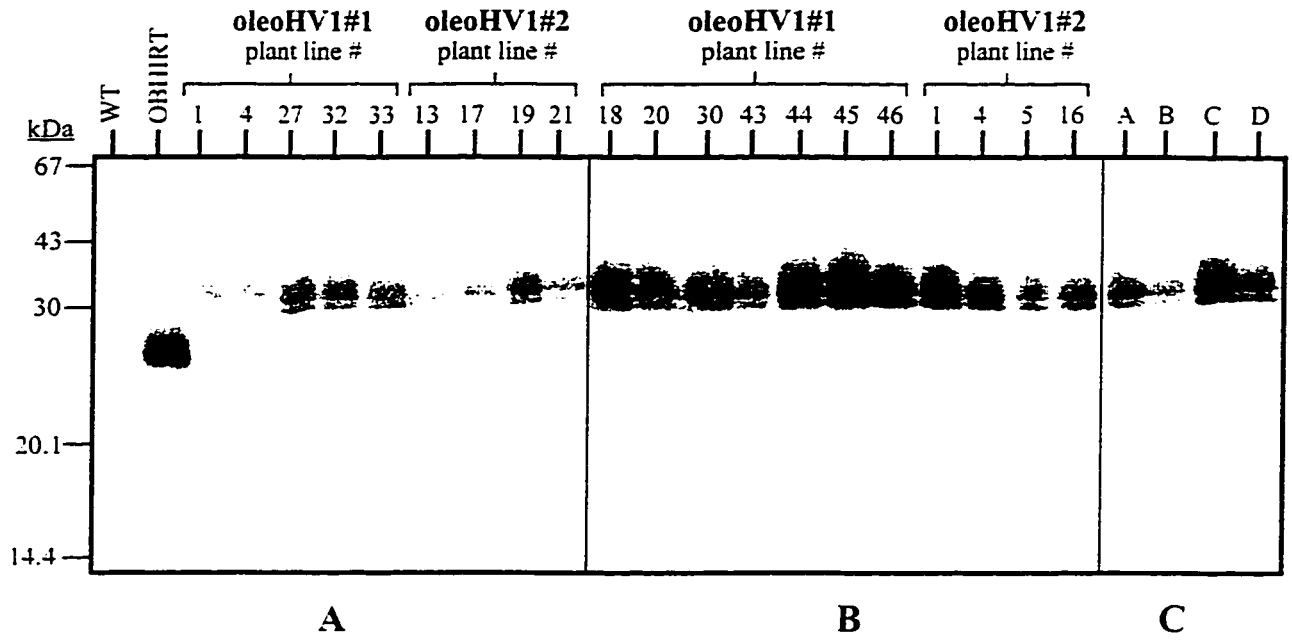
Figure 4.5 Determination of kanamycin resistance in *B. napus* seedlings. Wild-type (WT) seeds (panel A) and seeds derived from plants harbouring 1 insert of a gene conferring kanamycin-resistance (panel B) were germinated and grown for one month in sterile Murashige and Skoog media (section 2.2.7) containing, or lacking, 50mg/L kanamycin. In panel A, wild-type seeds germinated and grown in the absence of kanamycin (left plate) show abundant secondary root growth, and green first leaves and stems. In the presence of 50mg/L kanamycin, (panel A, right plate), little, if any secondary root growth occurs, and the stems and first leaves are chlorotic. In panel B, heterozygous plants harbouring 1 copy of a gene conferring kanamycin resistance exhibit the expected 3:1 ratio of resistance *vs.* susceptibility to the antibiotic.


A
WT, - Kan
WT, + Kan
B
TF
+ Kan
1 insert

kanamycin resitance *vs.* susceptibility, indicating that this plant harbours both copies of the construct at the same locus which do not undergo independent assortment.

Western blot analysis comparing the levels of oleosin-hirudin variant 1 fusion protein produced in seeds demonstrating a 3:1 *vs.* ≥ 15:1 ratio of kanamycin resistance:susceptibility was performed. As shown in Fig. 4.6, 5μg of oil body proteins prepared from wild type (WT) *B. napus* seed and seed from *B. napus* transformed with the oleoHV1#1, oleoHV1#2, or pCGN-OBHIRT (OBHIRT-9; section 3.2.2.2) fusion gene constructs were electrophoresed through two 15% polyacrylamide gels. Seed from different plants harbouring a particular construct are denoted by the construct they contain and the number that was allocated to each plant, i.e. oleoHV1#1, plant #1. The proteins were transferred onto PVDF membranes and incubated, using identical conditions, with mouse monoclonal antibodies raised against hirudin. With the exception of oleoHV1#1, plant #27 (explained above), the oil body protein samples of panel A, Fig. 4.6 are derived from seeds exhibiting a 3:1 ratio of kanamycin-resistance *vs.* susceptibility and contain one copy of the fusion gene construct. Panel B represents seeds which demonstrate a ≥15:1 ratio of resistance:susceptibility and possess ≥ 2 copies of the fusion gene construct. Panel C represents pooled oil body extracts from these sample types. Pooled samples were used in order to get an average expression level from plants representing different transformation events. Western blot analysis of oil body protein from 5, and 4 different plants expressing the oleoHV1#1 and #2 fusion gene construct, respectively, are shown in panel A, Fig. 4.6. As shown, the overall level of oleosin-hirudin variant 1 fusion protein detected by antibodies raised against hirudin is higher in seeds possessing the oleoHV1#1 *vs.* oleoHV1#2 fusion gene construct. Because the only difference between these two constructs lies in the presence (or absence) of the AMV leader sequence, these results suggest that the AMV leader sequence may provide translational enhancement, resulting in

Figure 4.6 Immunological detection of oleosin-hirudin variant 1 from oil body extracts derived from first-generation, transgenic *B. napus* harbouring the transgenes oleoHV1#1 and oleoHV1#2 (section 2.3.1.2, and 2.3.1.3, respectively). Five micrograms each of oil body protein from wild-type (WT) seed, or seed harboring the pCGN-OBHIRT, oleoHV1#1 or oleoHV1#2 constructs were subjected to SDS-PAGE, transferred onto PVDF membrane, and incubated with anti-hirudin antibodies. Panel A and B represent seed lines in which one, or multiple copies of the transgenes are inserted into the plant genome, respectively. In panel C, the oil body proteins from seed harbouring one insert or multiple inserts of the two different transgenes have been pooled. Lanes A and B represent pooled oil body protein from seed harbouring one copy of the oleoHV1#1 or oleoHV1#2 fusion gene, respectively. Lanes C and D represent pooled oil body protein from seed harbouring multiple copies of the oleoHV1#1 or oleoHV1#2 fusion gene, respectively. Note that the total accumulation of fusion protein in plants harbouring the oleoHV1#1 fusion gene is higher overall than that of oleoHV1#2. Furthermore, the average accumulation of fusion protein is higher in plants harbouring two inserts as opposed to one.


oleoHV1#1
plant line #
oleoHV1#2
plant line #
oleoHV1#1
plant line #
oleoHV1#2
plant line #
kDa
WT
OBIIRT
1 4 27 32 33
13 17 19 21
18 20 30 43 44 45 46
1 4 5 16
A B C D
67
43
30
20.1
14.4
A
B
C

increased protein accumulation. The overall level of fusion protein is higher when oil body protein from virtually non-segregating seeds (i.e. when ≤1/16 of the seeds are wild-type) is used. As shown in panel B, not only is the overall level of fusion protein higher in seeds expressing the oleoHV1#1 *vs.* the #2 construct, this level is higher than in seeds possessing only one copy of the same gene construct. This is confirmed by comparing the overall level of fusion protein from pooled samples of oil body protein derived from each construct and possessing either 1 or more copies of the fusion gene. As shown in Fig. 4.6, panel C, seeds expressing one segregating copy of oleoHV1#1 (i.e. oleoHV1#1, plant number 1, 4, 27, 32, and 33) (lane A) is higher than that of pooled oil body protein samples from similarly segregating lines of oleoHV1#2 (lane B). In addition, non- or virtually non-segregating seed lines harbouring the oleoHV1#1 gene construct (i.e. plants #-18, -20, -30, -43, -44, -45, -46), (Fig. 4.6, panel C, lane C) show higher levels of fusion protein accumulation than that of lines expressing the oleoHV1#2 counterpart (Fig. 4.6, panel C, lane D). Similarily, the level of oleoHV1#1 protein accumulation in plant #27, (panel A, Fig. 4.6), which contains two copies of the fusion gene at one locus, is higher than that of plants harbouring only one copy of that gene construct.

4.2.2.2 Anti-thrombin assays

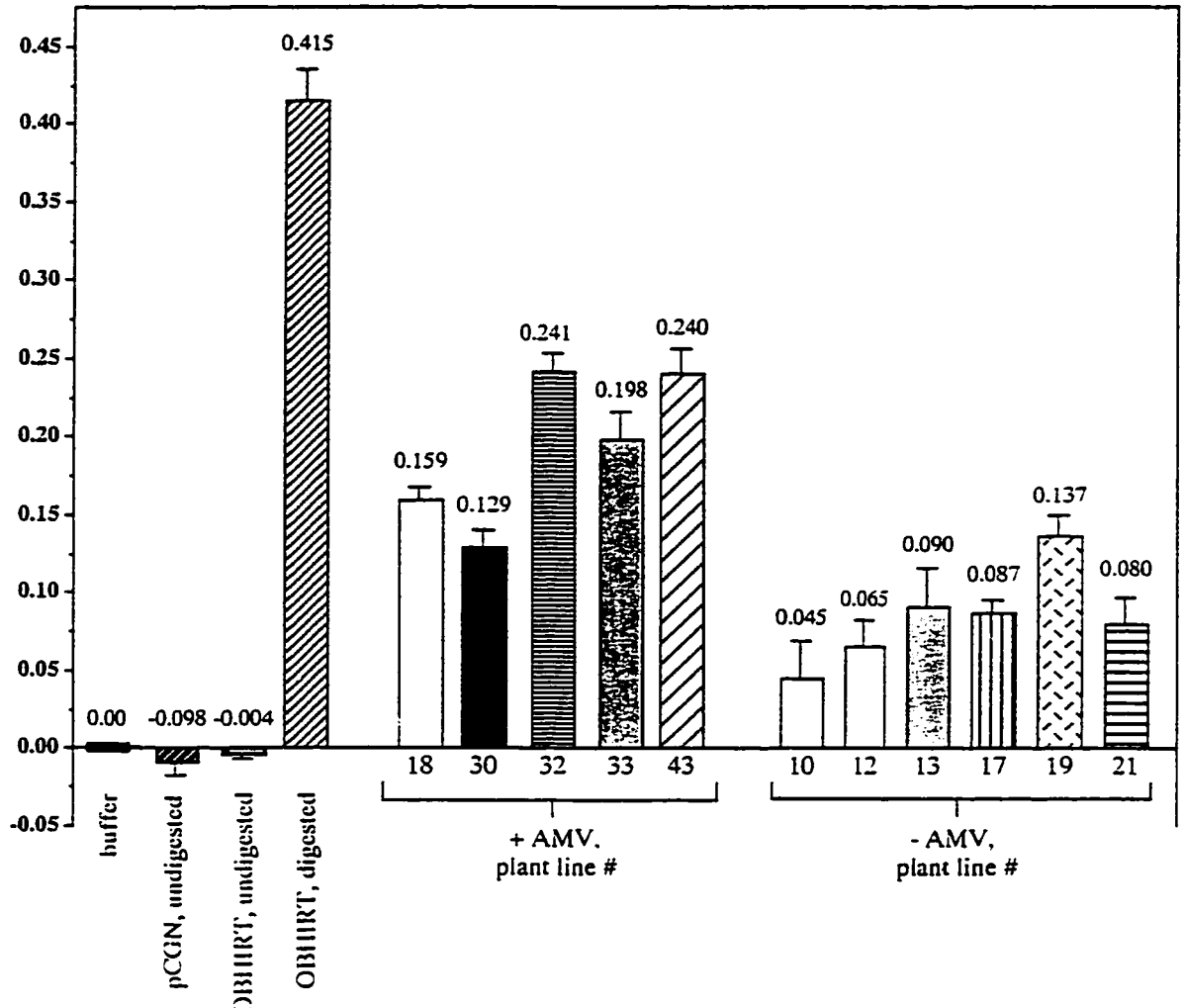
In order to determine the quantitative relationship between oleosin-hirudin variant 1 protein accumulation, and anti-thrombin activity in seed lines harbouring the oleoHV1#1 and #2 fusion gene constructs, colorimetric analysis (section 2.2.19) was performed using Factor Xa-treated seed extracts. One hundred and sixty microliters of oil body protein (9μg/μl) derived from seeds harbouring pCGN-1559 (negative control), OBHIRT-9 (true breeding; section 3.2.2.4), oleoHV1#1 or #2 were treated, or untreated with 5μg of FXa,

overnight. The unternatant from each sample was isolated and assayed for anti-thrombin activity. The anti-thrombin activity per milligram of oil body protein is given in Fig. 4.7. In this figure, samples were derived from control seed and seed from transformed plants which demonstrate a 3:1 ratio of kanamycin-resistance *vs.* susceptibility. In agreement with the results presented in Fig. 4.6, anti-thrombin activity is higher in oil body extracts derived from seed harbouring the oleoHV1#1 construct than that harbouring the oleoHV1#2 construct. The average anti-thrombin units (ATU) per mg of oil body measured in oleoHV1#1 and #2 transformants is 0.19 ATU/mg and 0.084 ATU/mg, respectively. Interestingly, this activity is lower than that derived from the highest expressing seed line harbouring pCGN-OBHIRT (plant line OBHIRT 9, section 3.2.2.1) (0.41 ATU/mg), which possesses a shorter (0.8kbp) promoter and no AMV leader sequence. Anti-thrombin activity was also determined in oil body extracts from seeds demonstrating ≥ 15:1 ratio of kanamycin-resistance *vs.* susceptibility. Factor Xa-treated extracts from these seeds demonstrated a significantly higher level of anti-thrombin activity. Two seed lines each harbouring multiple copies each of oleoHV1#1 (plant #44 and 45) and #2 (plant #4 and 16) demonstrated an average of 1.6 ATU/mg oil body protein, and 0.3 ATU/mg oil body protein, respectively.

4.2.2.3 Relationship between accumulation of fusion protein and steady-state transcript level in transformed seed

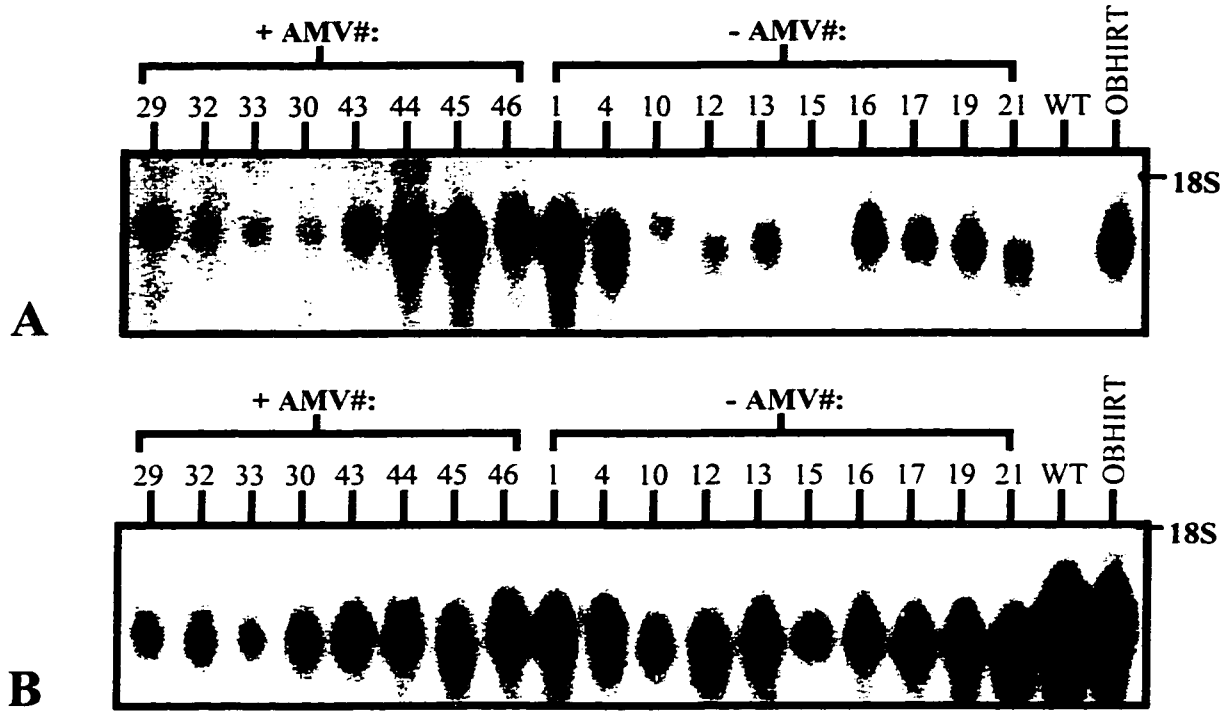
As demonstrated by both Western blot analysis (section 4.2.2.1) and anti-thrombin assays section 4.2.2.2), oleosin-hirudin fusion protein accumulation is higher in plants harbouring the oleoHV1#1 (+AMV) than the oleoHV1#2 (-AMV) fusion gene construct. In order to determine if differences in protein accumulation between these two constructs were due to variations in steady state levels of the oleosin-hirudin transcript, Northern blot

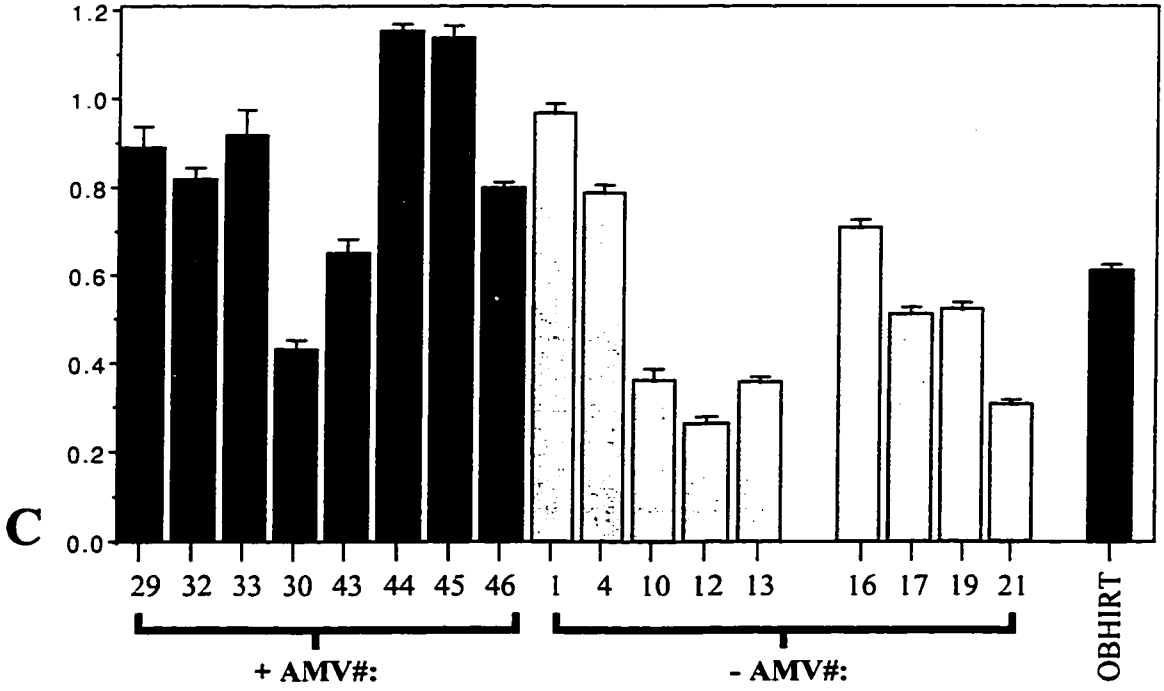
Figure 4.7 Measurement of anti-thrombin activity in unternatant of oil body extracts from *B. napus* transformed with either pCGN 1559 (negative control), oleosin-hirudin variant 2 (OBHIRT), or oleosin-hirudin variant 1 fusion gene constructs possessing or lacking an AMV leader sequence. Seed oil bodies were isolated and treated or not treated with Factor Xa. The unternatant from these samples, or buffer alone, was added to thrombin in buffer. The amount of anti-thrombin activity (in anti-thrombin units/mg oil body protein) is indicated above the appropriate column, and was determined as described in sections 2.2.20. As shown, anti-thrombin activity is only apparent in unternatant from Factor Xa-treated, transformed oil bodies.


0.45
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
-0.05
0.00
-0.098
-0.004
0.415
0.159
0.129
0.241
0.198
0.240
0.045
0.065
0.090
0.087
0.137
0.080
buffer
pCGN, undigested
OBHRT, undigested
OBHRT, digested
18
30
32
33
43
10
12
13
17
19
21
+ AMV,
plant line #
- AMV,
plant line #

analysis was performed. Forty micrograms of total RNA derived from late-cotyledonary embryos were electrophoresed and blotted onto Hybond N+. The blot was hybridized with radiolabelled hirudin-encoding DNA, subject to autoradiography (Fig. 4.8, panel A), stripped, then probed with oleosin cDNA (Fig. 4.8, panel B). As shown in panel B, Fig. 4.6., only the native oleosin, and not the oleosin-hirudin fusion gene, was detectable. As a result, the ratio of oleosin-hirudin transcript to that of native oleosin, could not be calculated. However, using computer-assisted densitometry, comparisons between plant lines of the relative ratio of oleosin-hirudin transcript to native oleosin, can be calculated as shown in Fig. 4.8, panel C. For example, the ratio of oleosin-hirudin: native oleosin transcript for +AMV, plant #44, is 1.18. That for +AMV, plant #46, is 0.81, and thus. has a lower proportion of oleosin-hirudin transcript, with respect to native oleosin, than plant #44. From Fig. 4.8, the average relative ratio of oleosin-hirudin transcript: native oleosin for plants possessing one insert of the oleoHV1#1 and oleoHV1#2 gene constructs are 0.7, and 0.4, respectively. The average relative ratio of oleosin-hirudin transcript: native oleosin for plants which are true-breeding or possess multiple inserts of the oleoHV1#1 or oleoHV1#2 gene construct are 0.99, and 0.82, respectively. It is interesting to note that the three highest expressers, oleoHV1#1 (+AMV), plant # 44, 45 and oleoHV1#2 (-AMV), plant # 1, demonstrate the highest level of fusion protein accumulation, as determined by Western blot analysis (Fig. 4.6). Furthermore, the lowest expressers of the oleosin-hirudin transcript, oleoHV1#2 (-AMV), plant # 10, 12, 13, and 21, show the lowest level of anti-thrombin activity (Fig. 4.7). These results suggest that variations in protein accumulation and anti-thrombin activity may not be due to differences in protein translation, but may be explained by differences in steady-state transcript accumulation. As a result, it is impossible to conclude that the increased levels of protein accumulation noted in the oleoHV1#1 transformants is due to the presence of the AMV leader sequence.

Figure 4.8 Relative ratio of oleosin-hirudin transcript to native oleosin transcript using Northern blot analysis and computer-assisted densitometry. Forty micrograms of total RNA from late-cotyledonary embryos was electrophoresed through a 1% agarose gel containing 6% formaldehyde. Tissue was derived from wild type (WT) plants or *B. napus* lines transformed with either pCGN-OBHIRT (section 2.3.1.1), pCGN-oleoHV1#1 (+AMV) (section 2.3.1.2), or pCGN-oleoHV1#2 (-AMV) (section 2.3.1.3). RNA was blotted onto Hybond N and the membrane was probed with ^{32}P-labeled hirudin-encoding DNA (panel A). The blot was stripped of radiolabel and reprobed with ^{32}P-labeled oleosin cDNA (panel B). To each milliliter of hybridization solution was added 1.75 X 10^4 becquerels of radiolabelled DNA. A numerical value corresponding to the intensity of each of these bands was determined by computer-assisted densitometry and the relative ratio of oleosin-hirudin *vs.* native oleosin transcript was calculated for each plant line expressing the oleosin-hirudin fusion transcript (panel C).


+ AMV#:
- AMV#:
29 32 33 30 43 44 45 46 1 4 10 12 13 15 16 17 19 21 WT OBHIRT
18S
A
+ AMV#:
- AMV#:
29 32 33 30 43 44 45 46 1 4 10 12 13 15 16 17 19 21 WT OBHIRT
18S
B


1.2
1.0
0.8
0.6
0.4
0.2
0.0
29 32 33 30 43 44 45 46 1 4 10 12 13 16 17 19 21 OBHIRT
+ AMV#:
- AMV#:
C

4.3 Discussion

The alfalfa mosaic virus (AMV) RNA4 leader sequence has been shown to increase steady state levels of cauliflower mosaic virus promoter-driven *E. coli* β-glucuronidase (GUS) expression, relative to similar constructs containing no AMV leader, by approximately 8-fold in transgenic tobacco leaves (Datla *et al.*, 1993). This increase in protein accumulation occurred independent of transcript abundance. Western blot analysis of transgenic plants possessing 1 or more copies of either the oleoHV1#1 or #2 gene constructs showed that, overall, oleosin-hirudin variant 1 fusion protein accumulated to a higher extent in seeds when the AMV leader was present 5' to the initiation codon (i.e. the oleoHV1#1 construct) (Fig. 4.6). Interestingly, from the results of both Western blot analysis, and anti-thrombin assays, when seeds harboured more than one copy of the insert, the fusion protein accumulation appeared to be additive for either construct (Fig. 4.6, and section 4.2.2.2). The level of protein accumulation, as determined by Western blot analysis, corresponds to the level of anti-thrombin activity present in the unternatant of oil body extracts treated with Factor Xa. Seeds harbouring 1 copy of the oleoHV1#1 or oleoHV1#2 fusion gene construct demonstrated an average of 0.19 and 0.084 anti-thrombin units (ATU)/mg of oil body protein, respectively. These seeds were derived from first-generation plants, and thus, demonstrated a 3:1 ratio of kanamycin resistance *vs.* susceptibility. Somewhat surprisingly, seeds harbouring the pCGN-OBHIRT construct (section 3.2.2.4; OBHIRT-9) demonstrated higher anti-thrombin activity (approximately 0.4 ATU/mg of oil body protein) than either oleoHV1#1 or #2. OBHIRT-9 seeds are true-breeding (i.e. every seed progeny harbours the oleosin-hirudin variant 2 fusion gene construct). Detailed segregation analysis of GUS shows that pooled homozygous lines will express 2X the level of activity as pooled heterozygous lines possessing one locus of the GUS gene (i.e. GUS functions as a codominant gene) (Scott Richmond and Dr. Gijs van

Rooijen, personal communication). If this is true for the oleosin-hirudin fusion gene construct, it may explain the fact that just over twice the level of anti-thrombin activity is detected in pooled homozygous lines (OBHIRT-9) than the pooled heterozygous lines harbouring either the oleoHV1#1 or #2 gene constructs. Furthermore, the OBHIRT-9 seed line used was one of the highest expressers tested in Chapter 3. Thus, the high level of expression observed in this seed line may be due to position effects. Of first-generation seeds harbouring multiple copies of the oleoHV1#1 or oleoHV1#2 fusion gene construct, the anti-thrombin activity was notably higher (1.6 and 0.3 ATU/mg oil body protein, respectively). From the results presented here, there appears to be a strong correlation between the level of protein accumulation (as determined by Western blot analysis) and the level of anti-thrombin activity. Analysis of FPLC-purified, Factor Xa-treated oil body unternatant from plants transformed with the oleosin-hirudin fusion gene construct have demonstrated that recovered hirudin protein is fully active (i.e. no hirudin protein was detected which did not retain its anti-thrombin qualities) (data not shown). This suggests that the process of oil body preparation and proteolytic release of hirudin from the oleosin carrier is sufficient for the correct formation of the three disulfide bridges.

The most obvious explanation for greater fusion protein accumulation in seeds harbouring the oleoHV1#1 construct than the oleoHV1#2 construct is the presence of the AMV leader sequence. However, as shown in Fig. 4.8., the steady-state levels of the transcripts from late-cotyledonary embryos expressing these constructs vary. Thus, such assumptions should be made with caution. This is surprising, as previous studies have shown that elevations in protein accumulation attributable to the presence of the AMV leader sequence are not associated with elevated transcript levels (Datla *et al.*, 1993, and references therein). There are a number of factors one must consider when assessing the reliability of the data presented in Fig. 4.8. Comparisons between the ratio of oleosin-

hirudin *vs.* native oleosin transcripts among plant lines is dependent upon the sensitivity and accuracy of the computer-assisted densitometry program, a low background on the autoradiograms, and a requirement that band intensity falls within a linear range. If time had permitted, this experiment would have been repeated and the relative abundance of the oleosin-hirudin *vs.* the native oleosin transcript would have been verified using quantitative reverse-transcriptase PCR technology.

Despite prolonged exposure times, it was surprising that both the oleosin-hirudin fusion transcript, and native oleosin transcripts could not be visualized on the same Northern blot probed with radiolabelled oleosin cDNA (Fig. 4.8, panel B). This suggests that further improvements can be made to increase the level of transcription of fusion gene constructs. The constructs presented in this chapter made use of an *Arabidopsis* oleosin promoter subcloned by Dr. Aine Plant (Plant *et al.*, 1994). This promoter was modified at its 3' end to include a Sma 1 (CCC/GGG) immediately downstream of the TATA box (Fig. 4.2). Because G-C bonds are stronger than A-T bonds (held by 3 *vs.* 2 hydrogen bonds, respectively), this restriction site may not be conducive to DNA denaturation by the RNA polymerase II holoenzyme during transcription initiation. Both the oleoHV1#1 and #2 constructs possessed this restriction site at the same proximity relative to the TATA box, and therefore, if it was affected by this, it is likely that the effect would have been equivalent for both the oleoHV1#1 and #2 constructs. If future experiments using the 2.5 kbp *Arabidopsis* oleosin promoter are to occur, it may be desirable to modify the 5' untranslated region such that its sequence more closely reflects that of the native oleosin promoter (Fig. 4.2, OBHIRT). Furthermore, the addition of DNA bases 5' to the ATG start codon, such as the insertion of an AMV leader sequence, could affect protein accumulation other than at the level of translational enhancement. This could include

enhanced mRNA stability. If future experiments involving the AMV leader sequence were to occur, it would be of value to test this possiblilty using nuclear run-on experiments.

Chapter 5:

Expression of oleosin-hirudin variant 1 concatamers in *Brassica napus*

5.1 Introduction

One of the challenges associated with the successful accumulation of high levels of oleosin-hirudin fusion proteins in plant seeds is the possible competition, provided by native oleosins, for a position on the oil body membrane. Fusion proteins possess no competetive advantage for targeting to the oil body over native oleosins. It was for this reason that the question was posed: if more than one hirudin-encoding sequence is fused, in tandem, to the 3'end of the oleosin cDNA, would this result in a higher overall accumulation of hirudin in plant seeds?

The fusion of large proteins to the C-terminal end of oleosins, such as the 68 kDa protein β-glucuronidase (GUS), has recently been reported (van Rooijen and Moloney, 1995b). These authors showed that oleosin-GUS fusions not only accumulated in *B. napus* seeds, but correctly targeted to the oil body fraction. The finding that oleosins could tolerate the addition of large proteins to its C-terminus was somewhat expected, as seed oleosins vary largely in both their size and amino acid composition at the C-terminal end. Thus, addition of higher molecular weight proteins to the C-terminus of oleosin, such as hirudin concatamers, is unlikely to affect correct targeting of the fusion protein, at least on the basis of molecular weight.

Section 2.3.2.2 describes the construction of such oleosin-hirudin concatamers. Three constructs were created, having 1, 2, or 4 hirudin concatamers fused, in-frame, to the 3'end of a *B. napus* oleosin cDNA. These constructs are named oleoHV1C, oleo2HV1, and oleo4HV1, respectively. Interposed between each of the hirudin concatamers, and between the oleosin-hirudin-encoding sequenced, is a Factor Xa recognition site. Factor Xa cleaves after the arginine residue in the sequence ile-glu-gly-arg. Because this recognition site will remain attached to the C-terminal end of each hirudin moiety joined in tandem

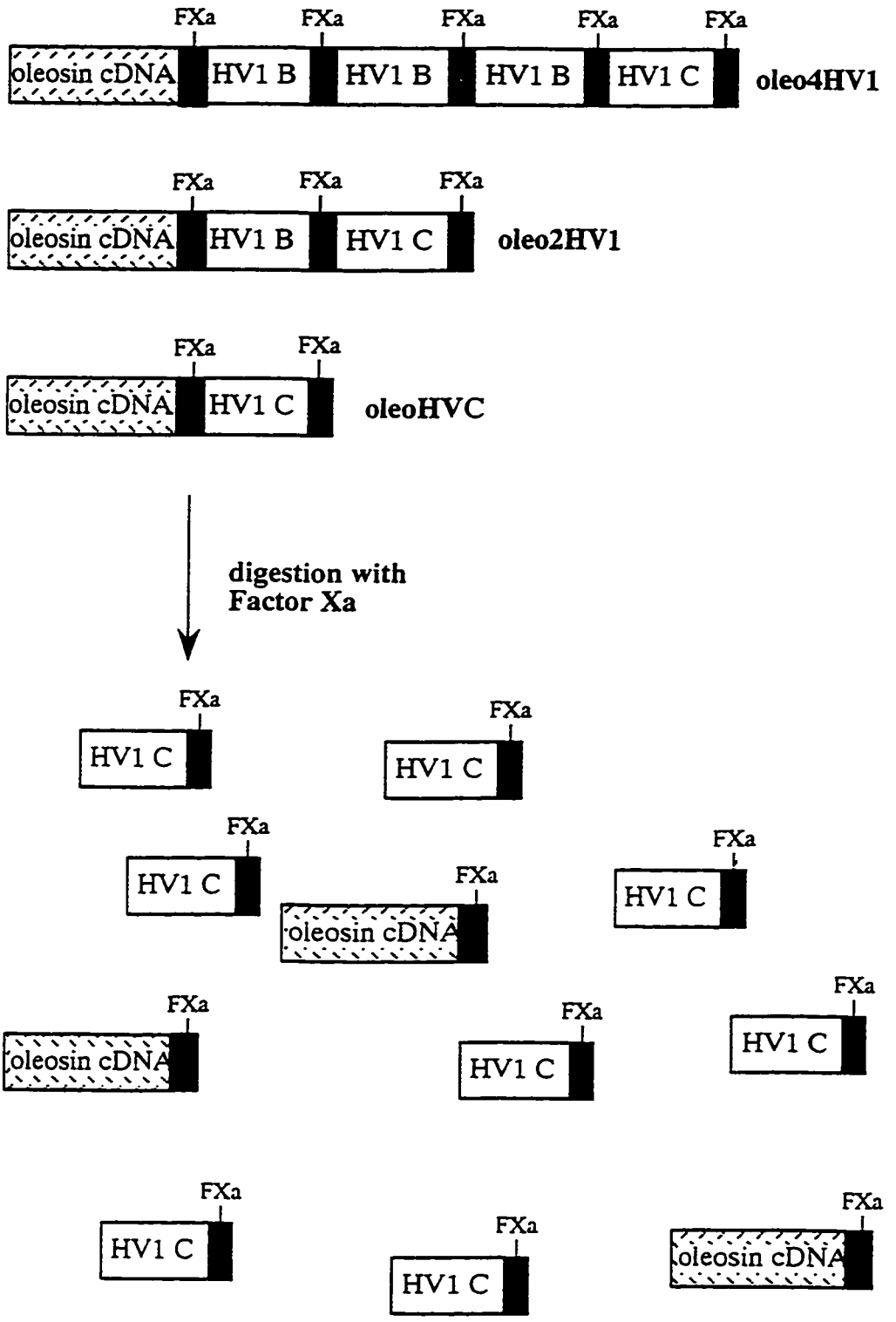
array, a Factor Xa recognition site was also added to the final hirudin sequence, such that all hirudin peptides derived from such fusions would be identical (Fig. 5.1). Thus, the construct oleoHV1C differs from the oleoHV1#1 and oleoHV1#2 constructs described in Chapter 4, by the presence of a Factor Xa recognition site at the 3' end of hirudin. All of the concatamer constructs are driven by a 2.5 kbp *Arabidopsis* oleosin promoter, have an AMV leader sequence, and are provided with a NOS transcriptional terminator. The following are preliminary results comparing the accumulation of fusion protein encoded by these three constructs using Western blot analysis and anti-thrombin assays.

5.2 Results

5.2.1 Plasmid Construction

The construction of hirudin concatamers is diagrammed in Fig. 2.7. Shown is the sequential, stepwise addition of hirudin-encoding moieties starting at the 5'end of nos transcriptional terminator in the plasmid vector, pUC 19. The sequential ligation of hirudin-encoding DNA at the 5' end of the growing concatameric sequence was performed such that, after each addition, the integrity of the sequence and its junctions could be assured by sequencing in from the 5'end. As the number of hirudin concatamers became larger (i.e. 5 hirudin-encoding sequences in tandem array), the colony size of the *E. coli* host strain into which they were transformed decreased. This resulted in very low plasmid yields and confounded the ability to identify clones containing the desired plasmid DNA. At this point, the choice of *E. coli* host strain was changed from DH5α to DH10B. This resulted in larger colonies containing the plasmid strain of interest, and higher plasmid yields. The differences between the genotypes of DH5α and DH10B gives no indication as to why the change in host strain would result in such marked differences in colony size or plasmid

Figure 5.1 Schematic representation of Factor Xa (FXa) cleavage of oleosin-hirudin concatamers. OleoHV1C, oleo2HV1, and oleo4HV1 represent oleosin fusion proteins having 1, 2, and 4 hirudin concatamers fused to the C-terminus of oleosin, respectively. Interposed between each hirudin molecule and between oleosin and hirudin, is a site recognized by the protease, Factor Xa. The last hirudin in the fusion protein also has a Factor Xa recognition site, such that after proteolytic cleavage, all hirudin proteins are identical. After cleavage, hirudin released into the unternatant while the oleosin remains attached to the oil body membrane.

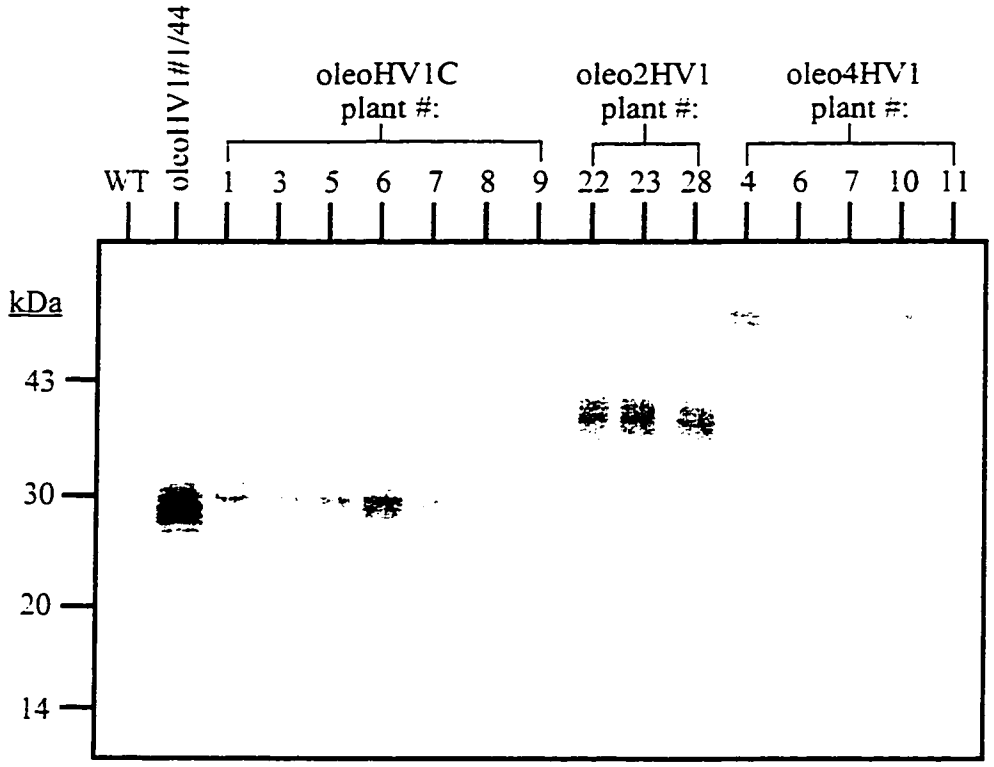

FXa
FXa
FXa
FXa
FXa
oleosin cDNA
HV1 B
HV1 B
HV1 B
HV1 C
oleo4HV1
FXa
FXa
FXa
oleosin cDNA
HV1 B
HV1 C
oleo2HV1
FXa
FXa
oleosin cDNA
HV1 C
oleoHVC
digestion with
Factor Xa
FXa
HV1 C
FXa
HV1 C
FXa
HV1 C
FXa
oleosin cDNA
FXa
HV1 C
FXa
oleosin cDNA
FXa
HV1 C
FXa
HV1 C
FXa
HV1 C
FXa
oleosin cDNA

yield. Using the DH10B strain, 8 hirudin-encoding concatamers were successfully cloned in front of the nos transcriptional terminator. The eight concatameric sequences and the nos transcriptional terminator were cloned behind the *Arabidopsis* oleosin promoter-*B. napus* oleosin cDNA using the cloning strategy given in Figs. 2.8-2.10. However, despite repeated attempts, this oleosin-8 hirudin-nos DNA sequence could not be cloned into pCGN 1559. Thus, only 1, 2, and 4 hirudin DNA concatamers fused, in-frame, to the 3'end of oleosin cDNA and terminated by a nos transcriptional terminator, were cloned into pCGN 1559 (Figs. 2.8-2.10). These 3 constructs were used to transform *B. napus* petioles (section 2.2.7) for the purpose of producing transgenic plants.

5.2.2 Western blot analysis of oleosin-hirudin variant 1 concatamers

B. napus petioles were transformed as described in section 2.2.7 with either the pCGNoleoHV1C, pCGNoleo2HV1, or pCGNoleo4HV1 fusion gene constructs (section 2.3.2). Dry seed (0.5g each) from *B. napus* transformed with the pCGN 1559 binary vector (negative control), oleoHV1#1, plant number 44 (sections 2.3.1.2 and 4.2.2.1) (positive control) and a number of plant lines harbouring the pCGNoleoHV1C, pCGNoleo2HV1, or pCGNoleo4HV1 fusion gene construct were ground in 10 ml of extraction buffer and the oil bodies isolated as described in section 2.2.13. Five micrograms of oil body protein from each sample was loaded onto a 15% SDS-polyacrylamide gel and electrophoresed. The proteins were transferred onto PVDF membranes and incubated with mouse monoclonal antibodies raised against hirudin. The expected molecular weight of oleoHV1 #1/44 and oleoHV1C is approximately 30 kDa while that of the oleosin-hirudin concatamer constructs, oleo2HV1 and oleo4HV1 is 37, and 52 kDa, respectively. As shown in Fig. 5.2, bands corresponding to the expected molecular weight of each protein type are evident. However, additional lower molecular

Figure 5.2 Immunological detection of oleosin-hirudin fusion proteins by antibodies raised against hirudin. Five micrograms of oil body protein from wild-type (WT) *B. napus* seeds and seed expressing oleoHV1#1/44, oleoHV1C, oleo2HV1, and oleo4HV1 were electrophoresed via SDS-PAGE, transferred onto a PVDF membrane, and incubated with antibodies raised against hirudin. As shown, bands of approximately 30, 37, and 52 kDa are apparent in seeds expressing oleoHV1C, oleo2HV1, and oleo4HV1 fusion proteins, respectively.


oleo1HV1#1/44
oleoHV1C
plant #:
oleo2HV1
plant #:
oleo4HV1
plant #:
WT
1 3 5 6 7 8 9
22 23 28
4 6 7 10 11
kDa
43
30
20
14

weight bands were observed below those of the predicted molecular weight. It is unclear whether this is due to aberrant mobility of the proteins through the gel, or if it is a result of protein degradation *in vivo* or during oil body isolation.

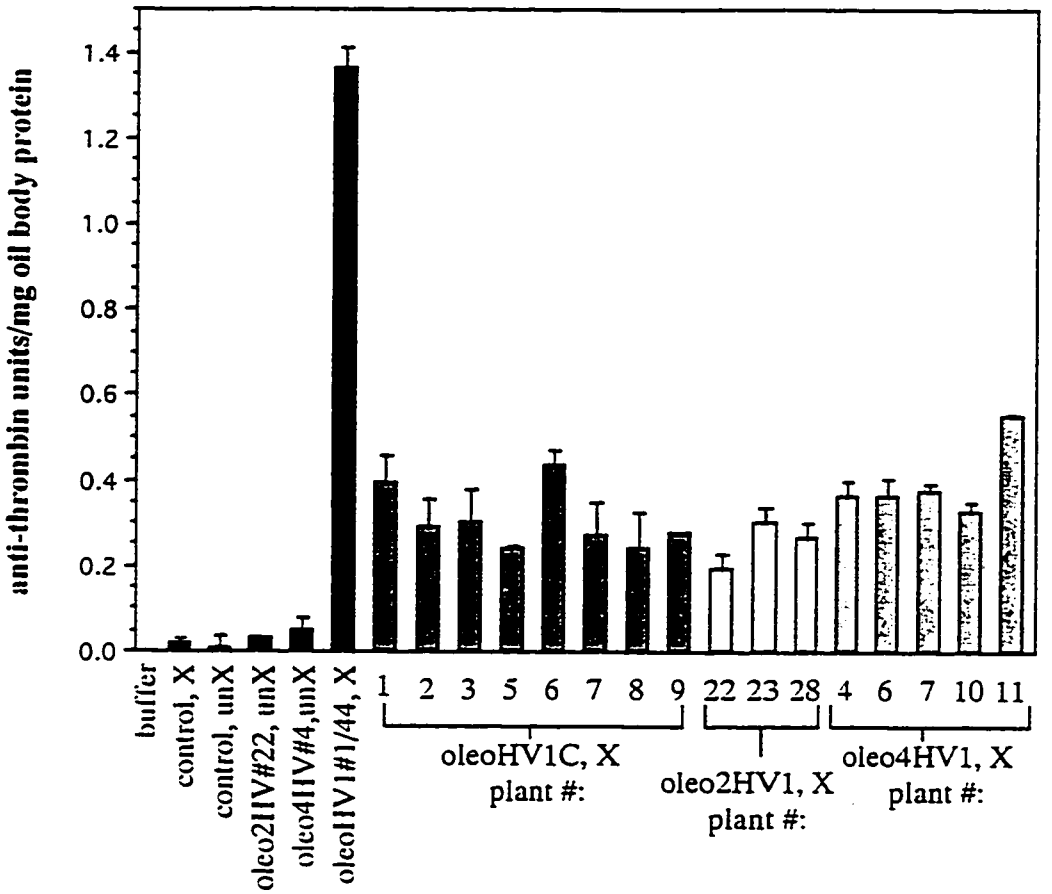
Compared to the levels of fusion protein apparent in oleoHV1#1/44, the level of fusion protein present in seeds harbouring the oleoHV1C construct appear to be much lower. It should be noted that the oleoHV1#1/44 plant line was the highest expresser of oleosin-hirudin variant 1 fusion protein (Fig. 4.4 and section 4.2.2.2), so this is somewhat expected. It was expected that if the fusion protein accumulated to equal levels in plants harbouring the oleosin-hirudin concatamer constructs, the bands detected by anti-hirudin antibodies would be 2 and 4 times as intense in plants harbouring the oleo2HV1 and oleo4HV1 construct, respectively, than oleoHV1C. This does not seem to be the case. The level of fusion protein accumulation in the oil body fractions from plants transformed with the oleo4HV1 construct appears to be somewhat lower than that of oil body extracts from plants transformed with either oleoHV1C or oleo2HV1. Possible reasons for this include position effects, resulting in lower levels of expression at the transcription level, or protein degradation. Northern blot analysis of embryo RNA from each of these plants lines would help distinguish between these two possibilities. The low levels of fusion protein accumulation in the oil body extracts of oleo4HV1 transformants may also be due to the construct, itself. As described in section 5.2.1, the bacterial host grew smaller as the number of hirudin-encoding concatamers increased in the plasmid vector. This may have been due to secondary structures which inhibited the replication of the plasmid vector, and, therefore, the plasmid copy number in the bacterial host. A reduction in the plasmid copy number would result in a reduction in the number of copies of the gene that confers antibiotic resistance (ampicillin). This may explain why the size of the bacterial colonies containing multiple hirudin concatameric sequences were small. *In planta*, the presence of

secondary structure in the messenger RNA corresponding to the oleo4HV1 transcript may have resulted in its degradation or reduced capacity for translation, resulting in lower fusion protein yields. Although the level of fusion protein accumulation in plants harbouring the oleo4HV1 gene construct is lower than expected, it is clear that each of the oleosin-hirudin concatamer fusions accumulate and target to the oil body fraction of transgenic *B. napus*.

5.2.3 Quantitation of biologically active hirudin in seeds expressing oleosin-hirudin concatamers

Oil body extracts (1mg/ml) derived from seed lines harbouring pCGN 1559 (negative control) or expressing either oleoHV1#1/44 (sections 2.3.1.2 and 4.2.2.1), oleoHV1C, oleo2HV1, or oleo4HV1 were treated (X) or untreated (unX) with the protease Factor Xa overnight at room temperature. Thrombin and a chromogenic substrate were added to the unternatant isolated from these samples and the change in absorption at 405nm was measured and converted into anti-thrombin units/mg oil body protein (section 2.2.19). The results are shown in Fig. 5.3. As predicted by Western blot analysis (Fig. 5.2), the highest level of anti-thrombin activity is present in the seed line, oleoHV1#1/44 (approximately 1.3 ATU/mg oil body protein). Although the level of oleo4HV1 fusion protein accumulation presented in Fig. 5.2 appears to be lower, overall, than that of oil body extracts from plants harbouring the oleoHV1C construct, the level of anti-thrombin activity is somewhat higher in seeds harbouring the oleo4HV1 construct (average of 0.39 ATU/mg oil body protein) than that expressing oleoHV1C (average of 0.3 ATU/mg oil body protein). If hirudin in the form of 4 concatamers fused to oleosin is not detected by antibodies with the same intensity as 4 oleosin-hirudin fusion proteins, four times as much hirudin activity (per mol of oleosin fusion) may be present in oil body extracts from plants harbouring oleo4HV1. However, while the level of oleo2HV1 fusion protein accumulation

Figure 5.3 Measurement of antithrombin activity in transformed and wild-type oil body extracts. Oil body protein (1mg/ml) from seeds harbouring pCGN 1559 (control), or expressing the fusion proteins oleoHV1#1/44 (section 4.2.2.2), oleoHV1C, oleo2HV1, or oleo4HV1 were treated (X) or untreated (unX) with the protease, Factor Xa. Fifty microliters of unternatant from these samples, or buffer alone, was added to thrombin in buffer. The amount of anti-thrombin activity (anti-thrombin units/ per mg of oil body protein) remaining after addition of oil body unternatant or buffer was determined as described in section 2.2.20. The thrombin activity present in the samples cleaved with Factor Xa have been adjusted to account for the presence of contaminating thrombin present in commercial Factor Xa. As shown, anti-thrombin activity is present only in the unternatant of transformed seed extract (with the exception of control seed) when the oil bodies have been treated (X) with Factor Xa.


anti-thrombin units/mg oil body protein
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
buffer
control, X
control, unX
oleo2IIV#22, unX
oleo4IIV#4,unX
oleo1IV1#1/44, X
1 2 3 5 6 7 8 9
oleoHV1C, X
plant #:
22 23 28
oleo2HV1, X
plant #:
4 6 7 10 11
oleo4HV1, X
plant #:

was higher than either oleoHV1C or oleo4HV1 (Fig. 5.2), the level of anti-thrombin activity was slightly lower (average of 0.25 ATU/mg oil body protein). These somewhat conflicting results may be accounted for, besides sampling error, by incomplete proteolytic digestion and release of the hirudin moiety from its oleosin "carrier", as hirudin is inactive while as an oleosin fusion (data not shown).

5.3 Discussion

The results presented in sections 5.2.1 and 5.2.2 demonstrate that oleosin-hirudin concatamers accumulate and target to the oil body fraction of transgenic seeds. In addition, hirudin is biologically active after it has been proteolytically released from its oleosin carrier. Although the level of fusion protein which accumulated in plants expressing the oleo2HV1 construct is higher than those expressing the oleoHV1C construct (Fig 5.2), the level of hirudin activity, as determined by anti-thrombin assays (Fig. 5.3) appears to be somewhat lower. In addition, oleo4HV1 fusion protein accumulated to slightly lower levels, overall, than oleoHV1C, however the level of hirudin activity was somewhat higher. The expression of the oleosin-hirudin concatamers may have been reduced at the transcriptional level, due to either position effects or mRNA turnover. The latter is less likely, since heterologous DNA constructs, such as that encoding β-glucuronidase (GUS) have been successfully expressed as fusions to the 3'end of oleosin-encoding sequences (van Rooijen and Moloney, 1995 a, b). Northern blot analysis of embryonic RNA from these transformants would give insight to this possibility. It is unlikely that the relatively low levels of oleosin-hirudin concatamers is a result of unstable insertion of the fusion protein into the oil body membrane, as it is believed that it is the central domain of the oleosin which is primarily responsible for anchoring the protein into the oil body membrane (van Rooijen and Moloney, 1995a). The Western blot (Fig. 5.2) shows a cluster of bands

in the range of the size expected for each construct. These lower molecular weight bands are not only present in the oil body extracts of seeds expressing oleosin-hirudin concatamer fusions, but also the positive control, oleoHV1#1/44. Thus, these lower molecular weight bands may represent degradation products which occurred during oil body preparation. This result emphasizes the need to repeat these experiments in order to ensure their validity.

The results presented here demonstrate that the fusion of concatameric proteins of interest to the C-terminus of oleosins represents a viable method for which heterologous proteins may be expressed. Fusion proteins consisting of multiple subunits of immunogenic epitopes have been used for the production of antisera (Kjerrulf *et al.*, 1997; Ryan *et al.*, 1997). However, searches through the Biological Abstract databases have revealed no previous publications in which concatameric fusion proteins have been used for the purpose of increased protein yield. It should be noted that care must be taken in the choice of heterologous protein wished to be expressed and the proteolytic cleavage site necessary for its release from the oleosin carrier. The hirudin proteins expressed using this system retain a Factor Xa cleavage site at their C-terminus (Fig. 5.1). Thus, if hirudin expressed in this manner were to be used therapeutically, it would have to undergo rigorous and costly clinical trials to ensure that this modification does not change such qualities as its immunogencity or its ability to inhibit thrombin. A way to circumvent this problem would be to find a protease which recognizes the C-terminal end of hirudin (i.e. glu-tyr-leu-gln) and cleaves after its last residue. This problem would likely be less important if the protein of interest was an industrial enzyme or was involved in improving meal quality.

Chapter 6:

Summary and General Discussion

6.1 Discussion

The results from the preceding chapters may be summarized as follows:

• The 0.8kbp *Arabidopsis* oleosin promoter drives the expression of the oleosin-hirudin variant 2 fusion gene in *B. napus* and tobacco seeds. The 2.5kbp *Arabidopsis* oleosin promoter was also shown to drive oleosin-hirudin variant 1 gene expression in *B. napus*

• Expression of oleosin-hirudin variant 2 fusion gene occurs in a seed-specific manner in transgenic *B. napus*

• Isolation and repeated "washing" of the oil body fat pad derived from *B. napus* seeds via flotation centrifugation results in the removal of the majority of storage seed protein with minimal loss of oleosins

• Oleosin-hirudin fusion protein accumulates in transgenic *B. napus* and tobacco seeds

• Oleosin-hirudin fusion protein is localized primarily to the oil body fraction (as opposed to the soluble or pelleted seed fraction) of transgenic seed ground and centrifuged in buffer

• Hirudin expressed as an oleosin fusion resides on the cytosolic side of the oil body

• The Factor Xa recognition site which lies between the oleosin and hirudin protein sequences is accessible to the protease in oil body extracts

• Hirudin is released from its oleosin carrier by the site-specific protease, Factor Xa

• Hirudin variant 2, having an activity of approximately 0.84 anti-thrombin units/mg of oil body, was recovered in the unternatant of Factor Xa-treated tobacco oil bodies

• Hirudin variant 2, having an activity of approximately 0.55 anti-thrombin units/mg of oil body, was recovered in the unternatant of Factor Xa-treated *B. napus* oil bodies

• Hirudin variant 1, having a maximal activity of approximately 1.8 anti-thrombin units/mg of oil body, was recovered in the unternatant of Factor Xa-treated *B. napus* oil bodies

• Hirudin variant 2 was purified to near-homogeneity (13,200 anti-thrombin units/mg protein) via anion exchange and reverse phase chromatography

• Oleosin-hirudin fusion proteins are stable in seeds for prolonged periods of time (years)

• The alfalfa mosaic virus RNA4 leader sequence may enhance overall oleosin-hirudin protein accumulation in transgenic seeds

• Higher copy numbers of the oleosin-hirudin variant 1 fusion gene in the *B. napus* genome results in higher levels of protein accumulation in transgenic seed

• Oleosin-hirudin variant 1 concatamers accumulate in transgenic *B. napus* seed and target to the oil body

The results presented in this thesis clearly demonstrate that hirudin accumulates as a C-terminal oleosin fusion in *B. napus* and tobacco seed. Some of these results have been previously published (Boothe *et al.*, 1988; Parmenter *et al.*, 1995, 1996). The results of van Rooijen and Moloney (1995b) suggest that the choice of heterologous protein fused to the C-terminus of oleosin in not limited by size. Higher anti-thrombin activity per milligram of oil body protein was measured in tobacco seed expressing the oleosin-hirudin variant 2 (OBHIRT) fusion gene construct than in *B. napus* harbouring the same construct. There are between 4-6 members of the oleosin gene family in *B. napus* (Keddie *et al.*, 1992; Dr. S. Bowra, personal communication). Little is known about tobacco oleosin genes as they have yet to be cloned. It is possible that tobacco oleosin genes are comprised of a smaller gene family than that found in *B. napus*. If this is true, then the addition of a foreign oleosin into the tobacco genome would represent a higher percentage of the total oleosin population in transgenic tobacco than in *B. napus*. This might explain, beyond reasons such as position effects, or systematic or sampling error, the differences in anti-thrombin activity recovered from the oil body fractions of these two plant species.

Hirudin has not only been successfully expressed in *B. napus* and tobacco, but also more recently in the species, *Brassica carinata* (Chaudhary *et al.*, 1998). These results suggest that this type of foreign gene expression may be feasible in other plant species which accumulates oil in the form of oil bodies. The ability to express foreign sequences as C-terminal oleosin fusions in *B. carinata* is significant, as this plant species demonstrates a low frequency of outcrossing (less than 8% pollination between forced crosses of *B. napus* and *B. carinata* (Chaudhary *et al.*, 1998 and references therein). The ability to use *B. carinata* for this purpose is significant, as it provides a biological barrier between potentially transformed varieties of this species, and the widely grown species, *B. napus*.

To date, large-scale production of recombinant hirudin has been most successfully accomplished in the heterologous expression system, *Saccharomyces cerevisiae*. Using a batch fermenter, hirudin secreted into 200 liters of culture media was isolated and purified in two chromatographic steps, followed by decolorization, to yield 3.6g of 96% pure protein (Lehman *et al.*, 1993). Our preliminary results have shown that, after proteolytic release from its oleosin carrier and purification using anion-exchange and reverse phase high performance liquid chromatography, 0.741mg of approximately 94% pure hirudin is recovered from 50g of *B. napus* seed harbouring the pCGN-OBHIRT construct (Jill Saponja and Dr. Joseph Boothe, personal communication). This recovery will likely improve with greater efficiency of oil body recovery, site-specific proteolysis and heightened expression levels. In fact, one plant (number 44) harbouring the construct pCGN-oleoHV1#1 (section 4.2.2.2) demonstrates more than three-fold higher levels of anti-thrombin activity in the unternatant of Factor Xa-treated oil bodies than that of the highest expresser harbouring the pCGN-OBHIRT construct.

As described in the introduction (section 1.2.6), a wide variety of heterologous proteins have been produced using plant-based expression systems. The accumulation of

human serum albumin in transgenic potatoes represented 0.02% of total leaf protein extracts (Sijmons *et al.*, 1990). Single chain antibodies expressed in transgenic tobacco have been shown to accumulate to 1.3% of the total leaf protein (Hiatt *et al.*, 1989). Preliminary estimates suggest that fusion protein accumulation in plants expressing the AMV-oleosin-hirudin variant 1 construct represent approximately 10% of oil body protein and, therefore, approximately 1% of total seed protein (data not shown). This level of protein accumulation compares well with other plant-based expression systems. Twenty percent of the total *B. napus* seed weight is protein. Assuming hirudin accumulates to a level of 1% total seed protein, it represents approximately 0.2% of total seed weight, or 2kg/tonne of seed. Canola seed costs approximately $250/tonne. Assuming downstream processing for the purification of hirudin would cost 10X this amount, this value would be $2750/tonne seed. Thus, the cost of purification of recombinant hirudin from seed, assuming 50% recovery, would be ($2750/tonne)X(1 tonne/2 kg hirudin)X(2)= $2.75/g hirudin. Currently, the cost of production of hirudin is at least $30 (US)/g (Ciba Geigy, personal communication). Thus, the production of hirudin in plant seeds using oleosin partitioning technology represents a highly competitive, and potentially lucrative system.

Of the four oleosin isoforms in the *Arabidopsis* genome, the most abundantly expressed gene in this family is the clone used in the creation of the oleosin-hirudin variant 2 fusion gene constructs described in Chapter 3 (Dr. Gijs van Rooijen and Scott Richmond, personal communication). Given the strength of the *Arabidopsis* oleosin promoter used in the work described here, it was expected that the fusion gene would represent a significant proportion of the mRNA produced. However, as shown in figures 3.5, 3.6, and 4.6, the fusion gene could not be detected on Northern blots probed with radiolabelled oleosin-encoding DNA. Thus, increasing the level of expression of the oleosin-hirudin fusion gene constructs represents an area of future improvement. Plant and

colleagues (1994) demonstrated that a region of the *Arabidopsis* oleosin promoter may be responsible in down-regulating gene expression. It is possible that removing this portion of the promoter may result in enhanced levels of expression of DNA sequences downstream of it while retaining correct spatial and temporal regulation of expression.

There are a number of studies which have employed the use of seed-specific promoters for the expression of heterologous proteins in plants. Some of the most notable of these have employed such promoters to direct the expression of protein-encoding sequences which would change the nutritive content of the seed meal. In a recent report, DNA sequences encoding a coiled-coil high-lysine, high-methionine peptide was driven by either the bean phaseolin promoter or the β-conglycinin promoter (Keeler *et al.*, 1997). Here, the heterologous protein accumulated to as much as 2% of the total tobacco seed protein. The phaseolin promoter has also been used to direct the expression of Brazil nut 2S albumin, which accumulated to between 1.7-4% of total seed protein in transgenic canola (Altenbach *et al.*, 1992). In an earlier study, the phaseolin promoter was used to direct the expression of a brazil nut protein in transgenic tobacco (Altenbach *et al.*, 1989). Here, the heterologous protein accumulated to up to 8% of the salt-extractable protein. An *Arabidopsis* 2S albumin promoter has also been used to direct the synthesis of brazil nut protein, which accumulated to up to 1.28% of the total seed protein (Conceicao *et al.*, 1994). A promoter that drives the seed-specific expression of the *B. napus* protein, napin (2S albumin), has also been used in heterologous expression studies. For example, this promoter was responsible for high levels of accumulation of a pea seed 2S albumin in transgenic *B. napus* (Stayton *et al.*, 1991). One of the reasons that these proteins accumulated to such high levels is the transcriptional activity of the promoter used. It should be noted, however, that care must be taken when choosing a promoter, as temporal, as well as spatial patterns of expression may be of great importance. For example, it is

expected that oleosin fusion mRNA must be present prior to or during oil body formation in order to synchronize with oil body biogenesis (Qu *et al.*, 1986; Tzen *et al.*, 1993). In this example, the transgene could be expressed under the control of a strong oleosin promoter from the same, or a different species. For example, a soybean oleosin promoter was used to drive the expression of a soybean oleosin in *B. napus* (Sarmiento *et al.*, 1997). Here, the foreign protein accumulated to approximately 20% of the native oleosin protein and in the same temporal manner. A possible alternative to the use of oleosin promoters to direct the expression of an oleosin fusion is the use of a seed-specific stearoyl-acyl carrier protein (ACP) desaturase promoter (Slocombe *et al.*, 1992). The desaturase promoter directs transcription of mRNA earlier than that of the oleosin promoter and peaks slightly earlier (45 days after anthesis (DAA)) than when oleosin transcripts begin to accumulate (45-50 DAA). The fact that these promoters are capable of driving high levels of expression of a transgene suggests that it is highly feasible that oleosin fusions may be expressed and accumulate to at least 1% of total seed protein.

An alternative approach for the enhancement of transgene expression in plants may be accomplished through the employment of scaffold attachment regions (SARs). These regions bind a proteinaceous nuclear scaffold which results in the formation of loops in the chromatin. It is believed that the formation of loops isolates the DNA present between the SARs, thus minimizing the effects of *cis*-regulatory elements in neighbouring chromatin (Grosveld *et al.*, 1987). Recently, a tobacco SAR clone flanking both ends of a β-glucuronidase (GUS) reporter gene was used to stably transform tobacco suspension cell cultures (Allen *et al.*, 1996). This group demonstrated that, using this construct, the overall level of GUS enzyme activity increased by nearly 140-fold relative to lines transformed with the same construct lacking SARs. Enhanced expression levels, although more moderate, have also been observed in plants transformed with SAR-flanked reporter gene

constructs using *Agrobacterium* -mediated transformation (Schöffl *et al.*, 1993; Mlynárová *et al.*, 1994, 1995; van der Geest *et al.*, 1994). By flanking the ends of an oleosin-hirudin fusion gene construct with SARs of tobacco, it might be possible to increase dramatically the level of protein accumulation in transgenic plants.

6.2 Future experiments

One of the challenges associated with achieving high levels of oleosin-hirudin fusion protein is the competition provided by native oleosins for a position on the oil body. One way this may be circumvented is to reduce the overall expression of native oleosins and replace them with the expression of an oleosin fusion gene. Reducing the levels of native oleosin mRNA may be achievable by antisense RNA technology. The *B. napus* clones aligned in Figures 6.1 and 6.2 share a high degree of sequence similarity (approximately 80%). Oleosin genes and cDNA sequences from a variety of plant species were identified through Genbank. Sequences sharing a low degree of identity at the DNA level were sought. A low degree of similarity at the nucleotide level is desirable so that antisense RNA specific to *B. napus* may be expressed without interfering with the accumulation of the heterologous oleosin. Two complete cDNA clones, one from rice (accession number U43930), and the other from sunflower (X78679), met these criteria. Previous studies have shown that heterologous oleosins can successfully target to the oil body membrane of transgenic plants (Sarmiento *et al,*. 1997; Lee *et al.*, 1991), thus the expression and targeting of either sunflower or rice oleosins in transgenic *B. napus* should be accommodated. Figure 6.1 is a DNA sequence alignment of the 5' coding regions of *B. napus* and rice. There is 27% sequence identity between bno20 and rice from bases 1 through 200. Similarly, sunflower and bno20 share only 26% sequence identity within this region (Fig. 6.2). In order for native oleosins to be "knocked out" by antisense RNA and

Figure 6.1 DNA sequence alignment of *B. napus* and rice oleosin coding regions. The 5' region encoding a rice oleosin clone (accession number U43930) is aligned to that of 4 different *B. napus* oleosin clones. The *B. napus* clones BNV (accession number X63779) BnnapII (accession number X58000) are partial clones, while bno20 (accession number M63985) and BnIII (accession number X61937) are full length cDNA clones. The *B. napus* clones share approximately 80% sequence identity. Boxes and bold print indicate nucleotides which are identical.

ClustalW Formatted Alignments

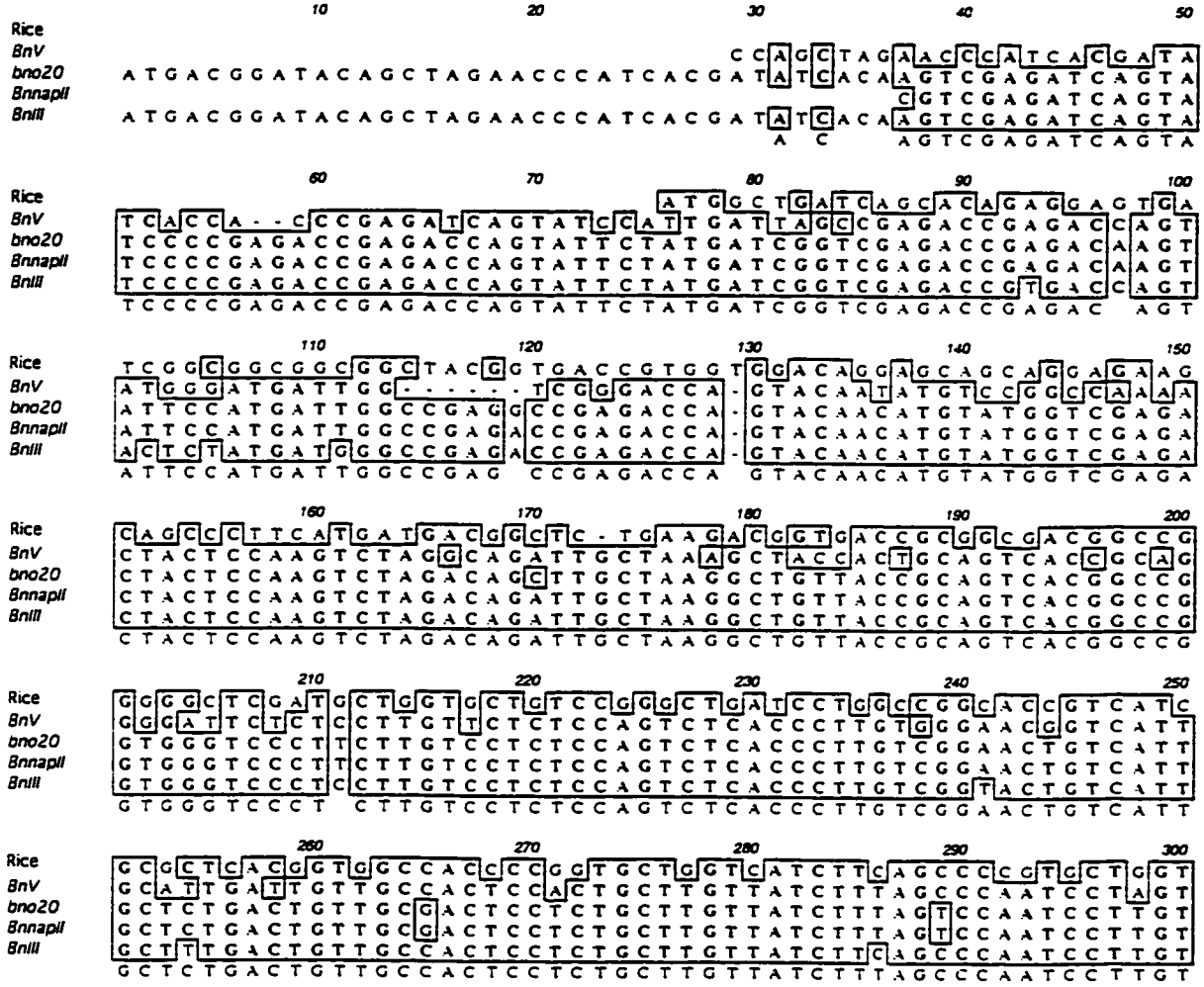
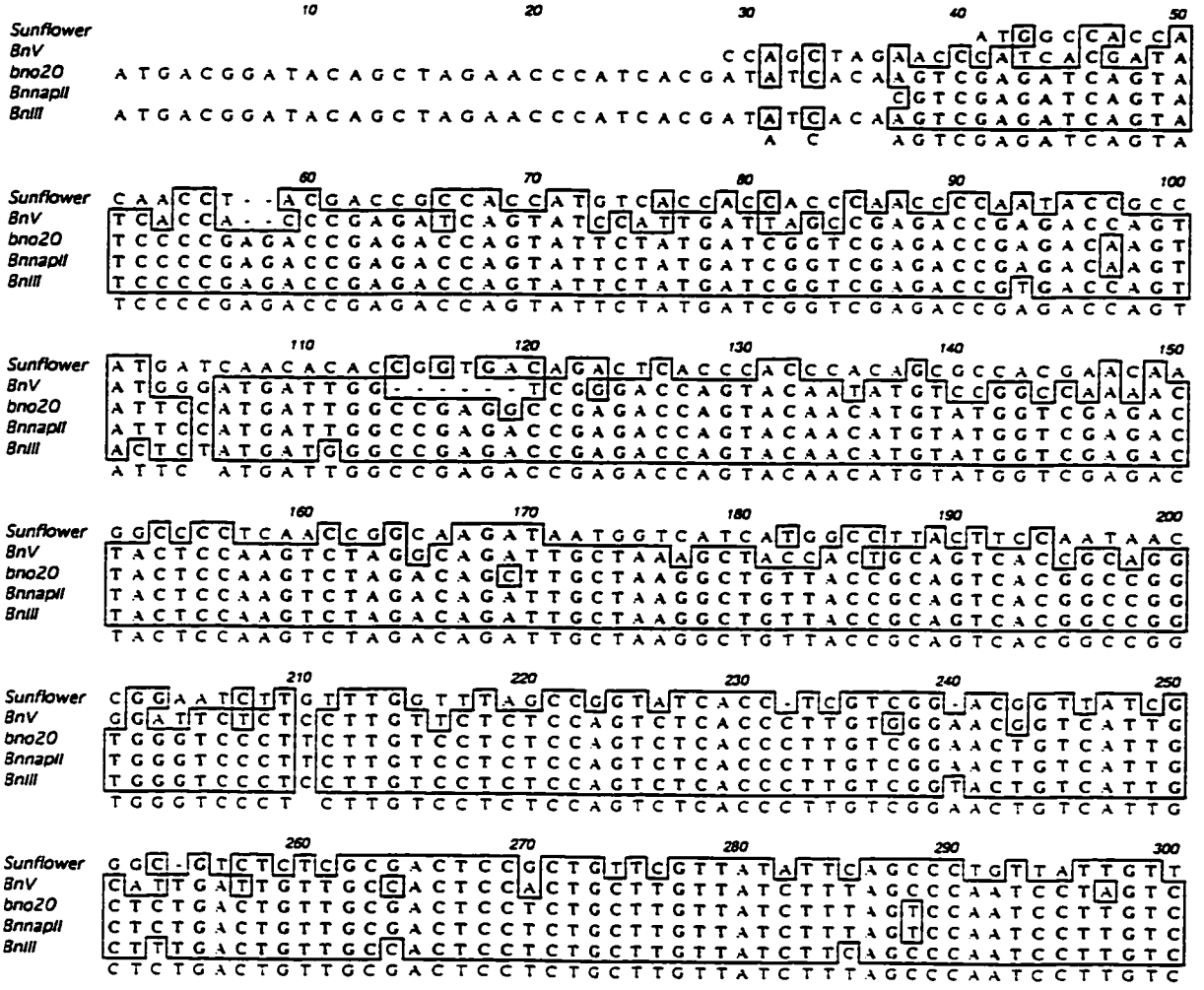


Figure 6.2 DNA sequence alignment of *B. napus* and sunflower oleosin coding regions. The 5' region encoding a sunflower oleosin clone (accession number X78679) is aligned to that of 4 different *B. napus* oleosin clones. The *B. napus* clones BNV (accession number X63779) BnnapII (accession number X58000) are partial clones, while bno20 (accession number M63985) and BnIII (accession number X61937) are full length cDNA clones. The *B. napus* clones share approximately 80% sequence identity. Boxes and bold print indicate nucleotides which are identical.

ClustalW Formatted Alignments



replaced with a heterologous oleosin-hirudin fusion gene in *B. napus*, two constructs would need to be made. One would consist of the antisense version of nucleotides 1 to 200 of the bno20 *B. napus* oleosin clone. This would be under the control of the 2.5kbp *Arabidopsis* oleosin promoter described in Chapter 4, and would be terminated by the nos transcriptional terminator. The second construct would consist of the full length sunflower or rice oleosin 5' untranslated region and coding region followed, in frame, with a gene encoding hirudin. These constructs would be driven by the 2.5kbp *Arabidopsis* oleosin promoter and terminated by a nos transcriptional terminator. Each construct, once introduced into *Agrobacterium*, would be used to transform *B. napus* petioles. Plants harbouring each gene would be selfed to promote homozygosity. Plants harbouring the antisense construct would be cross-pollinated with plants harbouring either the rice oleosin-hirudin or sunflower oleosin-hiridin fusion gene construct. These experiments would determine if native oleosin gene expression can be repressed using antisense technology, and if the rice- or sunflower oleosin-hirudin fusion protein could accumulate and target to higher levels than that demonstrated in this thesis. It is possible that suppression of gene expression using antisense RNA would result in inviable seeds. If this is the case, double transformation events with each plasmid type (facilitated by binary plasmids having genes conferring resistance to different antibiotics) may be of value.

An interesting alternative to using antisense RNA to "knock-out' gene expression is to study the effects of downregulating oleosin expression and accumulation through the use of T-DNA tagged *Arabidopsis* mutants. *Arabidopsis thaliana* is a convenient system for study because it has a short life cycle, is easily transformed, and has a small genome containing little repetitive DNA. The *Arabidopsis* genome and its cDNAs are rapidly being sequenced, and extensive mutant seed libraries are readily available. T-DNA insertional mutagenesis is a well-established experimental procedure, which results in the insertion of foreign DNA

(derived from *Agrobacterium)* into random sites of the plant genome (McKinney *et al.*, 1995). This causes the perturbation of gene expression or protein accumulation in the plant. The small size of the *Arabidopsis* genome, coupled with the efficiency of transformation with T-DNA, results in a very high likelihood of creating a plant line containing a T-DNA insert in an *Arabidopsis* oleosin gene. Seed stocks of thousands of mutant lines are readily available. The identification of a seed line containing a T-DNA insert of the *Arabidopsis* oleosin may be achieved by PCR analysis. Briefly, primers complementary to the top strand of a highly conserved region of oleosin, such as the central domain, are used in conjunction with primers complementary to both strands of the T-DNA insert. Primers complementary to the top strand, and not both strands of the oleosin will be used because an insertion downstream of the central domain will unlikely affect targeting of the oleosin (van Rooijen and Moloney, 1995a). By analyzing the PCR product (via gel electrophoresis), it will be possible to determine if the T-DNA tag lies within the oleosin gene. Once such a mutant has been identified, it will be transformed with an oleosin-hirudin fusion gene construct, such as oleoHV1#1 (section 2.3.1.2). Wild-type *Arabidopsis* will also be transformed with this construct. By comparing the levels of fusion protein accumulation between these two types of transformants, in addition to the T-DNA tagged mutant which does not harbour the oleosin-hirudin fusion gene, it will be possible to determine if the loss of the native oleosin results in heightened accumulation of the fusion protein, and what, if any, phenotypic alteration results from the interruption of the oleosin gene by the T-DNA tag.

Industry typically favors the expression of well characterized systems like *E. coli*, yeast, and Chinese Hamster Ovary cell culture, because their use has already met FDA approval. The cost of changing from older technology to newer production systems can be prohibitive for biotechnology companies. The choice of production system is determined

by a variety of factors such as the properties of protein to be expressed, the purpose for which it is intended, and economic factors which directly relate to its ease of purification and recovery. Although *B. napus* seeds are inexpensive to produce, some of the downstream processes involved in the isolation of hirudin from *B. napus* seed are costly. One is the need to proteolytically release hirudin from its oleosin "carrier". This is a common problem for systems expressing fusion proteins. The need for site-specific proteolysis is particularly problematic, economically, when it is required in large quantities. The cost of proteases may be alleviated with the advent of recombinant DNA technology. Site-specific proteases such as enterokinase and protease 3C have been successfully produced by heterologous expression systems (Collins-Racie *et al,*. 1995; Walker *et al,*. 1994). These proteases were not only shown to be active, but were active as fusion proteins. Thus, it may be possible to express these proteases as oleosin fusion proteins. To test this hypothesis, a gene encoding one of these proteases could be fused in-frame to an oleosin-encoding sequence. This construct would be under the control of the *Arabidopsis* oleosin promoter and terminated by a nos transcriptional terminator. Oil bodies which are derived from plants harbouring this construct would be mixed with oil bodies containing an oleosin fusion gene having a site recognized by the protease interposed between the two protein domains. This system would be advantageous, not only for economic reasons, but because only the protein of interest, and not the protease, would be present in the oil body unternatant.

A number of methods exist by which heterologous proteins and peptides may be produced. These include heterologous systems such as bacteria, yeast, and mammalian cell cultures. As described earlier, these systems incur a number of problems which can interfere with the large-scale production of proteins or peptides of interest. The use of oleosins as carrier proteins for the production of pharmaceuticals, industrial enzymes or

other high-value peptides represents a novel and economically viable method by which such peptide sequences may be produced in large-scale.

Appendix 1. Table 1) Determination of significant difference in thrombin activity between preparations of wild-type and transformed tobacco oil body extracts. All assays were performed in triplicate. Individual pairwise analysis of thrombin activity in three preparations of WT (wild type) or TF (transformed) oil body unternant treated (X) or untreated with Factor Xa. The asterisk (*) indicates samples which are not significantly different using analysis of variance (ANOVA) at 95% confidence. Significant differences are only apparent between TF, X and the other samples assayed.

Appendix 1, Table 1

sample	*Source of Variation*	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*	
WT vs WT.X	Between Groups	0	1	0	0	1	4.493998063	*
	Within Groups	0.000772282	16	4.82676E-05				
WT.X vs TF	Between Groups	4.57291E-06	1	4.57291E-06	0.104807376	0.750328329	4.493998063	*
	Within Groups	0.000698105	16	4.36316E-05				
WT vs TF	Between Groups	4.57291E-06	1	4.57291E-06	0.10715967	0.747645231	4.493998063	*
	Within Groups	0.000682781	16	4.26738E-05				
TF vs TF.X	Between Groups	0.029108004	1	0.029108004	447.2715906	4.03781E-13	4.493998063	
	Within Groups	0.001041265	16	6.5079E-05				

Appendix 2. Table 1) Determination of significant difference in thrombin activity between varying amounts of wild-type and transformed oil body extracts. All assays were performed in triplicate. Individual pairwise analysis between thrombin activity in 20, 40, or 60µl volumes of buffer, WT (wild type) or TF (transformed) oil body unternant treated (X) or untreated with clostripain. The asterisk (*) indicates samples which are not significantly different using analysis of variance (ANOVA) at 99% confidence. As shown, addition of 20, 40, or 60µl of WT, WT,X or TF are not significantly different. Significant differences in thrombin activity was only observed between the addition of 20, 40, and 60µl of TF,X. Table 2) Determination of significant difference in thrombin activity between oil body extracts from wild type and transformed plants treated or untreated with clostripain. Individual pairwise analysis between thrombin activity in 20-60µl of buffer, WT and TF oil body preparations treated or untreated with clostripain. The asterisk (*) indicate samples which are not significantly different using analysis of variance (ANOVA) at 99% confidence. Significant differences are only apparent between TF, X and the other samples assayed.

Appendix 2, Table 1

sample	*Source of Variation*	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*	
buffer: 20ul vs 40ul	Between Groups	9.24822E-05	1	9.24822E-05	2.706751355	0.175269576	21.19759301	*
	Within Groups	0.000136669	4	3.41672E-05				
buffer: 40ul vs 60ul	Between Groups	6.97386E-05	1	6.97386E-05	6.825480302	0.05926027	21.19759301	*
	Within Groups	4.08696E-05	4	1.02174E-05				
buffer: 20ul vs 60ul	Between Groups	0.000322839	1	0.000322839	10.87524272	0.029996244	21.19759301	*
	Within Groups	0.000118743	4	2.96857E-05				
WT: 20ul vs 40ul	Between Groups	1.39467E-06	1	1.39467E-06	0.190247041	0.685219737	21.19759301	*
	Within Groups	2.93233E-05	4	7.33083E-06				
WT: 20ul vs 60ul	Between Groups	0.000112968	1	0.000112968	12.11608785	0.025333171	21.19759301	*
	Within Groups	3.72952E-05	4	9.32381E-06				
WT: 40ul vs 60ul	Between Groups	8.92587E-05	1	8.92587E-05	9.10185812	0.039281667	21.19759301	*
	Within Groups	3.92266E-05	4	9.80665E-06				
WT.X: 20ul vs 40ul	Between Groups	1.29705E-07	1	1.29705E-07	0.014473712	0.910040989	21.19759301	*
	Within Groups	3.58456E-05	4	8.96141E-06				
WT.X: 20ul vs 60ul	Between Groups	3.63239E-05	1	3.63239E-05	1.66396959	0.266592924	21.19759301	*
	Within Groups	8.73186E-05	4	2.18296E-05				
WT.X: 40ul vs 60ul	Between Groups	4.07947E-05	1	4.07947E-05	1.595752468	0.275120491	21.19759301	*
	Within Groups	0.000102258	4	2.55646E-05				
TF: 20ul vs 40ul	Between Groups	1.37742E-05	1	1.37742E-05	0.641199191	0.468139574	21.19759301	*
	Within Groups	8.59279E-05	4	2.1482E-05				
TF: 20ul vs 60ul	Between Groups	1.70912E-05	1	1.70912E-05	0.700068835	0.449834556	21.19759301	*
	Within Groups	9.76543E-05	4	2.44136E-05				
TF: 40ul vs 60ul	Between Groups	1.78746E-07	1	1.78746E-07	0.010635877	0.922823118	21.19759301	*
	Within Groups	6.72239E-05	4	1.6806E-05				
TF.X: 20ul vs 40ul	Between Groups	0.003434385	1	0.003434385	1089.889238	5.02036E-06	21.19759301	
	Within Groups	1.26045E-05	4	3.15113E-06				
TF.X: 40ul vs 60ul	Between Groups	0.000686999	1	0.000686999	272.0112589	7.91421E-05	21.19759301	
	Within Groups	1.01025E-05	4	2.52563E-06				
TF.X: 20ul vs 60ul	Between Groups	0.007193465	1	0.007193465	4722.381408	2.68668E-07	21.19759301	
	Within Groups	6.09308E-06	4	1.52327E-06				

Appendix 2, Table 2

	Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*	
buffer vs WT	Between Groups	2.37142E-05	1	2.37142E-05	0.5747285	0.459407655	8.530946616	*
	Within Groups	0.000660184	16	4.12615E-05				
WT vs WT,X	Between Groups	0.000100327	1	0.000100327	4.548945778	0.048776228	8.530946616	*
	Within Groups	0.00035288	16	2.2055E-05				
WT,X vs TF	Between Groups	1.34671E-05	1	1.34671E-05	0.694385652	0.416943241	8.530946616	*
	Within Groups	0.000310309	16	1.93943E-05				
buffer vs WT,X	Between Groups	0.000221595	1	0.000221595	5.577129391	0.031210999	8.530946616	*
	Within Groups	0.000635724	16	3.97328E-05				
buffer vs TF	Between Groups	0.000262401	1	0.000262401	8.262360407	0.011009591	8.530946616	*
	Within Groups	0.000508138	16	3.17586E-05				
WT vs TF	Between Groups	0.000123063	1	0.000123063	7.877550816	0.012665134	8.530946616	*
	Within Groups	0.000249952	16	1.5622E-05				
TF vs TF,X	Between Groups	0.041152659	1	0.041152659	85.4705944	8.09434E-08	8.530946616	
	Within Groups	0.007703732	16	0.000481483				
WT,X vs TF,X	Between Groups	0.041152659	1	0.041152659	85.4705944	8.09434E-08	8.530946616	
	Within Groups	0.007703732	16	0.000481483				
WT vs TF,X	Between Groups	0.035787211	1	0.035787211	73.91853721	2.1514E-07	8.530946616	
	Within Groups	0.007746303	16	0.000484144				
buffer vs TF,X	Between Groups	0.033968466	1	0.033968466	67.6903115	3.84756E-07	8.530946616	
	Within Groups	0.008029147	16	0.000501822				

References

Abell, B.M., Holbrook, L.A., Abenes, M., Murphy, D.J., Hills, M.J., and M.M. Moloney (1997) Role of the proline knot motif in oleosin endoplasmic reticulum topology and oil body targeting. Plant Cell. **9**: 1481-93.

Abelson, P.H. (1996) Pharmaceuticals based on biotechnology. Science. **13**: 719.

Allen, G.C., Hall, Jr., G., Michalowski, S., Newman, W., Spiker, S., Weissinger, A.K. and W.F. Thompson (1996) High-level transgene expression in plant cells: effects of a strong scaffold attachment region from tobacco. Plant Cell. **8**: 899-913.

Altenbach, S.B., Kuo, C., Staraci, L.C., Pearson, K.W., Wainwright, C., Georgescu, A. and J. Townsend (1992) Accumulation of a brazil nut protein in seeds of transgenic canola results in enhanced levels of seed protein methionine. Plant Mol. Biol. **18**: 235-45.

Altenbach, S.B., Pearson, K.W., Meeker, G., Staraci, L.C. and S.S.M. Sun (1989) Enhancement of the methionine content of seed proteins by the expression of a chimeric gene encoding a methionine-rich protein in transgenic plants. Plant Mol. Biol. **13**: 513-22.

Amersham (1987) Membrane transfer and detection methods. Amersham Canada Ltd.

Anonymous (1997a) Gen's annual top biotech and pharma companies. Gen. Engin. News. **17**: 23.

Anonymous (1997b) Gen's molecular millionaires. Gen. Engin. News. **17**: 10.

Batchelder, C., Ross, J.H.E. and D.J. Murphy (1994) Synthesis and targeting of *Brassica napus* oleosin in transgenic tobacco. Plant Sci. **104**: 39-47.

Bebbington, C. and C. Hentschel (1985) The expression of recombinant DNA molecules in mammalian cells. TiBtech. **3**: 314-318.

Bebbington, C.R., Renner, G., Thomson, S., King, D., Abrams, D. and G.T. Yarranton (1992) High-level expression of a recombinant antibody from myeloma cells using a glutamine synthetase gene as an amplifiable selectable marker. Bio/Tech. **10**: 169-75.

Benatti, L., Scacheri, E., Bishop, D.H.L., and P. Sarmientos (1991) Secretion of biologically active leech hirudin from baculovirus-infected insect cells. Gene. **101**: 255-260.

Bender, E., Vogel, R., Koller, K-P. and J. Engels (1990) Synthesis and secretion of hirudin by *Streptomyces lividans*. Appl. Microbiol. Biotechnol. **34**: 203-207.

Bergfeld, R., Hong, Y-N., Kuhnl, T., and P. Schopfer (1978) Formation of oleosomes (storage lipid bodies) during embryogenesis and their breakdown during seedling development in cotyledons of *Sinapis alba* L. Planta. **143**: 297-307.

Bessen, D. and V.A. Fischetti (1988) Passive acquired mucosal immunity to group A streptococci by secretory immunoglobulin A. J. Exp. Med. **167**: 1945-50.

Bettigole, R.E. (1992) Drugs acting on the blood and blood-forming organs. In: *Textbook of Pharmacology*. C.M. Smith (ed.) pp 784-801, W.B. Saunders Company, Philadelphia, P.A.

Bischoff, R., Clesse, D., Whitechurch, O., Lepage, P., and C. Roitsch (1989) Isolation of recombinant hirudin by preparative high-performance liquid chromatography. J. Chromatogr. **476**: 245-55.

Boothe, J.G., Saponja, J.A., and D.L. Parmenter (1998) Molecular farming in plants: Oilseeds as vehicles for the production of pharmaceutical proteins. Drug Dev. Res. (submitted)

Boyajian, G.E., and L.H. Carreira (1997) Phytoremediation: A clean transition from laboratory to marketplace? Nature Biotech. **15**: 127-28.

Brown, J.E., Baugh, R.F. and C. Hougie (1980) The inhibition of the intrinsic generation of activated factor X by heparin and hirudin. Thrombosis Res. **17**: 267-272.

Buckholz, R.G. and M.A.G. Gleeson (1991) Yeast systems for the commercial production of heterologous proteins. Bio/tech. **9**: 1067-72.

Carter, C.M., Rhee, K.C., Hagenmaier, R.D. and K.F. Mattil (1974) Aqueous extraction-an alternative oilseed milling process. J. Am. Oil Chem. Soc. **51**: 137-41.

Chang, J. (1983) The functional domain of hirudin, a thrombin specific inhibitor. FEBS Lett. **164**: 307-313.

Chang, J-Y. (1991) Stability of hirudin, a thrombin-specific inhibitor. J. Biol. Chem. **266**: 10839-43.

Chatrenet, B., and J-Y Chang (1992) The folding of hirudin adopts a mechanism of trial and error. J. Biol. Chem. **267**: 3038-43.

Chatrenet, B., and J-Y Chang (1993) The disulfide folding pathway of hirudin elucidated by stop/go folding experiments. J. Biol. Chem. **268**: 20988-20996.

Chaudhary, S., Parmenter, D.L. and M.M. Moloney (1998) Transgenic *Brassica carinata* as a vehicle for the production of recombinant proteins in seeds. Plant Cell Rep. in press.

Chrispeels, M.J. (1984) Biosynthesis, processing and transport of storage proteins and lectins in cotyledons of developing legume seeds. Philos.Trans. R. Soc. Lond. **304**: 309-322.

Chrispeels, M.J., Higgins, T.J.V. and D. Spencer (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.

Close, P. Bichler, J., Kerry, R., Ekman, S., Bueller, H.R., Kienast, J., Marbet, G.A., Schramm, W., and M.D. Verstraete (1994) Weak allergenicity of recombinant

hirudin, CGP 39393 in immunocompetent volunteers. Coronary Artery Dis. **5**: 943-49.

Collins-Racie, L.A., McColgan, J.M., Grant, K.L., DiBlasio-Smith, E.A., McCoy, J.M. and E.R. LaVallie (1995) Production of recombinant bovine enterokinase catalytic subunit in *Esherichia coli* using the novel secretory fusion partner DsbA. Bio/Tech. **13**: 982-87.

Conceicao, A. da S., Van Vliet, A. and E. Krebbers (1994) Unexpectedly higher expression of a chimeric 2S albumin seed protein transgene from a tandem array construct. Plant Mol. Biol. **26**: 1001-05.

Cornelissen, B.J.C., Hooft van Huijsduijnen, R.A.M., and J.F. Bol (1986) A tobacco mosaic virus-induced tobacco protein is homologous to the sweet-tasting protein thaumatin. Nature **321**: 531-32.

Cummins, P.J., Hills, M.J., Ross, J.H.E., Hobbs, D.H., Watson, M.D., and D.J. Murphy (1993) Differential, temporal and spatial expression of genes involved in storage oil and oleosin accumulation in developing rapeseed embryos: implications for the role of oleosins and the mechanisms of oil-body formation. Plant Mol. Biol. **23**: 1015-27.

Da Silva, N.A., and J.E. Bailey (1991) Influence of plasmid origin and promoter strength in fermentations of recombinant yeast. Biotechnology and Bioengineering. **37**: 318-324.

Dalsgaard, K., Uttenhal, A., Jones, T.D., Xu, F., Merryweather, A., Hamilton, W.D.O., Langeveld, J.P.M., Boshuizen, R.S., Kamstrup, S., Lomonossoff, G.P., Porta, C., Vela, C., Casal, J.I., Meloen, R.H. and P.B. Rodgers (1997) Plant-derived vaccine protects target animals against a viral disease. Nature Biotech. **15**: 248-52.

Datar, R.V., Cartwright, T. and C-G. Rosen (1993) Process economics of animal cell and bacterial fermentations: a case study analysis of tissue plasminogen activator. Bio/Tech. **11**: 349-56.

Datla, R.S.S., Bekkaoui, F.,. Hammerlindl, J.K., Pilate, G., Dunstan, D.I., and W.L. Crosby (1993) Improved high-level constitutive foreign gene expression in plants uning an AMV RNA4 untranslated leader sequence. Plant Sci. **94**: 139-149.

Davis, L.G., Dibner, M.D., and J.F. Battey (1986) *Basic methods on molecular biology.* Elsevier, N.Y. USA p. 123-24.

Dellaporta, S.L., Wood, J., and J.B. Hicks (1983). A plant DNA minipreparation : version II. Plant Mol. Biol. Rep. **1**: 19-23

De Nobel, J.G. and J.A. Barnett (1991) Passage of molecules through yeast cell walls: a brief essay-review. Yeast. **7**: 313-23.

Dickman, S. (1996) Germany joins the biotech race. Science. **274**: 1454-55.

Dodt, J. (1995) Anticoagulatory substances of bloodsucking animals: from hirudin to hirudin mimetics. Angew. Chem. Int. Ed. Engl. **34**: 867-80.

Dodt, J., Muller, H.P., Seemuller, U. and J-Y Chang. (1984). The complete amino acid sequence of hirudin, a thrombin specific inhibitor. FEBS Letters. **165**: 180-183.

Dodt, J., Schmitz, T., Schafer, T. and C. Bergmann (1986) Expression, secretion and processing of hirudin in E. coli using the alkaline phosphatase signal sequence. FEBS Lett. **202**: 373-377.

Dower, W.J., Miller, J.F., and C.W. Ragsdale (1988) High efficiency transformation of *E coli* by high voltage electroporation. Nucl. Acids Res. **16**: 6125-6145.

Edgington, S., Francisco, M., Johnson, E., Marshall, A., Nasto, B. and G Sinha (1997) Boehringer Mannheim's new clot buster. Nature Biotech. **15**: 114.

Echelard, Y. (1996) Recombinant protein production in transgenic animals. Curr. Opin. Biotechnol. **7**: 536-40.

Erlich, H.A. (ed.) (1989). *PCR Technology- Principles and Applications for DNA Amplification* Stockton Press, New York.

Eskin, N.A.M., McDonald, B.E., Przybylski, L.J., Scarth, R., Mag, T., Ward, K. and D. Adolph (1996) Canola oil. In *Edible Oil and Fat Products: Oils and Oil Seeds.* Y.H. Hui (ed). John Wiley and Sons Inc. pp. 14-21.

Faber, K.N., Harder, W., Ab, G., and M. Veenhuis (1995) Methylotrophic yeasts as factories for the production of foreign proteins. Yeast. **11**: 1331-44.

Feinberg, A.P. and B. Vogelstein (1984) A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. Addendum. **137**: 266-67.

Fenton, J.W., Fasco, M.J., Stackrow, A.B., Armon, D., Young, A.M., and J. Findlayson (1977) Human thrombins. J. Biol. Chem. **252**: 3587-3589.

Fenton, J.W., Olson, T.A., Zabinski, M.P., and G.D. Wilner (1988) Anion-binding exosite of human α–thrombin and fibrin(ogen) recognitions. Biochemistry. **27**:7106-7112.

Ferrari, E., Jarnagin, A.S. and B.F. Schmidt (1993) Commercial production of extracellular enzymes. In: *Bacillus subtilis and Other Gram-Positive Bacteria: Biochemistry, Physiology, and Molecular Genetics.* A.L. Sonenshein, J.A. Hoch, and R. Losick (eds.) pp 917-37, American Society for Microbiology. Washington, DC.

Fortkamp, E., Rieger, M., Heisterberg-Moutses, G., Schweitzer, S., and R. Sommer (1986) Cloning and Expression in *Escherichia coli* of a synthetic DNA for hirudin, the blood coagulation inhibitor in the leech. DNA. **5**: 511-517.

Frey-Wyssling, A.E., Grieshaber, E., and K. Muhlethaler (1963) Origin of spherosomes in plant cells. J. Ultrastruct. Res. **8**: 506-16.

Gallie, D.R., Sleat, D.E., Watts, J.W., Turner, P.C. and T.M.A. Wilson (1987). A comparison of eukaryotic viral 5'-leader sequences as enhancers of mRNA expression *in vivo*. Nucl. Acids Res. **15**: 8693-709.

Goddijn, O.J.M and J. Pen (1995) Plants as bioreactors. Tibtech. **13**: 379-87.

Goeddel, D.V., Heyneker, H.L., Hozumi, T., Arentzen, R., Itakura, K., Yansura, D.G., Ross, M.J., Miozzari, G., Crea, R., and P.H. Seeburg (1979) Expression in *Escherichia coli* of a chemically synthesized gene for human insulin. Nature. **281**: 544-548.

Goeddel, D.V., Yelverton, E., Ullrich, A., Heyneker, H.L., Miozzari, G., Holmes, W., Seeburg, P.H., Dull, T., May, L., Stebbing, N., Crea, R., Maeda, S., McCandlliss, R., Sloma, A., Tabor, J.M., Gross, M., Familettis, P.C., and S. Pestka (1980) Human leukocyte interferon produced by *E. coli* is biologically active. Nature. **287**: 411-416.

Goff, S.A., and A.L. Goldberg (1985) Production of abnormal proteins in *E. coli* stimulates transcription of *lon* and other heat shock proteins. Cell. **41**: 587-595.

Goff, S.A., and A.L. Goldberg (1987) An increased content of protease La, the *lon* gene product, increases protein degradation and blocks growth in *E. coli*. J. Biol. Chem. **262**: 4508-4515.

Goldberg, R.B., Barker, S.J. and Perez-Grau (1989) Regulation of gene expression during plant embryogenesis. Cell. **56**: 149-60.

Gosse, M.E. and T.F. Nelson (1997) Approval times for supplemental indications for recombinant proteins. Nature Biotech. **15**: 130-134.

Gossen, M.F.A. (1992) Large-scale insect cell culture. Curr. Opin. Biotech. **3**: 99-104.

Grinna, L.S. and J.F. Tschopp (1989) Size distribution and general structural features of N-linked oligosaccharides from the methylotrophic yeast, *Pichia pastoris.* Yeast **5**: 107-115.

Grosveld, F., van Assendelft, G.B., Greaves, D.R. and G. Kollias (1987) Position-independent, high-level expression of the human β-globin gene in transgenic mice. Cell. **51**: 975-85.

Gruenwald, M.D. and M.S. Heitz (1996) *Baculovirus Expression Vector System: Procedures and Methods Manual.* Pharmingen, San Diego, CA.

Grutter, M.G., Priestle, J.P., Rahuel, J. Grossenbacher, H., Bode, W., Hofsteenge, J., and S.R. Stone (1990) Crystal structure of the thrombin-hirudin complex: A novel mode of serine protease inhibition. EMBO J. **9**: 2361-2365.

Hanahan, D. (1983). Studies on transformation of *E. coli* with plasmids. J. Mol. Biol. **166**: 557-560.

Harvey, R.P., Degryse, E., Stefani, L., Schamber, F., Cazenave, J.P., Courtney, M., Tolstoshev, P. and J.P. Lecocq. (1986). Cloning and expression of cDNA coding for anticoagulant hirudin from blood-sucking leech, *Hirudo medicinalis.* Proc. Natl. Acad. Sci. USA. **83**: 1084-1088.

Haq, T.A., Mason, H.S., Clements, J.D. and C.J. Arntzen (1995) Oral immunization with a recombinant bacterial antigen produced in transgenic plants. Science. **268**: 714-16.

Heinrikson, R.L. and F.J. Kezdy (1991) Introduction. In: *Purification and Analysis of Recombinant proteins.* R. Seethram and S.K. Sharma (eds.) p. v-x. Marcel Dekker, Inc., New York.

Heinrikson, R.L. and A.G. Tomasselli (1991) Purification and characterization of recombinant proteins. In: *Purification and Analysis of Recombinant proteins.* R. Seethram and S.K. Sharma (eds.) p. 3-28, Marcel Dekker, Inc., New York.

Hensing, M.C.M., Rouwenhorst, R.J., Heijnen, J.J., van Dijken, J.P. and J.T. Pronk (1995) Physiological and technological aspects of large-scale heterologous-protein production with yeasts. Antonie van Leewenhoek. **67**: 261-79.

Herman, E.M. (1987) Immunogold localization and synthesis of an oil-body membrane protein in developing soybean seeds. Planta **172**: 336-45.

Herman, E.M., Shannon, L.M. and M.J. Chrispeels (1986) The Golgi apparatus mediates the transport and post-translational modification of protein body proteins. In . *Molecular Biology of Seed Storage proteins and lectins.* L.M. Shannon and M.J. Chrispeels (eds.) The American Society of Plant Physiologists, Rockville, Md. pp. 163-173.

Hiatt, A. (1990) Antibodies produced in plants. Nature. **344**: 469-70.

Hiatt, A., Cafferkey,R. and K. Bowdish (1989) Production of antibodies in transgeneic plants. Nature. **342**: 76-78.

Hiei, Y., Ohte, S., Komari, T., and T. Kumashiro (1994) Efficient transformation of rice (*Oryza sativa* L.) mediated by *Agrobacterium* and sequence analysis of the boundaries of the T-DNA. Plant J. **3**: 271-82.

Hirsh, J. (1991) Heparin. New Engl. J. Med. **324**: 1565-74.

Hodgson, J. (1993) Expression systems: a user's guide. Bio/tech. **11**: 887-93.

Hoffman, L.M., Donaldson, D.D. and E.M. Herman (1988) A modified storage protein is synthesized, processed, and degraded in the seeds of transgenic plants. Plant Mol. Biol. **11**: 717-729.

Hogg, P.J. and C.M. Jackson (1989) Fibrin monomer protects thrombin from inactivation by heparin-antithrombin III: implications for heparin efficacy. Proc. Natl. Acad. Sci. USA **86**: 3619-23.

Holbrook, L.H., van Rooijen, G.J.H., Wilen, R.W. and M.M. Moloney (1991) Oil body proteins in microspore derived embryos of *Brassica napus*: Hormonal, osmotic and developmental regulation of synthesis. Plant Physiol. **97**: 1051-58.

Hood, E.E., Helmer, G.L., Fraley, R.T., and M-D Chilton (1986) The hypervirulence of *Agrobacterium tumefaciens* A281 is encoded in a region of pTiBO542 outside of T-DNA. J. Bacteriol. **168**: 1291-1301.

Horsch, R.B., Fry, J.E., Hoffman, N.L., Wallroth, M., Eichholtz, D., Rogers, S.G., and R.T. Fraley (1985) A simple and general method for transferring genes into plants. Science **227**: 1229-1231.

Itakura, K., Hirose, T., Crea, R., Riggs, A.D., Heyneker, H.L., Bolivar, F., and H.W. Boyer (1977) Expression in *Escherichia coli* of a chemically synthesized gene for human somatostatin. Science. **198**: 1056-63.

Jobling, S.A. and L. Gehrke (1987) Enhanced translation of chimaeric messenger RNAs containing a plant viral untranslated leader sequence. Nature **325**: 622-25.

Johnson, P.H. (1994) Hirudin: Clinical potential of a thrombin inhibitor. Annu. Rev. Med. **45**: 165-77.

Kaiser, B., Simon, A. and F. Markwardt (1990) Antithrombotic effects of recombinant hirudin in experimental angioplasty and intravascular thrombolysis. Thromb. Haemost. **63**: 44-47.

Keddie, J.S., Hubner, G., Slocombe, S.P., Jarvis, R.P., Cummins, I., Edwards, E-W., Shaw, C.H. and D.J. Murphy (1992) Cloning and characterisation of an oleosin gene from *Brassica napus*. Plant Mol. Biol. **19**: 443-53.

Keeler, S.J., Maloney, C.L., Webber, P.Y., Patterson, C., Hirata, L.T., Falco, S.C. and J.A. Rice (1997) Expression of *de novo* high-lysine α-helical coiled-coil proteins

may significantly increase the accumulation levels of lysine in mature seeds of transgenic tobacco plants. Plant Mol. Biol. **34**: 15-29.

Kennedy, E.P. (1961) Biosynthesis of complex lipids. Proc. Fed Am. Soc. Exp. Biol. **20**: 934-40.

King, L.A. and R.D. Possee (1992) *The Baculovirus Expression System, A Laboratory Guide.* Chapman and Hall, London.

King, D.J., Walton, E.F. and G.T. Yarranton (1989) The production of proteins and peptides from *Saccharomyces cerevisiae.* In: *Molecular and Cellular Biology of Yeasts.* E.F. Walton and G.T. Yarranton (eds.) pp.107-133, Blackie and Son Ltd., New York.

Kjerrulf, M., Lowenadler, B., Svanholm, C. and N. Lycke (1997) Tandem repeats of T helper epitopes enhance immunogenicity of fusion proteins by promoting processing and presentation. Mol. Immunol. **34**: 599-608.

Klee, H., Horsch, R., and S. Rogers (1987) *Agrobacterium*-mediated plant transformation and its further applications to plant biology. Annu. Rev. Plant Physiol. **38**: 467-86.

Klocking, H-P. (1991) Toxicology of hirudin. Sem. Thromb. Hemost. **17**: 126-129.

Klocking, H-P., Guttner, J. and F. Fink (1988) Toxicological studies with recombinant hirudin. Folia Haematol. **115**: 75-82.

Krebbers, E., Herdies, L., De Clercq, A., Leemans, J. Van Damme, J., Segura, M., Gheysen, G., Van Montagu, M. and J. Vandekerckhove (1988) Determination of the processing site of an Arabidopsis 2S albumin and characterization of the complete gene family. Plant Physiol. **87**: 859-866.

Kühnel, B., Holbrook, L.A., Moloney, M.M., and G.J.H. van Rooijen (1996) Oil bodies of transgenic *Brassica napus* as a source of immobilized β-glucuronidase. JAOCS. **73**: 1533-38.

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

Laroche, Y., Storme, V., DeMeutter, J.,Messens, J. and M. Lauwereys (1994) High-level secretion and very efficient isotopic labeling of tick anticoagulant peptide (TAP) expressed in the methylotrophic yeast, *Pichia pastoris.* Bio/Tech. **12**: 1119-24.

Lee, K., and A.H.C. Huang. (1991). Genomic nucleotide sequence of a *Brassica napus* 20kDa oleosin gene. Plant Physiol. **96**: 1395-1397.

Lee, W.S., Tzen, J.T.C., Kridl., J.C., Radke, S.F., and A.H.C Huang (1991) Maize oleosin is correctly targeted to seed oil bodies in *Brassica napus* transformed with the maize oleosin gene. Proc. Natl. Acad. Sci. **88**: 6181-85.

Lehman, E.D., Joyce, J.G., Bailey, F.J., Markus, H.Z., Schultz, L.D., Dunwillei, C.T., Jacobson, M.A. and W.J. Miller (1993) Expression, purification and characterization of multigram amounts of a recombinant hybrid HV1-HV2 hirudin variant expressed in *Saccharomyces cerevisiae.* Prot. Expr. Purif. **4**: 247-255.

Lloyd-Evans, M.L.P. and P. Barfoot (1996) EU boasts good science base and economic prospects for crop biotechnology. Genet. Engin. News. **16**: 16.

Loison, G., Findeli, A., Bernard, S., Nguyen-Juilleret, M., Marquet, M., Riehl-Bellon, N., Carvallo, D., Guerra-Santos, L., Brown, S.W., Courtney, M., Roitsch, C. and Y. Lemoine (1988) Expression and secretion in *S. cerevisiae* of biologically active leech hirudin. Bio/Tech. **6**: 72-77.

Luckow, V.A. and M.D. Summers (1988) Trends in the development of a baculovirus expression system. Bio/Tech. **6**: 47-55.

Ma, J.K-C. and M.B. Hein (1996) Antibody production and engineering in plants. Ann. N.Y. Acad. Sci. **792**: 72-81.

Ma, J.K-C., Hunjan, M., Smith, R., Kelly, C., and T. Lehner (1990) An investigation into the mechanism of protection by local passive immunization with monoclonal antibodies against *Streptococcus mutans*. Infect. Immun. **58**: 3407-14.

Makrides, S.C. (1996) Strategies for achieving high-level expression of genes in *Eschericia coli*. Microbiol. Rev. **60**: 512-38.

Mann, K.G. (1987) The assembly of blood clotting complexes on membranes. Trends Biol. Sci. **12**: 229-233.

Markwardt, F. (1956) Die Isolierung und chemische Charakterisierung des Hirudins. Hoppe Seylers Z Physiol. Chem. **308**: 147-156.

Markwardt, F. (1970) Hirudin as an inhibitor of thrombin. Meth. Enz. **19**: 924-932.

Markwardt, F. (1994) The development of hirudin as and antithrombin drug. Thrombosis Res. **74**: 1-23.

Markwardt, F, Hauptmann, J., Nowak, G., Kleβen, C. and P. Walsmann (1982) Pharmacological studies on the antithrombotic action of hirudin in experimental animals. Thromb. Haemost. **47**: 226-29.

Markwardt, F., Nowak, G. and J. Sturzebecher (1991) Clinical pharmacology of recombinant hirudin. Haemostasis (Suppl. 1) **21**: 133-136.

Markwardt, F. and P. Walsmann (1967) Reindarstellung und analyse des thrombin-inhibitors hirudin. Hoppe-Sylers Z. Physiol. Chem. **348**: 1381-86.

McBride, K.E., and K.R. Summerfelt (1990) Improved binary vector for *Agrobacterium*-mediated plant transformation. Plant Mol Biol **14**: 269-276.

McElroy, D. (1996) The industrialization of plant transformation. Nature Biotech. **14**: 715-716.

McMaster, D. (1993) University of Calgary, Dept. Medical Biochemistry. Personal Communication.

McKinney, E.C., Ali, N., Traut, A., Feldmann, K.A., Belostotsky, D.A., McDowelll, J.M. and R.B. Meagher (1995) Sequence-based identification of T-DNA insertion mutations in *Arabidopsis:* actin mutants act2-1 and act4-1. Plant J. **8**: 613-22.

McPherson, A. (1982) *Preparation and Analysis of Protein Crystals.* p. 77 John Wiley and Sons, New York.

Mendoza-Vega, O., Hebert, C., and S.W. Brown (1994) Production of recombinant hirudin by high density fed-batch cultivations of a *Saccharomyces cerevisiae* strain: physiological considerations during the bioprocess design. J. Biotechnol. **32**: 249-59.

Mlynárová, L., Loonen, A., Heldens, J., Jansen, R.C., Keizer, P., Steikema, W.J. and J-P. Nap (1994) Reduced position effect in mature transgenic plants conferred by the chicken lysozyme matrix-associated region. Plant Cell. **6**: 417-26.

Mlynárová, L., Jansen, R.C., Conner, A.J., Steikema, W.J. and J-P. Nap (1995) The MAR-mediated reduction in position effect can be uncoupled from copy-number-dependent expression in transgenic plants. Plant Cell. **7**: 599-609.

Mohankumar, C., Arumughan, C., and R.K. Raj (1990) Histological localization of oil plam fruit lipase. J. Am. Oil. Chem. Soc. **10**: 665-69.

Moloney, M.M., Walker, J.M., and K.K. Sharma (1989) High efficiency transformation of *Brassica napus* using *Agrobacterium* vectors. Plant Cell Reports **8**: 238-242.

Murashige, T. and F. Skoog (1962) A revised medium for rapid growth and bioassays with tobaco tissue cultures. Plant Physiol. **15**: 473-97.

Murphy, D.J. (1993) Structure, function and biogenesis of storage lipid bodies and oleosins in plants. Prog. Lipid Res. **32**: 247-280.

Murphy, D.J. (1996) Engineering oil production in rapeseed and other oil crops. Tibtech. **14**: 206-13.

Murphy, D.J., and I. Cummins (1989) Purification and immunogold localization of the major oil-body membrane protein of oilseed rape. Plant Sci. **60**: 47-54.

Murphy, D.J., Cummins, I., and A.S. Kang (1989) Synthesis of the major oil-body membrane protein in developing rapeseed (*Brassica napus*) embryos. Biochem. J. **258**: 285-93.

Murphy, D.J., Keen, J.N., O'Sullivan, J.N., Au, D.M.Y., Edwards, E.W., Jackson, P.J., Cummins, I., Gibbons, T., Shaw, C.H., and A.J. Ryan (1991) A class of amphiphathic proteins associated with lipid storage bodies in plants. Possible similarities with animal serum apolipoproteins. Biochim. Biophys. Acta. **1088**: 86-94.

Nawrath, C., Poirier, Y. and C.R. Somerville (1994) Targeting of the polyhydroxybutyrate biosynthetic pathway to the plastids of *Arabidopsis thaliana* results in high levels of polymer accumulation. Proc. Natl. Acad. Sci. **91**: 12760-64.

Oerke, E.C. (1994) Estimated crop losses due to pathogens, animal pests and weeds. In: *Crop production and crop protection: estimated losses in major food and cash crops.* Oerke, E.C., Dehne, H.W., Schonbeck, F. and A. Weber (eds.) pp 72-88. Elsevier. Amsterdam.

Parcy, F., Valon, C., Raynal, M., Gaubier-Comella, P., Delseny, M. and J. Giraudat (1994) Regulation of gene expression programs during *Arabidopsis* seed development: roles of the ABI3 locus and of endogenous abscisic acid. Plant Cell **6**: 1567-82.

Parmenter, D.L., Boothe, J.G., and M.M. Moloney (1996) Production and purification of recombinant hirudin from plant seeds. In *Transgenic plants: a production system for industrial and pharmaceutical proteins.* Owen, M.R.L. and J Pen (eds). John Wiley and Sons, Toronto pp 280-261.

Parmenter, D.L., Boothe, J.G., van Rooijen, G.J.H., Yeung, E.C., and M.M. Moloney. (1995) Production of biologically active hirudin in plant seeds using oleosin partitioning. Plant Mol. Biol. 29: 1167-1180.

Patterson, R.M., Selkirk, J.K. and B.A. Merrick (1995) Baculovirus and insect cell gene expression: review of baculovirus biotechnology. Envir. Health Persp. **103**: 756-59.

Pen, J., Verwoerd, T.C., Van Paridon, P.A., Beudeker, R.F., Van den Elzen, P.J.M., Gearse, K., Van der Klis, J.D., Versteegh, H.A.J., Van Ooyen, A.J.J. and A. Hoekema (1993) Phytase-containing transgenic seeds as a novel feed additive for improved phosphorous utilization. Bio/Tech. **11**: 811-14.

Persidis, A. and A. Persidis (1996) Biotechnology consortia versus multifirm alliances: Paradigm shift at work? Nature Biotech. **14**: 1657-60.

Plant, A.L., van Rooijen, G.J.H., Anderson, C.P. and M.M.Moloney (1994) Regulation of an *Arabidopsis* oleosin gene promoter in transgenic *Brassica napus*. Plant Mol. Biol. **25**: 193-205.

Poirier, Y., Dennis, D.E., Lomparens, K., and C.R. Somerville (1992) Production of polyhydroxybutyrate, a biodegradable thermoplastic, in higher plants. Science. **256**: 520-23.

Qu, R., Wang, S., Lin, Y., Vance, V. and A. Huang (1986) Characteristics and biosynthesis of membrane proteins of lipid bodies in the scutella of maize. Biochem. J. **235**: 57-65.

Radke, S.E., B.M. Andrews, M.M. Moloney, M.L. Crouch, J.C. Kridl, and V.C. Knauf. (1988). Transformation of Brassica napus L using *Agrobacterium tumefaciens*: developmentally regulated expression of a reintroduced napin gene. Theor. Appl. Genet. **75**: 685-694.

Ridder, R.M., Schmit, R., Legay, F. and H. Gram (1995) Generation of rabbit monoclonal antibody fragments from a combinatorial phage display library and their production in the yeast *Pichia pastoris*. Bio/Tech. **13**: 255-60.

Riehl-Bellon, N., Carvallo, D., Acker, M., van Dorsselaer, A., Marquet, M., Loison, G., Lemoine, Y., Brown, S.W., Courtney, M., and C. Roitsch (1989) Purification and biochemical characterization of recombinant hirudin produced by *Saccharomyces cerevisiae*. Biochemistry. **28**: 2941-2949.

Ross, J.H.E., Sanchez, J., Millan, F. and D.J. Murphy (1993) Differential presence of oleosins in oleogenic seed and mesocarp tissues in olive *(Olea europaea)* and avocado *(Persea americana)*. Plant Sci. **93**: 203-10.

Rugh, C.L., Wilde, H.D., Stack, N.M., Thompson, D.M., Summers, A.O. and R.B. Meagher (1996) Mercuric ion reduction and resistance in transgenic *Arabidopsis thaliana* plants expressing a modified bacterial merA gene. Proc. Natl. Acad. Sci. USA. **93**: 3182-87.

Ryan, E.T., Butterton, J.R., Zhang, T., Baker, M.A., Stanley, S.L.Jr. and S.B. Calderwood (1997) Oral immunization with attenuated vaccine strains of *Vibrio cholerae* expressing a dodecapeptide repeat of the serine-rich *Entamoeba histolytica* protein fused to the cholera toxin B subunit induces systemic and mucosal antiamebic and anti-*V. cholera* antibody responses in mice. Infect. Immun. **65**: 3118-25.

Rydel, T.J., Ravichandran, K.G. Tulinsky, A., Bode, W., Huber, R., Roitsch, C., and J.W. Fenton II (1990) The structure of a complex of recombinant hirudin and human α-thrombin. Science. **249**: 277-80.

Sarmiento, C., Ross, J.H.E., Herman, E. and D.J. Murphy (1997) Expression and subcellular targeting of a soybean oleosin in transgenic rapeseed. Implications for the mechanism of oil-body formation in seeds. Plant J. **11**: 783-96.

Sambrook, J., Fritsch, E.F. and T. Maniatis (1989) *Molecular cloning - a laboratory manual*, 2nd ed. Cold Springs Harbor Laboratory Press, U.S.A.

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

Schagger, H., and G. Von Jagow (1987) Tricine-sodium dodecyl sulfate-polyacrylamide gel electrophoresis for the seaaration of proteins in the range from 1 to 100 kDa. Anal. Biochem. **166**: 368-379.

Schimke, R.T. (1988) Gene amplification in cultured cells. J. Biol Chem. **263**: 5989-92.

Schöffl, F., Schröder, G., Kleim, M. and M. Rieping (1993) An SAR sequence containing 395 bp DNA fragment mediates enhanced, gene-dosage-correlated expression of a chimaeric heat shock gene in transgenic tobacco plants. Transgenic Res. **2**: 93-100.

Schwarzenbach, A.M. (1971) Observations on spherosomal membranes. Cytobiologie **4**: 415-17.

Sherwood, R. (1991) Protein fusions: bioseparation and application.Tibtech. **9**: 1-3.

Sijmons, P.C., Dekker, B.M., Schrammeijer, B., Verwoerd, T.C., Van der Elzen, P.J.M., and A. Hoekema (1990) Production of correctly processed human serum albumin in transgenic plants. Biotechnology. **8**: 217-221.

Simonsen, C.C. and M. McGrogan (1994) The molecular biology of production cell lines. Biologicals. **22**: 85-94.

Slocombe, S.P., Cummins, I, Jarvis, R.P. and D.J. Murphy (1992) Nucleotide sequence and temporal regulation of a seed-specific *Brassica napus* cDNA encoding a stearoyl-acyl carrier protein (ACP) desaturase. Plant Mol. Biol. **20**: 151-55.

Stayton, M., Harpster, M., Brosio, P. and P. Dunsmuir (1991) High-level, seed-specific expression of foreign coding sequences in *Brassica napus*. Aust. J. Plant Physiol. **18**: 507-17.

Stone, S.R. and J. Hofsteenge (1986) Kinetics of the inhibition of thrombin by hirudin. Biochem. **25**: 4622-28.

Sudbery, P.E. (1996) The expression of recombinant proteins in yeasts. Curr. Opin. Biotech. **7**: 517-24.

Thanavala, Y., Yang, Y-F., Lyons, P., Mason, H.S., and C. Arntzen (1995) Immunogenicity of transgenic plant-derived hepatitis B surface antigen. Proc. Nal. Acad. Sci. USA. **92**: 3358-61.

Thomas, J.A. (1995) Recent developments and perspectives of biotechnology-derived products. Toxicol. **105**: 7-22.

Ting, J.T.L., Balsamo, R.A., Ratnayake, C., and A.H.C. Huang (1997) Oleosin of plant seed oil bodies is correctly targeted to the lipid bodies in transformed yeast. J. Biol. Chem. **272**: 3699-06.

Turner, R., and G.D. Foster (1995) The potential exploitation of plant viral translational enhancers in biotechnology for increased gene expression. Mol. Biotech. **3**: 225-236.

Tzen, J.T.C and A.H.C Huang (1992) Surface structure and properties of plant seed oil bodies. J. Cell Biol. **117**: 327-35.

Tzen, J.T.C., Lie, G.C., and A.H. Huang (1992) Characterization of the charged components and their topology on the surface of plant seed oil bodies. J. Biol. Chem. **267**: 15626-34.

Tzen, J.T.C., Cao, Y-z., Laurent, P., Ratnayake, C., and A.H. Huang (1993) Lipids, proteins, and structure of seed oil bodies from diverse species. Plant Physiol. **101**: 267-76.

Urlaub, G. and L.A. Chasin (1980) Isolation of Chinese hamster cell mutants deficient in dihydrofolate reductase activity. Proc. Natl. Acad. Sci. **77**: 4216-20.

van der Geest, A.H.M., Hall, G.E., Spiker, S., and T.C. Hall (1994) The β-phaseolin gene is flanked by matrix attachment regions. Plant J. **6**: 413-23.

van Rooijen, G.J.H. (1993) Molecular biolody of oil body proteins in the *Brassicaceae* : structure, function, and biotechnological applications. PhD thesis. University of Calgary.

van Rooijen, G.J.H. and M.M. Moloney (1995a) Structural requirements of oleosin domains for subcellular targeting to the oil body. Plant Physiol. **109**: 1353-61.

van Rooijen, G.J.H. and M.M. Moloney (1995b) Plant seed oil-bodies as carriers for foreign proteins. Bio/Tech. **13**: 72-77.

van Rooijen, G.J.H., Terning, L.I., and M.M. Moloney (1992) Nucleotide sequence of *Arabidopsis thaliana* oleosin gene. Plant Mol. Biol. **18**: 1177-1179

Vance, V.B. and A.H.C. Huang (1987) The major protein from lipid bodies of maize. J. Biol. Chem. **262**: 11275-79.

Vandekerckhove, J., Van Damme, J., Van Lijsebettens, M., Botterman, J., DeBlock, M., Vandeweile, M., De Clercq, A., Leemans, J., Van Montagu, M. and E. Krebbers (1989) Enkephalins produced in transgenic plants using modified 2S seed storage proteins. Bio/Technology. **7**: 929-932.

Verstraete, M., Nurmohamed, M., Kienast, J., Siebeck, M., Silling-Engenhardt, G., Buller, H., Hoet, B., Bichler, J., and P. Close (1993) Biologic effects of recombinant hirudin (CGP 39393) in human volunteers. J. Am. Coll. Cardiol. **22**: 1080-88.

Verwoerd, T.C., Dekker, B.M.M., and A. Hoekema (1989).A small-scale procedure for the rapid isolation of plant RNAs. Nuc. A. Res. **17**: 2362

Vogelstein, B. and D. Gillespie (1979) Preparative and analytical purification of DNA from agarose. Proc. Natl. Acad. Sci. **76**: 615-19

Walden, R. and R. Wingender (1995) Gene-transfer and plant regeneration techniques. Tibtech. **13**: 324-31.

Walker, P.A., Leong, L.E., Ng, P.W., Tan, S.H., Waller, S., Murphy, D. and A.G. Porter (1994) Efficient and rapid affinity purification of proteins using recombinant fusion proteases. Bio/Tech. **12**: 601-05.

Wan, Y. and P.G. Lemaux (1994) Generation of large numbers of independently transformed fertile barley plants. Plant Physiol. **104**: 37-48.

Wanner, G., Formanck, H., and R.R. Theimer (1981) The ontogeny of lipid bodies in plant cells. Planta **151**: 109-23.

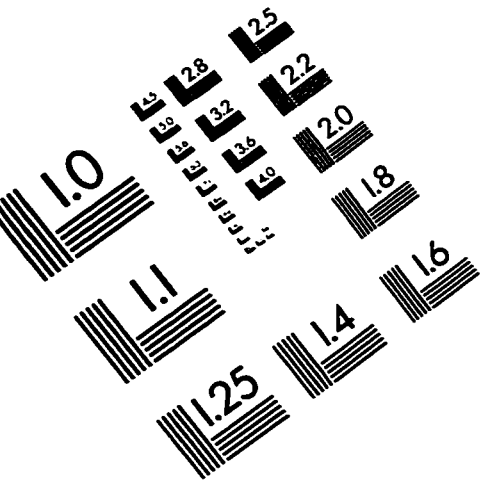
Weydemann, U., Keup, P., Piontek, M., Strasser, A.W.M., Schweden, J., Gellissen, G. and Z.A. Janowicz (1995) High-level secretion of hirudin by *Hansenula polymorpha*-authentic processing of three different preprohirudins. Appl. Micro. Biotech. **44**: 377-85.

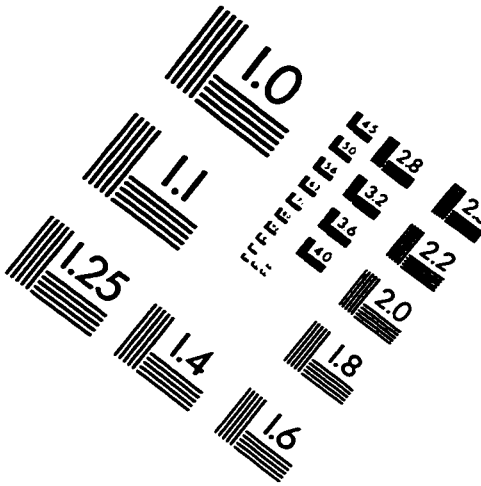
Wilson, K.A. (1986) In: *Plant Protyolytic Enzymes, vol. II.* Dalling, M.J. (ed) pp. 19-47, CRC Press, London.

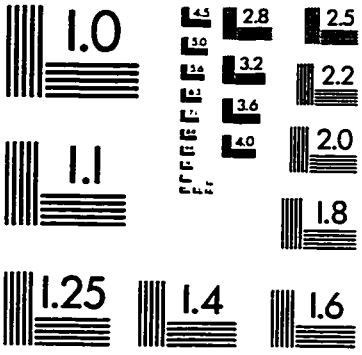
Wong, S-L. (1995) Advances in the use of *Bacillus subtilis* for the expression and secretion of heterologous proteins. Curr. Opin. Biotech. **6**: 517-22.

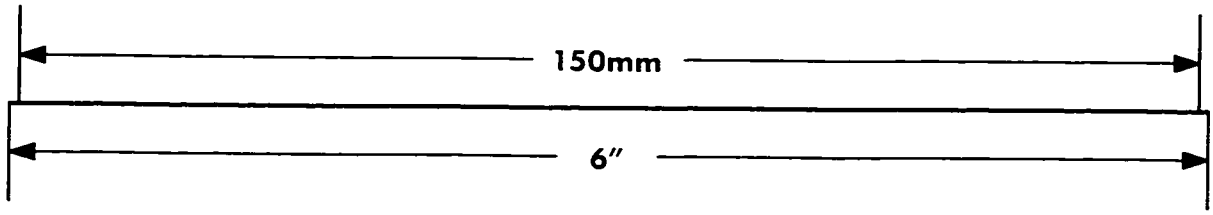
Wright, G., Carver, A., Cottom, D., Reeves, D., Scott, A., Simons, P., Wilmut, I., Garner, I, and A. Colman (1991) High level expression of active human alpha-1-antitrypsin in the milk of transgenic sheep. Bio/Tech. **9**: 830-34.

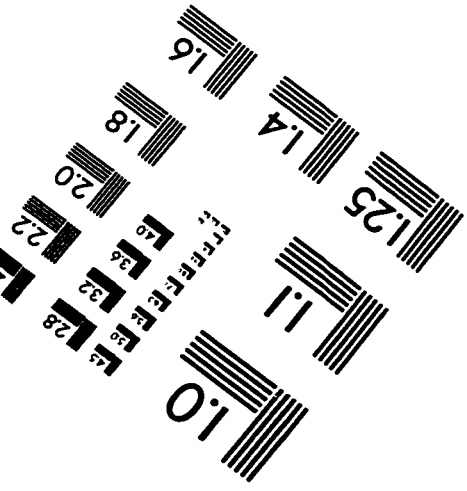
IMAGE EVALUATION TEST TARGET (QA-3)

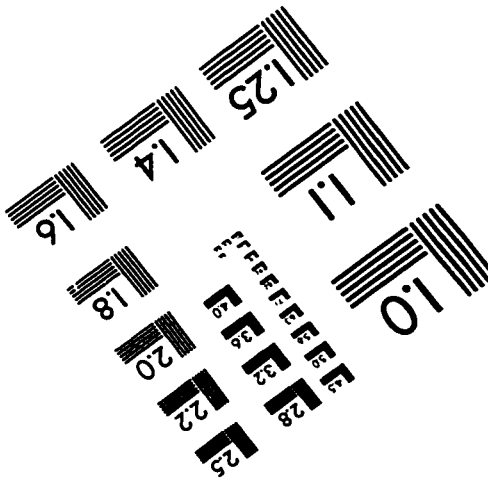












APPLIED IMAGE, Inc
1653 East Main Street
Rochester, NY 14609 USA
Phone: 716/482-0300
Fax: 716/288-5989