

Durham E-Theses

Expression of functional plant lectins in heterologous systems

Raemaekers, Romaan J.M.

How to cite:

Raemaekers, Romaan J.M. (2000) *Expression of functional plant lectins in heterologous systems*, Durham theses, Durham University. Available at Durham E-Theses Online:
<http://etheses.dur.ac.uk/4621/>

Use policy

The full-text may be used and/or reproduced, and given to third parties in any format or medium, without prior permission or charge, for personal research or study, educational, or not-for-profit purposes provided that:

- a full bibliographic reference is made to the original source
- a [link](#) is made to the metadata record in Durham E-Theses
- the full-text is not changed in any way

The full-text must not be sold in any format or medium without the formal permission of the copyright holders.

Please consult the [full Durham E-Theses policy](#) for further details.

Expression of Functional Plant Lectins in Heterologous Systems

A thesis submitted by Romaan J.M. Raemaekers (Bio-Engineer) in
accordance with the requirements of the University of Durham for the
degree of Doctor of Philosophy.

**The copyright of this thesis rests with
the author. No quotation from it should
be published in any form, including
Electronic and the Internet, without the
author's prior written consent. All
information derived from this thesis
must be acknowledged appropriately.**

Department of Biological Sciences

University of Durham

(U.K.)

2000



13 JUL 2001

Thesis

2000/
RAE

Abstract

The mannose-binding lectin from snowdrop (*Galanthus nivalis* agglutinin; GNA) was produced in *Escherichia coli* and purified as a functional protein after denaturation/renaturation. Incorporation of the four extra C-terminal residues recently revealed from X-ray crystallographic data demonstrated that these residues increase binding to the glycoprotein carboxypeptidase Y. However, no differences in activities were observed in haemagglutination assays when compared to native GNA and toxicity towards rice brown planthopper (*Nilaparvata lugens*; BPH) in artificial diet bioassays was unaltered. Site-directed mutagenesis of the carbohydrate-binding site of GNA provided evidence of a direct correlation between the binding potential of GNA to BPH gut glycoprotein 'receptors' and the toxicity levels of GNA towards BPH nymphs.

Functional recombinant plant lectins GNA and PHA (*Phaseolus vulgaris* agglutinin) were expressed in *Pichia pastoris* using native signal peptides or the *Saccharomyces* α -factor prepro-sequence to direct secretion. The α -factor prepro-sequence was inefficiently processed unless Glu-Ala repeats were added at the C-terminal end. In the latter case, removal of the Glu-Ala repeats was itself inefficient leading to recombinant lectins with heterogenous N-termini. In contrast, PHA expressed with the native signal peptide was secreted, correctly processed and fully functional. No expression of GNA from a construct containing the native GNA signal peptide was observed. The PHA-E signal peptide directed correct processing and secretion of both GNA and green fluorescent protein (GFP) when used in expression constructs in *Pichia*.

A fusion protein containing both GNA and GFP (GNA-GFP) was expressed in *Pichia pastoris*. Simultaneous dual activities (i.e. carbohydrate binding and fluorescence) of recombinant GNA-GFP were demonstrated. Partial cleavage in the linker region resulted in co-purification of GNA which increased the binding activity of the fusion protein. Selective binding of GNA-GFP to haemocytes in the haemolymph of *Lacania oleracea* was observed, both *in vitro* and when the protein was fed to insects in diet.

Table of Contents

Abstract.....	2
Table of Contents	4
List of Figures & Tables.....	10
Declaration.....	15
Statement of Copyright.....	15
Abbreviations.....	16
Acknowledgements	19
Publications.....	20
Chapter 1.....	21
Plant Lectins - Introduction	21
1.1 Lectins - Terminology.....	21
1.2 Occurrence and Distribution of Plant Lectins	22
1.3 Physiological Role - Plant Defense	24
1.4 Structure - Function Relationship of Plant Lectins.....	29
1.4.1 Classification	29
1.4.1.1 The Legume Lectins.....	31
1.4.1.2 The Chitin-Binding Lectins with Hevein Domains.....	32
1.4.1.3 The Type 2 Ribosome-Inactivating Proteins	33
1.4.1.4 The Monocot Mannose-Binding Lectins.....	34
1.4.1.5 Jacalin and Related Lectins.....	35
1.4.1.6 Lectins from <i>Amaranthus</i> Species.....	36
1.4.1.7 Cucurbitaceae Phloem Lectins.....	36

1.4.2 Strategies for Generation of Carbohydrate Binding and Specificities of Plant Lectins	37
Hydrogen bonds	38
Hydrophobic Interactions and van der Waals Forces	42
Extended Sites and Secondary Sites	45
Quaternary Associations.....	48
1.5 The Snowdrop Lectin.....	52
1.6 Phytohaemagglutinin	56
1.7 Secretory Pathways in Plants and Yeast (<i>Pichia pastoris</i>)	57
1.8 Research Objectives.....	62
Chapter 2.....	64
Materials & Methods	64
2.1 Microorganisms	64
2.2 Chemicals and Reagents	65
2.3 Standard Molecular Techniques	65
2.3.1 Bacterial Culture	65
2.3.2 Competent Cells.....	65
2.3.3 Transformation of <i>E. coli</i>	66
2.3.4 Plasmid DNA Isolation	67
2.3.5 DNA Concentration Determination	67
2.3.6 DNA Digestion with Restriction Enzymes.....	67
2.3.7 Phenol Extraction and Ethanol Precipitation of DNA	68
2.3.8 DNA Ligation	68
2.3.9 Oligonucleotides	69

2.3.10 Standard DNA Amplification using PCR (Polymerase Chain Reaction)	69
2.3.11 Colony PCR	69
2.3.12 Agarose Gel Electrophoresis of DNA	70
2.3.13 DNA Purification from Agarose Gel	70
2.3.14 DNA Sequencing	71
2.3.15 Glycerol Stocks	71
2.4 Standard Protein Analysis	72
2.4.1 SDS-PAGE	72
2.4.2 Immunoblotting	73
2.4.3 Electroblothing of Proteins to PVDF Membranes	74
2.4.4 N-Terminal Sequencing	75
2.4.5 Gel Filtration	75
2.4.6 Deglycosylation of Protein	75
2.4.7 Protein Concentration Determination	76
2.5 Construction of Expression Vectors	76
2.5.1 Expression Constructs for GNA in <i>Escherichia coli</i>	77
2.5.2 Site-Directed Mutagenesis	77
2.5.3 Expression Constructs for GNA in <i>Pichia pastoris</i>	78
2.5.4 Expression Constructs for PHA-E and PHA-L in <i>Pichia pastoris</i>	82
2.5.5 Expression Constructs for GFP in <i>Pichia pastoris</i>	85
2.5.6 Expression Constructs for GNA-GFP in <i>Pichia pastoris</i>	85
2.6 Expression of Recombinant Protein in <i>E. coli</i> and <i>P. pastoris</i>	88
2.6.1 Induction of Expression of Recombinant GNA in <i>Escherichia coli</i>	88
2.6.2 Transformation of <i>Pichia</i>	91

2.6.3 Induction of Expression of Recombinant Proteins in <i>Pichia pastoris</i> ...	92
2.6.4 Preparation of Protein Extracts from <i>Pichia</i> Cells.....	93
2.7 Purification of Recombinant Lectin.....	93
2.7.1 Purification of Recombinant Wild-Type GNA from <i>E. coli</i> by Affinity Chromatography	93
2.7.2 Purification of Recombinant Wild-Type and Altered GNA from <i>E. coli</i> by Anion-Exchange Chromatography	94
2.7.3 Affinity-Purification of PHA and GNA from <i>Pichia</i> Culture Supernatant.....	94
2.8 Recombinant Protein Functional Assays	95
2.8.1 Plant Lectins and Green Fluorescent Protein	95
2.8.2 Hemagglutination Assays.....	96
2.8.3 GNA-Carboxypeptidase-Y-Binding Assay	96
2.8.4 Direct ELISA of GNA Variants.....	97
2.8.5 Probing of BPH Total Gut Protein Extract with Recombinant GNA and Detection by Immunoblotting	98
2.8.6 BPH Bioassays.....	99
2.8.7 Extraction of Haemolymph from <i>Lacania oleracea</i>	99
2.8.8 Preparation of Haemocyte Monolayers.....	99
2.8.9 Fluorescence Microscopy	100
2.8.10 Confocal Microscopy	100
2.8.11 Fluorometric Assays.....	101
Chapter 3.....	102
Expression of Functional GNA in <i>E. coli</i>	102
3.1 Introduction	102

5.1 Introduction	173
5.2 Results	174
5.2.1 Expression of GNA-GFP in <i>P. pastoris</i>	174
5.2.2 Purification of GNA-GFP	185
5.2.3 Functional Activities of Recombinant GNA-GFP	186
5.2.3.1 Haemagglutination Activity of Recombinant GNA-GFP	188
5.2.3.2 Binding of Recombinant GNA-GFP to Mannan-Agarose Beads .	190
5.2.3.3 Binding of Recombinant GNA-GFP to Haemocytes <i>in vitro</i>	190
5.2.3.4 Binding of Recombinant GNA-GFP to Haemocytes <i>in vivo</i>	196
5.3 Discussion	199
Chapter 6	208
General Discussion & Future Prospects	208
Chapter 7	217
Appendix	217
References	219

List of Figures & Tables

Figure 1.1 Schematic diagram of an example of hydrogen binding in lectin- monosaccharide interactions.....	39
Figure 1.2 Three-dimensional structure of the tetrameric snowdrop lectin.	53
Figure 1.3 Yeast secretory pathway.	58
Table 1.1 Oligonucleotide sequences of primers used to prepare <i>Pichia</i> expression constructs.	79
Figure 2.1 Development of expression constructs for GNA.	80
Figure 2.2 Development of expression constructs for GNA <i>continued</i>	81
Figure 2.3 Development of expression constructs for PHA-E and PHA-L.....	83
Figure 2.4 Development of expression constructs for GFP.....	86
Figure 2.5 Development of expression constructs for GFP <i>continued</i>	87
Figure 2.6 Development of expression constructs for GNA-GFP.....	89
Figure 2.7 Development of expression constructs for GNA-GFP <i>continued</i>	90
Figure 3.1 Amino acid sequence of the mature GNA polypeptide.....	106
Figure 3.2 Region of high-affinity mannose-binding site I of GNA.....	107
Figure 3.3 Analysis of total protein extracts of pGNA105-containing <i>E. coli</i> after induction.	108
Figure 3.4 Analysis of total protein extracts of pGNA11-containing <i>E. coli</i> after induction.	110
Figure 3.5 Analysis of denatured/renatured rGNA11 and rGNA105.	112
Figure 3.6 Purification of rGNA105 by affinity chromatography on a mannose- agarose column.....	114

Figure 3.7 Purification of rGNA11 by affinity chromatography on a mannose-agarose column.	115
Figure 3.8 Anion-exchange chromatography of rGNA105 and rGNA109.	116
Figure 3.8 <i>continued</i> . Anion-exchange chromatography of rGNA105(Y97F) and rGNA109(Y97F).	117
Figure 3.8 <i>continued</i> . Anion-exchange chromatography of rGNA105(D91L) and rGNA105(N93L).	118
Figure 3.8 <i>continued</i> . Anion-exchange chromatography of rGNA105(D91L/N93L).	119
Figure 3.9 SDS-PAGE analysis of IEC-purified rGNA109(Y97F).	120
Figure 3.10 SDS-PAGE and immunoblot analysis of IEC-purified recombinant wild-type and mutant GNA.	121
Figure 3.11 SDS-PAGE and immunoblot analysis of IEC-purified recombinant GNA variants.	122
Figure 3.12 Haemagglutination and haemagglutination inhibition assays of recombinant wild-type and mutant GNA.	125
Figure 3.13 Resolution of IEC-purified rGNA105 and rGNA105(Y97F) by gel filtration.	126
Figure 3.13 <i>continued</i> . SDS-PAGE analysis of IEC- and IEC/GF-purified rGNA105 and rGNA105(Y97F).	127
Figure 3.14 Haemagglutination assays of IEC-purified recombinant wild-type and mutant GNA.	129
Figure 3.15 Binding of GNA variants to CpY in the presence of varying amounts of mannose.	130

Figure 3.15 <i>continued</i> . Inhibition of binding of GNA variants to CpY by mannose.	131
Figure 3.16 Recognition of GNA variants by polyclonal anti-nGNA antibodies by direct ELISA.	132
Figure 3.17 Binding of rGNA109 and rGNA109(Y97F) to CpY in the presence of varying amounts of mannose.	134
Figure 3.17 <i>continued</i> . Inhibition of binding of rGNA109 and rGNA109(Y97F) to CpY by mannose.	135
Figure 3.18 Immunoblot analysis of BPH gut glycoproteins probed with rGNA105 or rGNA105(Y97F).	136
Figure 3.19 Effects of recombinant wild-type and altered GNA to BPH survival. ..	138
Figure 4.1 Summary of expression and purification of PHA in <i>Pichia</i> and silver stained gel showing the affinity purification of recombinant PHA-L on a thyroglobulin-agarose column expressed with the native signal peptide.	149
Table 4.1 Haemagglutination activity of rabbit erythrocytes by recombinant PHA-E and GNA; lowest concentrations of lectins to give complete agglutination are shown.	151
Figure 4.2 Summary of expression and purification of GNA in <i>Pichia</i> expressed using the α -factor prepro- sequence.	153
Figure 4.3 GNA and GFP expression.....	158
Figure 4.4 Summary of processing and N-terminal sequences of lectin and GFP constructs expressed in <i>Pichia pastoris</i>	159
Figure 4.5 Agglutination of <i>Pichia</i> cell suspension in the presence and absence of native GNA.	163

Figure 4.6 Binding of affinity-purified <i>Pichia</i> -expressed PHA-E:GNA to CpY in the presence of varying amounts of mannose.....	165
Figure 4.7 Binding to CpY of varying concentrations of native GNA dissolved in <i>Pichia</i> BMMY culture medium or in PBS alone.	166
Figure 4.8 Binding to CpY of native GNA dissolved in PBS, 1% yeast extract, 1% beef extract and 1.34% yeast nitrogen base supplement.	167
Figure 5.1 Distribution of rGFP and GNA-GFP in <i>Pichia</i> GS115 cells.	177
Figure 5.1 <i>continued</i> . Distribution of rGFP and GNA-GFP in <i>Pichia</i> GS115 cells.	178
Figure 5.1 <i>continued</i> . Distribution of rGFP and GNA-GFP in <i>Pichia</i> GS115 cells.	179
Figure 5.1 <i>continued</i> . Distribution of rGFP and GNA-GFP in <i>Pichia</i> GS115 cells.	180
Figure 5.2 Immunoblot analysis of <i>Pichia</i> culture medium containing secreted GFP and GNA-GFP.....	183
Figure 5.3 Immunoblot analysis of <i>Pichia</i> cell extracts containing GFP and GNA-GFP.....	184
Figure 5.4 Immunoblot analysis of affinity-purified rGNA-GFP.....	187
Figure 5.5 Immunoblot analysis of treated and untreated GNA-GFP.....	189
Figure 5.6 Elution peak profile of GNA-GFP during affinity purification and concentration ratios of GNA-GFP to GNA in eluted fractions collected.	191
Figure 5.7 Haemagglutination assays of affinity-purified recombinant GNA-GFP.	193
Figure 5.8 Binding of affinity-purified recombinant GNA-GFP to mannan-agarose beads.	194
Figure 5.8 <i>continued</i> . Binding of affinity-purified recombinant GNA-GFP to mannan-agarose beads.	195
Figure 5.9 Binding of affinity-purified recombinant GNA-GFP to 6 th instar <i>Lacania oleracea</i> haemocytes <i>in vitro</i> – crude technique.	197

Figure 5.9 <i>continued</i> . Binding of affinity-purified recombinant GNA-GFP to 6 th instar <i>Lacania oleracea</i> haemocytes <i>in vitro</i> – crude technique.....	198
Figure 5.10 Binding of affinity-purified recombinant GNA-GFP to 6 th instar <i>Lacania oleracea</i> haemocytes <i>in vitro</i> – monolayer technique.....	200
Figure 5.11 Binding of affinity-purified recombinant GNA-GFP to 4 th instar <i>Lacania oleracea</i> haemocytes <i>in vivo</i>	201
Figure 5.12 Immunoblot analysis of pooled haemolymph extracted from 4 th instar <i>Lacania oleracea</i>	202

Declaration

No material here has previously been submitted for any other degree. Except where acknowledged, all material is the work of the author.

Statement of Copyright

The copyright of this thesis rests with the author. No quotation from it should be published without his prior written consent and information derived from it should be acknowledged.

Abbreviations

AMA	<i>Amaranthus caudatus</i> agglutinin
BMGH	buffered glycerol medium with histidine
BMGY	buffered glycerol complex medium
BMMH	buffered methanol medium with histidine
BMMY	buffered methanol complex medium
BPH	brown planthopper (<i>Nilaparvata lugens</i>)
BSA	bovine serum albumin
CAPS	3-(cyclohexylamino)-1-propanesulfonic acid
cDNA	complementary or copy DNA
ConA	concanavalin A (Jack bean)
CpY	carboxypeptidase Y
DAP	1,3-diaminopropane
DBL	<i>Dolichos biflorus</i> (horde gram) lectin
DMSO	dimethylsulphoxide
ECL	enhanced chemiluminescence
EcorL	<i>Erythrina corallodendron</i> lectin
EDTA	ethylenediaminetetracyclic acid
ELISA	enzyme-linked immunosorbent assay
ER	endoplasmatic reticulum
FITC	fluorescein isothiocyanate
Fuc	fucose
Gal	galactose
GalNAc	<i>N</i> -acetylgalactosamine

GalNDns	<i>N</i> -dansylgalactosamine
GF	gel filtration
GFP	green fluorescent protein
Glc	glucose
GlcNac	<i>N</i> -acetylglucosamine
GNA	<i>Galanthus nivalis</i> (snowdrop lectin) agglutinin
GS	<i>Griffonia simplicifolia</i>
IEC	ion-exchange chromatography
IPTG	isopropyl- β -D-thiogalactoside
KM+	a lectin isolated from the jackfruit
LenL	lentil lectin
LOA	<i>Listera ovata</i> agglutinin
LOL	<i>Lathyrus ochrus</i> agglutinin
Man	mannose
Me	methyl
MPA	<i>Maclura pomifera</i> (Osage orange tree) agglutinin
NeuAc	sialic acid
nGNA	native GNA; GNA isolated from snowdrop
NPA	<i>Narcissus pseudonarcissus</i> agglutinin
PBS	phosphate-buffered saline
PCR	polymerase chain reaction
PHA	<i>Phaseolus vulgaris</i> (red kidney bean) agglutinin
PNA	peanut (<i>Arachis hypogaea</i>) agglutinin
PSL	<i>Pisum sativum</i> (garden pea) lectin
PVDF	polyvinylidenedifluoride

rGNA	recombinant GNA
RIP	ribosome inactivating protein
SBA	soybean (<i>Glycine max</i>) agglutinin
SDS-PAGE	sodiumdodecylsulfate polyacrylamide gel electrophoresis
SNA	<i>Sambucus nigra</i> (elderberry) agglutinin
TBS	tris-buffered saline
Temed	<i>N,N,N',N'</i> -tetramethylethylenediamine
UDA	<i>Urtica dioica</i> (stinging nettle) agglutinin
UEA	<i>Ulex europeaus</i> (Furze gorse) agglutinin
WGA	wheat germ (<i>Triticum vulgare</i>) agglutinin
X-Gal	5-bromo-4-chloro-3-indolyl- β -D-galactoside
YNB	yeast nitrogen base

List of constructs for expression in *E. coli*

rGNA105	recombinant GNA consisting of 105 amino acid residues
rGNA105(D91L)	rGNA105 with mutation Asp -> Leu at position 91
rGNA105(N93L)	rGNA105 with mutation Asn -> Leu at position 93
rGNA105(D91L/N93L)	rGNA105 with mutations Asp -> Leu, & Asn -> Leu at positions 91 & 93, respectively
rGNA105(Y97F)	rGNA105 with mutation Tyr -> Phe at position 97
rGNA109	rGNA105 with 4 extra C-terminal amino acid residues
rGNA109(Y97F)	rGNA109 with mutation Tyr -> Phe at position 97

Acknowledgements

I would like to thank everyone that helped me in producing this manuscript.

I would like to thank my supervisor, Dr. John Gatehouse, for giving me the opportunity to work on a very interesting topic, for his continued guidance, advice, encouragement, enthusiasm and support throughout the project, and for critically reading this and other manuscripts. This work would not have been possible without the generous funding of Dr. John Gatehouse and the Biological Sciences Department, University of Durham. I would like to thank Dr. Tony Fordham-Skelton for his invaluable help with some of the *Pichia* work and for the helpful discussions, Dr. Dave Bown for his technical know-hows, Drs. Xavier Foissac and Jinping Du both for their advice, discussions and motivation. Many thanks extend to all the colleagues in Lab 1 for creating a good working atmosphere.

I am very grateful to Dr. Elaine Fitches for assistance in preparation of caterpillar haemolymph samples, Steve Woodhouse for help with the confocal microscopy images, Dr. Jackie Spence for her time with the fluorescent microscopic images, Dr. Mathusuma Bharathi for the laborious brown planthopper bioassays, Peter Hunter (University of Durham) and CSL (York) for the rabbit blood samples, John Gilroy for N-terminal protein sequencing, Julia Bartley, Vicky Kelly and Gillian Storey for their excellence in DNA sequencing, and Paul Sydney and David Hutchinson for the photographic work.

Finally, I would like to thank my parents and Tim for their support and persistence.

Publications

Parts of this work have been included in the following publications:

Down, R.E., Ford, L., Woodhouse, S.D., Raemaekers, R.J.M., Leitch, B., Gatehouse, J.A. & Gatehouse, A.M.R. (2000). Snowdrop lectin (GNA) has no acute toxic effects on a beneficial insect predator, the 2-spot ladybird (*Adalia bipunctata* L.). *J. Insect Physiol.* **46**, 379-391.

Raemaekers, R.J.M., de Muro, L., Gatehouse, J.A. & Fordham-Skelton, A.P. (1999) Functional phytohaemagglutinin (PHA) and *Galanthus nivalis* agglutinin (GNA) expressed in *Pichia pastoris*. *Eur. J. Biochem.* **265**, 394-403.

Longstaff, M., Powell, K. S., Gatehouse, J. A., Raemaekers, R., Nevell, C. A. & Hamilton, W. D. O. (1998) Production and purification of active snowdrop lectin in *Escherichia coli*. *Eur. J. Biochem.* **252**, 59-65.

Chapter 1

Plant Lectins - Introduction

1.1 Lectins - Terminology

Lectins are carbohydrate-binding proteins of non-immune origin that bind reversibly to mono- or oligosaccharides. They are widely distributed in nature, being found in viruses, micro-organisms, plants and animals. Because of their binding specificity, lectins may serve as recognition molecules within a cell, between cells, or between organisms (Sharon, 1993).

All too often, the term lectin implies that it is synonymous with agglutinin (agglutinating lectin). However, not all lectins are able to agglutinate cells or precipitate glycoconjugates. Some lectins consist exclusively of a single carbohydrate-binding domain and due to their monovalent nature, these lectins, or merolectins (Van Damme *et al.*, 1998a), cannot agglutinate cells or precipitate glycoconjugates. For example, the chitin-binding hevein from the latex of the rubber tree (Van Parijs *et al.*, 1991), and the mannose-binding proteins from orchids (Van Damme *et al.*, 1994) do not show agglutinating activity. Also, others, such as the so-called type 2 ribosome-inactivating proteins ricin and abrin, agglutinate very poorly, even though they contain lectinic subunits (Kocourek & Horejsi, 1983).

The emphasis on non-immune origin is necessary to distinguish lectins from anti-carbohydrate antibodies which also bind reversibly to carbohydrates. However, only

the lectins are found in plants and bacteria. In addition, antibodies are structurally similar to each other, whereas lectins are structurally diverse.

Lectins are also distinguished from many carbohydrate-binding proteins in that they do not alter the covalent structure of recognized glycosyl ligands. However, sometimes what is perceived to be an enzyme is actually, per definition, a lectin. For example, Class I chitinases are lectins that consist of a chitin-binding domain linked through a hinge region to a catalytic domain (Collinge *et al.*, 1993). Similarly, ricin and abrin, are fusion proteins of an N-terminal toxic A-chain with *N*-glycosidase activity and a C-terminal carbohydrate-binding domain (Barbieri *et al.*, 1993).

1.2 Occurrence and Distribution of Plant Lectins

Plant lectins are ubiquitous in the plant kingdom, being found in both dicotyledons and monocotyledons, with, for example, Class I chitinases seeming to be present in all plant species (Collinge *et al.*, 1993). However, when only the agglutinating lectins (agglutinins) are taken into consideration then the occurrence of these in plants is the exception rather than the rule (Van Damme *et al.*, 1998a).

Lectins can be found in seeds, and in virtually all kinds of vegetative tissues such as leaves, stems, bark, bulbs, tubers, corms, rhizomes, roots, fruits, flowers, ovaries, phloem sap and nectar (Peumans & Van Damme, 1995). There are striking differences in the location and relative abundance of seed and non-seed lectins in plants (Etzler, 1986). For instance, seed lectins are located in the cotyledons (e.g. in legumes), the endosperm (e.g. castor bean) or embryo (e.g. wheat). Typically, lectins

usually account for 0.1 - 5 % of the total seed protein. However, some seed lectins are predominant proteins representing up to 50 % of the total protein (e.g. in some *Phaseolus* species). Non-seed lectins also account for 0.1 - 5 % of the total protein content of the tissue in which they occur. However, there are instances of lectins accounting for greater than 50 % of the total protein content in vegetative storage tissues, for example, in garlic (*Allium sativum*) cloves (Smeets *et al.*, 1997) and ground elder (*Aegopodium podagraria*) rhizomes (Peumans *et al.*, 1985). Others occur in very small quantities that are hardly detectable (e.g. in leaves of the leek (*Allium porrum*) (Van Damme *et al.*, 1993)). Non-seed lectins may occur in different tissues of the same plant. For example, the snowdrop (*Galanthus nivalis*) and daffodil (*Narcissus pseudonarcissus*) lectins have been found in virtually all vegetative tissues, although the lectin is most abundant in the bulbs (Van Damme & Peumans, 1990). The potato (*Solanum tuberosum*) lectin occurs in tubers, stems, leaves and fruits (Kilpatrick, 1980). However, there are some exceptions. For instance, the ground elder lectin is confined to the rhizome (Peumans *et al.*, 1985). Lectins from tulips (*Tulipa*) are present in the bulbs but are almost undetectable in the stems and leaves (Van Damme & Peumans, 1989). There are also a few examples of lectins which occur both in seed and in vegetative tissues. For example, some legume lectins are found in seeds as well in bark tissue. However, these lectins are not identical and are encoded by different though highly homologous genes (Van Damme *et al.*, 1995a).

On the subcellular scale, most lectins are secretory proteins, meaning that they enter the secretory system and subsequently accumulate either in vacuoles or in the cell wall and intercellular spaces (Chrispeels & Raikhel, 1991). Vacuolar lectins from

legume seeds, such as phytohaemagglutinin, concanavalin A, soybean agglutinin, pea lectin and favin, are primarily located in protein bodies of the cotyledons (Etzler, 1986). In the thorn apple *Datura stramonium* the seed lectin is not only present in protein bodies (Kilpatrick *et al.*, 1979) but some of the lectin is also associated with cell walls (Broekaert *et al.*, 1988). The structurally related potato tuber lectin is also found associated with the cell walls (Casaloungue & Lezica, 1985). Similarly, a cell wall lectin has been found in mung bean (*Phaseolus aureus*) hypocotyls (Kauss & Glaser, 1974; Haasz *et al.*, 1981).

1.3 Physiological Role - Plant Defense

Since lectins are found in many different species and many different tissues and organs, it is assumed that they play fundamental biological roles in plants. Lectins have been implicated in a wide spectrum of roles including seed maturation/germination (Howard *et al.*, 1972), maintenance of seed dormancy (Peumans *et al.*, 1983), cell wall extension (Kauss & Glaser, 1974), cell-cell recognition (Knox *et al.*, 1976), growth regulation (Howard *et al.*, 1972) and carbohydrate transport (Boyd, 1963). A number of lectins have been found to be mitogens for lymphocytes (Goldstein & Hayes, 1978) ((Lis & Sharon, 1977;1981), although they have been found to have little (Howard *et al.*, 1977; Del Campillo *et al.*, 1981) or no (Nagl, 1972; Vasil & Hubbell, 1977) mitogenic effects on plant cells of the same species from which these lectins have been obtained. Surprisingly, there are a number of lectins that have a hydrophobic-binding site which bind to adenine and adenine derivatives with cytokinin activity (Roberts & Goldstein, 1983) as well as indoleacetic acid (Edelman & Wang, 1978). Legume root lectins may be involved

in the specific recognition of bacteria (*Rhizobium* and *Bradyrhizobium* species) for the purpose of establishing symbiosis (Bohloul & Schmidt, 1974). However, the actual species-species recognition appears to be mediated by small fatty acylated and sulfated tetrasaccharides generated by the bacteria (Lerouge *et al.*, 1990), not by the lectins themselves. Instead, their role may be to accumulate by agglutination large numbers of bacteria to the root hair surface.

Plant lectins are thought to be important in plant defense. Many plant lectins are capable of recognizing and binding with high affinity glycoconjugates not common or absent in plants, but present on the surfaces of micro-organisms (i.e. bacteria and fungi), and intestinal tracts of insects or mammalian predators. Most lectins maintain their integrity under unfavourable conditions; they are stable over a wide pH range, are able to withstand heat, and are resistant to animal, including insect, proteases. The whole plant is continually exposed to pathogens and pests, and inevitably some vital tissues and organs, such as storage organs and seeds, require extra protection. Preferential accumulation of lectins in these metabolic inactive resting organs could thus be advantageous for protecting the plant.

There is indirect evidence that some plant lectins are involved in defence against bacteria. Ayoub *et al.* (1994) demonstrated that several legume lectins were able to bind to specific bacterial cell wall peptidoglycan components, such as muramic acid, *N*-acetylmuramic acid, and muramyl dipeptide. Broekaert and Peumans (1986) showed that the presence of thorn apple (*Datura stramonium*) seed lectin (GlcNAc-specific) resulted in the loss of movement of motile bacteria at the air-water interface. Lectins that bind chitin have been found to affect the growth and development of

organisms that contain chitin (i.e. fungi and insects). For example, UDA, a lectin from the stinging nettle, has been shown to have strong antifungal properties against several chitin-producing fungi (Broekaert *et al.*, 1989). Also, it was observed that a preparation of chitin affinity-purified wheat germ agglutinin (WGA) inhibited spore germination and hyphal growth of *Trichoderma viride* (Mirelman *et al.*, 1975). However, these effects were due to contaminating chitinases, rather than WGA, since Schlumbaum *et al.* (1986) demonstrated that *in vitro*, chitinase-free preparations of WGA do not inhibit fungal growth, whereas chitinase does inhibit fungal growth. The exact mechanism of the nettle lectin is not known but it appears that only the cell wall synthesis is affected as a result of disturbed chitin synthesis (Van Parijs *et al.*, 1992). Since the lectin is not able to kill germinating spores or mycelium, unlike Class I chitinases which are considered fungicidal, there is doubt as to whether they actually take part in plant defensive roles. Some authors believe that the nettle lectin, naturally found in rhizomes and seeds, is involved in the control of the colonization of rhizomes by endomycorrhiza (Peumans & Van Damme, 1995).

There are many clear examples of lectins providing a protective role in plants against insects (Czapla & Lang, 1990; Gatehouse *et al.*, 1995). Although the mode of action is still unknown, the toxicity of the lectins is most probably based on their specific binding to glycoconjugates on the luminal side of the gut. Four types of interactions are possible: (i) binding of lectins to carbohydrate structures in the peritrophic membrane (chitin for chitin-binding lectins) (Harper *et al.*, 1998), (ii) binding of lectins to carbohydrate structures on the epithelial cells along the digestive tract (Pusztai & Bardocz, 1996), (iii) binding of lectins to glycosylated digestive enzymes, and (iv) binding to glycosylated proteins from the host plant preventing the protein's

digestion. Screening of hundreds of purified plant lectins through feeding experiments utilizing artificial diets were focussed on economically important crop pests and have resulted in promising candidates towards the development of transgenic crops (Czapla & Lang, 1990; Gatehouse *et al.*, 1995). The first report on the lethal effects of a purified lectin sample from the bean *Phaseolus vulgaris* towards the cowpea weevil (*Callosobruchus maculatus*) larvae (Janzen *et al.*, 1976) were wrongly attributed to the lectin PHA - the active material was actually the contaminating α -amylase inhibitor (Huesing *et al.*, 1991), thus emphasizing on the purity of the lectin preparations. Since then, a number of plant lectins were identified for their effectiveness against the cowpea weevil (Murdoch *et al.*, 1990; Huesing *et al.*, 1991; Gatehouse *et al.*, 1992), European corn borer (*Ostrinia nubilalis*) (Czapla & Lang, 1990), corn rootworm (*Diabrotica* species) (Czapla & Lang, 1990), legume pod borer (*Maruca vitrata*) (Machuka *et al.*, 1999), rice brown planthopper (*Nilaparvata lugens*) (Powell *et al.*, 1993; 1995a; 1995b), potato leafhopper (*Empoasca fabae*) (Habibi *et al.*, 1993), pea aphid (*Acyrtosiphon pisum*) (Rahbe & Febvay, 1993; Rahbe *et al.*, 1995), peach-potato aphid (*Myzus persicae*) (Sauvion *et al.*, 1996), blowfly (*Lucilia cuprina*) (Eisemann *et al.*, 1994), and tomato moth (*Lacanobia oleracea*) (Fitches *et al.*, 1997). Clear evidence emerging from these results show that there is no direct correlation between lectin-carbohydrate specificity and toxicity. Two examples illustrate this point. (i) although the sialic acid-binding lectins from elderberry (*Sambucus nigra*; SNA-I) (Shibuya *et al.*, 1987) and *Maackia amurensis* (Knibbs *et al.*, 1991) exhibit similar specificity (i.e. bind to 2,6-neuraminyl-gal/GalNAc, a sugar that is absent in plants but a major carbohydrate component of animal glycoproteins), SNA-I was found to be extremely toxic to *C. maculatus*, whereas the other lectin was relatively innocuous to the same insect. (ii)

the toxicity of GlcNAc-specific lectins may be attributed to binding of these lectins to the chitin-containing peritrophic membrane. However, such a mechanism neither explains the toxicity of lectins such as the mannose-specific GNA (*Galanthus nivalis* agglutinin) and LOA (*Listera ovata* agglutinin), nor the lack of toxicity of GlcNAc-specific WGA (wheat germ agglutinin) towards *Maruca*-pod borer (Machuka *et al.*, 1999). It is speculated that since the peritrophic membrane contains other carbohydrate moieties besides chitin, lectins with different specificities may cause membrane blockage by binding to the different glycoproteins. Also, GNA has been shown to pass through the insect midgut epithelium and enter into the haemolymph of the rice brown planthopper, thus exerting its toxic effects systemically (Powell *et al.*, 1998).

Since plant viruses do not contain carbohydrate moieties on their surfaces, it is logical to assume that plant lectins cannot directly exert anti-viral activity. However, insecticidal lectins may reduce or prevent the spread of insect-transmitted viral diseases.

The digestive tract in higher animals has a wide spectrum of glycoconjugates exposed in the luminal side presenting a myriad of targets for lectin binding. Unlike with insects, the mechanisms by which the lectins exert their toxicity is well understood because of studies assessing the possible health risk of lectins present in plants used in food and feed production. For example, PHA, the lectin from the kidney bean *Phaseolus vulgaris*, has an extraordinarily high resistance to gut proteolysis. Once ingested, it binds to complex glycoconjugates in the brush border cell membrane of the small intestine, where it is rapidly endocytosed. Once in the cell, the lectin sets off

an enhanced metabolic activity that eventually leads to hyperplasia and hypertrophy of the small intestine culminating in severe discomfort of the animal (Pusztai *et al.*, 1990). Other common features of the anti-nutritive effects in lectin-gut interaction include damage to the microvillus membrane, shedding of cells, reduction in the absorptive capacity of the small intestine, interference with the immune system, hypersensitivity reactions, interference with the microbial ecology of the gut, selective overgrowth, and effects (i.e. hormones) on systemic metabolism. Well documented deleterious effects of other lectins on cells of the small intestine include soybean (*Glycine max*) agglutinin, concanavalin A (from jack bean, *Canavalia ensiformis*), other lectins from *Phaseolus* genus, and wheat germ (*Triticum aestivum*) agglutinin (Pusztai, 1991). However, there is no evidence for GNA toxicity towards higher animals (Pusztai *et al.*, 1990) so this protein would be suitable for incorporation into a transgenic crop.

1.4 Structure - Function Relationship of Plant Lectins

1.4.1 Classification

Plant lectins form a very large and heterogenous group of proteins. They may be grouped in three different ways: (i) according to their overall structure, (ii) according to their carbohydrate-specificities, or (iii) according to their evolutionary and structural relatedness.

There are four main types of plant lectins based on the overall structure of the lectin subunits. These are the mero-, holo-, chimero-, and super-lectins. Merolectins are monovalent lectins that are exclusively built of only one carbohydrate-binding motif.

Thus, they are not able to agglutinate cells or precipitate glycoconjugates. Examples include the monomeric mannose-specific proteins from orchids (*Listera ovata* and *Epipactis helleborine*) (Van Damme *et al.*, 1994) and the chitin-binding hevein (*Hevea brasiliensis*) (Van Parijs *et al.*, 1991). Hololectins contain two or more carbohydrate-binding sites. This is realised by consisting of (i) only one monomer containing two or more binding sites (e.g. *Urtica dioica* agglutinin) (Peumans *et al.*, 1984), (ii) di - oligomer containing only one site per subunit (e.g. dimer *Pisum sativum* agglutinin, and tetramer phytohaemagglutinin PHA) (Rini *et al.*, 1993; Hamelryck *et al.*, 1996) or (iii) di - oligomer containing two or more sites per subunit (e.g. tetramer with three sites per subunit *Galanthus nivalis* agglutinin) (Hester *et al.*, 1995). Because of their subunit multivalency, the hololectins have agglutination properties. Chimerolectins are proteins with a carbohydrate-binding domain fused to a completely unrelated domain. Chimerolectins may act as (i) merolectins, for example, Class I plant chitinases which consist of only one N-terminal chitin-binding site per molecule linked to a catalytic domain (Collinge *et al.*, 1993), or (ii) hololectins, for example, Type 2 ribosome-inactivating proteins, such as the dimeric ricin, that possess an N-terminal catalytic domain connected to a C-terminal carbohydrate-binding chain consisting of two sites (Barbieri *et al.*, 1993). Superlectins are chimerolectins that consist of two structurally different carbohydrate-binding domains with unrelated sugar specificities. For example, a lectin from tulip (*Tulipa*) bulbs, TxLC-I, consists of an N-terminal mannose-binding domain fused to a GalNAc-binding domain (Cammue *et al.*, 1986; Van Damme *et al.*, 1996a).

Plant lectins may also be divided into groups according to their preferential binding to sugars. To date, at least eight carbohydrate specificity groups have been

distinguished, namely, mannose, mannose/glucose, mannose/maltose, fucose, GlcNAc, Gal/GalNAc, sialic acid, and complex glycan groups (Van Damme *et al.*, 1998a).

In general, the most common type of secondary structure of vacuolar plant proteins is the β -sheet fold (Efimov, 1994) interconnected with a number of loops, with the conspicuous absence of the α -helix motif. However, recently, two related lectins isolated from *Dolichos biflorus* also contain a small C-terminal α helix for quaternary association and adenine/hormone binding (Hamelryck *et al.*, 1999). Plant lectins exhibit considerable diversity in their structure, yet they may be classed into at least seven subgroups of structurally and evolutionarily related families (Van Damme *et al.*, 1998b). The legume lectins, the Type 2 ribosome-inactivating proteins (RIPs), the chitin-binding lectins and the monocot mannose-binding lectins are considered large families, whereas the jacalin-related lectins, the amarantins, and the Cucurbitaceae phloem lectins are considered small families.

1.4.1.1 The Legume Lectins

Legume lectins occur exclusively within the plant family Leguminosae. Over a 100 legume lectins have been characterised in detail. All share a common polypeptide fold known as the β -sandwich (or jelly-roll motif). They are the only group of plant lectins that actually utilize divalent cations in their carbohydrate-binding specificity; the lectins exhibit markedly conserved Ca^{2+} and Mn^{2+} binding sites. Although they are built of protomers with high sequence similarities and 3D structures, legume lectins differ strongly in their sugar-specificities (Sharon & Lis, 1990). The majority of lectins are mannose/glucose-binding lectins (e.g. *Canavalia ensiformis* lectin, *Pisum*

sativum lectin) and Gal/GalNAc-binding lectins (e.g. *Arachis hypogaea* lectin, *Sophora japonica* lectin). Others are specific for L-fucose (e.g. *Ulex europaeus* lectin), Neu5Ac α (2,3)Gal/GalNAc (e.g. *Maackia amurensis* lectin) or for more complex oligosaccharides (e.g. *Phaseolus vulgaris* lectin) (Van Damme *et al.*, 1998a). This difference in specificity is a result of mutational events in the combining site (Sharma & Surolia, 1997), whereby the length and orientation of loop regions that provide the framework of the carbohydrate-binding site is altered.

Surprisingly, the β -sandwich structure is also shared by some animal lectins, such as galectins, pentraxins and some other unrelated proteins from animal and microbial sources (Srinivasan *et al.*, 1996). The amino acid sequences are unrelated and the sugar-binding sites lie on different parts of the structure, suggesting that this similarity arises as a result of convergence to a stable structural motif rather than from a common evolutionary ancestor (Emsley *et al.*, 1994).

1.4.1.2 The Chitin-Binding Lectins with Hevein Domains

Chitin-binding lectins composed of hevein domains are widespread throughout the plant kingdom, occurring in taxonomically unrelated plant families such as Gramineae, Urticaceae, Solanaceae, Papaveraceae, Euphorbiaceae, Phytolaccaceae (Raikhel *et al.*, 1993) and Viscaceae (Peumans *et al.*, 1996). These lectins are built up of one, two, three, four or seven hevein domains. The name of the domain is derived from hevein, a small chitin-binding protein from the latex of the rubber tree (*Hevea brasiliensis*). Several structurally unrelated chitin-binding lectins do not possess a hevein domain, e.g. GlcNAc-binding legume lectins and the chitin-binding phloem lectins of Cucurbitaceae, and thus do not belong to this category.

The hevein sequence contains only 43 amino acids and is very rich in cysteine and glycine. The chitin-binding lectins have evolved by gene duplication of an ancestral disulfide-rich domain (Wright *et al.*, 1991). The domain is predominantly free of any regular secondary structure. Structure determination of the dimeric WGA have shown that each subunit consists of four independently folded and helically assembled hevein domains with each domain stabilised by four disulphide bridges at identical positions (Wright, 1989). Both the amino acid sequences and three-dimensional structures of the hevein domains are markedly conserved, explaining why all chitin-binding lectins have similar carbohydrate-binding specificities.

1.4.1.3 The Type 2 Ribosome-Inactivating Proteins

Type 2 RIPs are a superfamily of structurally related lectins found in members of taxonomically unrelated plant families. Well-known examples of this family are ricin (from the seeds of *Ricinus communis*) and abrin (from the seeds of *Abrus precatorius*). Type 2 RIPs differ from Type 1 RIPs in their molecular structure. Type 1 RIPs are monomeric proteins consisting of a singly catalytically active subunit (approx. 30 kDa), whereas Type 2 RIPs are composed of one (ricin), two (SNA-V) (from bark of *Sambucus nigra* or elderberry) (Van Damme *et al.*, 1996b) or four (SNA-I) (from bark of *Sambucus nigra* or elderberry) (Van Damme *et al.*, 1996c) identical units with each unit consisting of an A and B chain covalently linked through a disulphide bridge. The A chain possesses N-glycosidase activity and exhibits sequence homology to the Type 1 RIP, whereas the B chain is composed of a carbohydrate-binding domain. As is the case with the chitin-binding WGA, the ricin B chain is stabilised through multiple disulphide bonds.

Like legume lectins, Type 2 RIPs differ from each other in their sugar specificities even though they are structurally similar. Most Type 2 RIPs are specific for galactose or GalNAc, however, SNA-I recognizes NeuAc α (2,6)Gal/GalNAc (Van Damme *et al.*, 1996c).

1.4.1.4 The Monocot Mannose-Binding Lectins

The monocot mannose-binding lectins are an extended superfamily of structurally and evolutionarily related proteins which have been isolated from species of the Alliaceae, Amaryllidaceae, Araceae, Bromeliaceae, Iridaceae, Liliaceae, and Orchidaceae (Van Damme *et al.*, 1998b). These lectins show exclusive specificity towards mannose which makes them different from the mannose/glucose-binding lectins isolated from leguminous species (Goldstein & Poretz, 1986). The subunit structure, as exemplified by the known three-dimensional structures of the snowdrop (Hester *et al.*, 1995) and amaryllis (Chantalat *et al.*, 1996) lectins, represents a highly conserved novel lectin fold, namely, the β -barrel (β -prism II). This fold consists of three four-stranded antiparallel β -sheets arranged to form a triangular shaped prism. The monocot mannose-binding lectins exhibit a marked structural diversity both in terms of number and overall structure of the protomers. Most of the lectins are composed of two or four identical protomers of approx. 12 kDa, which are synthesised as separate proteins. There are also several types of heteromeric forms: (i) built up of two different, but highly homologous, subunits derived from separate precursors (e.g. dimeric *Allium ursinum* lectin I) (Smeets *et al.*, 1994); (ii) built up of two different subunits derived from a single precursor with two tandemly arrayed mannose-binding domains which share high homology (e.g. dimeric *Allium sativum*

lectin I) (Van Damme *et al.*, 1992) or low homology (e.g. tetrameric *Arum maculatum* lectin) (Van Damme *et al.*, 1995b); (iii) built up of two unrelated tandemly arrayed domains derived from a single precursor (e.g. tetrameric tulip lectin TxLC-I with each protomer consisting of a mannose-binding domain and GalNac-binding domain which is partly processed giving rise to both cleaved and uncleaved protomers) (Van Damme *et al.*, 1996a).

1.4.1.5 Jacalin and Related Lectins

A novel mode of carbohydrate recognition was discovered in jacalin, a *Moraceae* plant lectin derived from the jackfruit (*Artocarpus integrifolia*) with a β -prism I fold (Sankaranarayanan *et al.*, 1996). This involves three Greek-key β sheets with the strands running parallel to the threefold axis (in comparison to the snowdrop lectin with a β -prism II fold with three four-stranded β sheets that are nearly perpendicular to the threefold axis). It is worth mentioning that this fold has striking similarity to the β -prism fold that occurs in domain II of the insecticidal *Bacillus thuringiensis* δ -endotoxin (Li *et al.*, 1991).

Jacalin is a glycosylated tetramer with each subunit composed of an α (133 residues) chain and a short β (20 residues) chain as a result of complex post-translational processing of a single precursor. The presence of different isoforms is due to expression by different genes of a lectin gene family and also to differences in post-translational modifications (Young *et al.*, 1995). The jacalin fold has one galactose-binding site per subunit located at the loop region at one end of the β -prism fold and involving the N-terminus (Sankaranarayanan *et al.*, 1996). Jacalin binds to the T-antigen disaccharide (Gal- β 1,3-GalNac) with high specificity (Kabir & Daar, 1994).

This observation is ratified by the crystal structure determination of another member of the jacalin family, *Maclura pomifera* agglutinin (MPA), in complex with the disaccharide. Another lectin from the jacalin family, KM+, isolated from the jackfruit, is mannose-specific (Rosa *et al.*, 1999). Unlike the other lectins, KM+ does not undergo proteolysis into α and β chains.

1.4.1.6 Lectins from *Amaranthus* Species

Amaranthin (ACA) is a dimeric Gal-specific plant lectin isolated from the seeds of *Amaranthus caudatus* (Rinderle *et al.*, 1990). Like jacalin, amaranthin is T-antigen-specific. However, the protein utilizes the β -trefoil fold, a protein motif first observed in the soybean trypsin inhibitor (Murzin *et al.*, 1992) and also found in the B-chains of the Type 2 RIPs ricin and abrin (Rutenber & Robertus, 1991; Tahirov *et al.*, 1995). Despite structural similarity, there is no sequence similarity to other β -trefoil proteins, and the (QxW)₃ peptide, reminiscent in many sugar-binding proteins (Hazes, 1996), is absent. Other *Amaranthus* species contain closely related seed lectins similar in sequence to ACA (e.g. lectin isolated from *Amaranthus leucocarpus*) (Raina & Datta, 1992).

1.4.1.7 Cucurbitaceae Phloem Lectins

Cucurbitaceae phloem lectins are a small group of chitin-binding proteins confined to the phloem sap of a few genera of the Cucurbitaceae family. A dimeric lectin was isolated from a *Cucurbita* species and characterized (Sabnis & Hart, 1978). Despite its carbohydrate-binding properties, the pumpkin lectin does not share any sequence

homology with the chitin-binding hevein domain containing lectins (Wang *et al.*, 1994).

1.4.2 Strategies for Generation of Carbohydrate Binding and Specificities of Plant Lectins

Although many examples of lectin-carbohydrate interactions are covered in the literature, only those pertaining to plant lectins will be discussed here. Animal and viral lectins were reviewed recently (Weis, 1997; Drickamer, 1999; Rini & Lobsanov, 1999).

How do lectins interact with sugar ligands? In recent years, an upsurge of lectin-sugar complex analyses by a variety of biophysical techniques, including X-ray crystallography and nuclear magnetic resonance, has led to a better understanding of the molecular basis of this interaction. More than 100 three-dimensional structures of plant lectins and their complexes are currently available on the 3D Lectin Data Bank on the World Wide Web (URL: <http://www.cermav.cnrs.fr/databank/lectine>). From these data, it is apparent that different strategies exist for lectins to recognize sugar structures. As was discussed above, plant lectins could be organised into a small number of groups according to their tertiary folds. Lectins with different folds are able to recognize structurally similar ligands (e.g. the monocot mannose-specific lectins and the mannose/glucose-specific legume lectins). Also, lectins with similar folds are able to distinguish different sugar structures (e.g. legume lectins). Carbohydrate-binding sites may be found on completely different frameworks altogether, as exemplified by the legume lectins with their large β -structures and low

cysteine content and the cereal WGA with its high cysteine content and absence of known regular secondary structure.

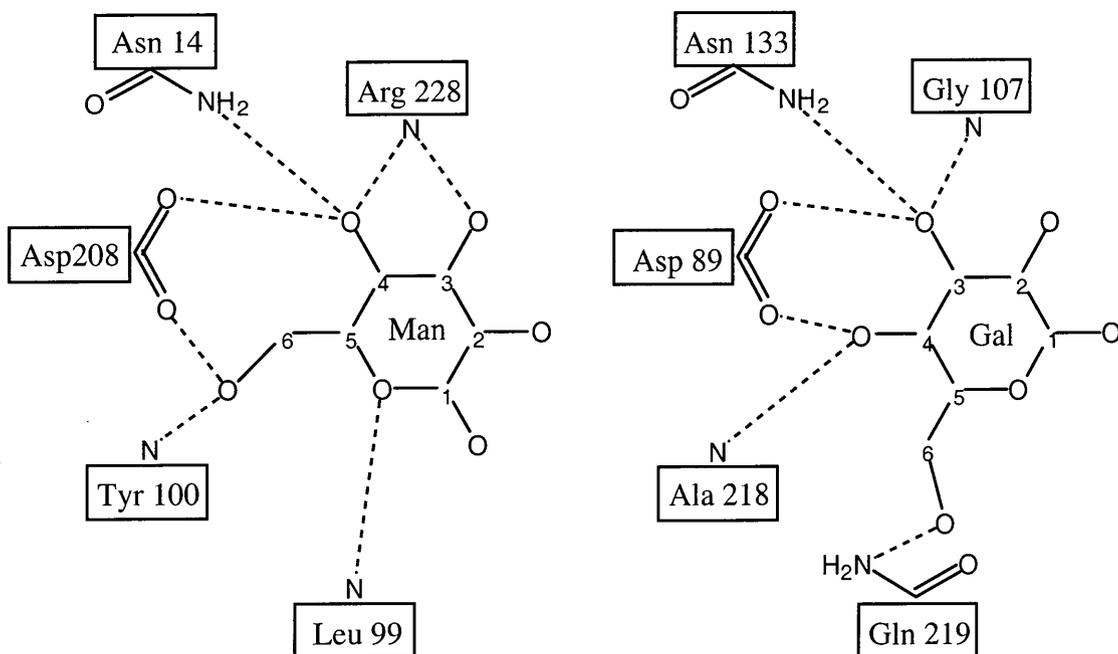
The basic principles that allow different lectins to bind selected saccharide ligands through hydrogen bonds in cooperation with van der Waals and hydrophobic interactions with appropriate affinity will be discussed.

Hydrogen bonds

The monosaccharide, usually a hexose, presents a number of hydroxyl groups that can serve as both hydrogen bond donors and acceptors in lectin-carbohydrate interactions. The oxygen atom of each hydroxyl group is sp^3 hybridised; the oxygen atom has two lone electron pairs, and is bound to a hydrogen and carbon atom in tetrahedral geometry. Thus, the OH group can accept two and donate one hydrogen bond. The ring oxygen atom is also sp^3 hybridised and harbours two lone electron pairs; it can only accept hydrogen in hydrogen bonds. Different protein side-chain (carboxyl, hydroxyl and amide groups) and main-chain (carbonyl and amide groups) groups can participate in this hydrogen bonding. In general, an amide group acts as hydrogen bond donor and the carbonyl group as acceptor ($(NH)_{1 \text{ or } 2} \rightarrow OH \rightarrow O=C$) (Weis & Drickamer, 1996). Less frequently, protein OHs, such as tyrosine, threonine and serine, can also take part in hydrogen bonding (Elgavish & Shaanan, 1997).

The legume lectins are widely used as a model system for studying protein-carbohydrate interactions. Hydroxyl groups of a particular sugar that mediate hydrogen binding in lectin-saccharide interactions determine the specific recognition of the lectin for the saccharide. The legume lectins ConA (Naismith *et al.*, 1994), pea

(A)



(B)

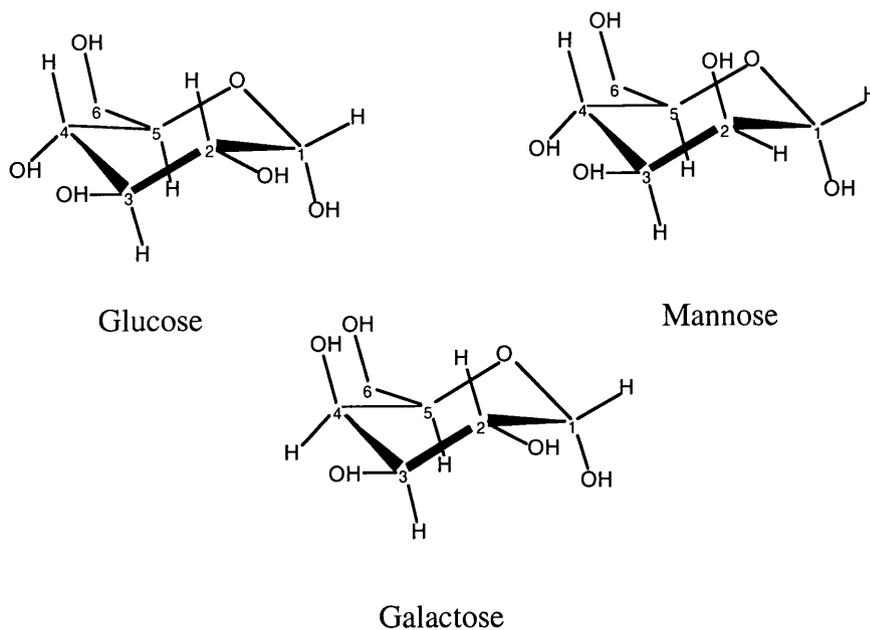


Figure 1.1 (A) Schematic diagram of an example of hydrogen binding in lectin-monosaccharide interactions: binding of mannose/glucose and galactose to concanavalin A and EcorL, respectively (modified from Loris *et al.* (1998)); dashed lines represent hydrogen bonds. (B) Conformations of three hexoses.

lectin (Rini *et al.*, 1993) and LOL1 (Bourne *et al.*, 1990) all bind the C2 epimers mannose and glucose. They form hydrogen bonds with the 3-, 4- and 6-OHs of the sugars, but not the 2-OH, so that discrimination of the two monosaccharides by the lectins is not possible. In contrast, the snowdrop lectin, GNA, is subjected to hydrogen bonding at the 2-OH as well as the other OHs so that only mannose is recognized; this axial hydroxyl group is conformationally constrained by two hydrogen bonds (Hester *et al.*, 1995). GNA completely lacks specificity towards glucose (Shibuya *et al.*, 1988); hydrogen bonding with an equatorial OH would be much less favourable.

The lectins from the legumes *Lathyrus ochrus* (LOL1) and *Erythrina corallodendron* (EcorL) differ in their specificities, the former, as already mentioned, being Man/Glc-specific and the latter Gal/GalNAc-specific. Man/Glc and Gal are C4 epimers; both mannose and glucose have an equatorial 4-OH, whereas galactose has an axial 4-OH. These lectins have quite similar sugar-binding domains and are part superimposable (Sharon, 1993). The amino acid residues that participate in hydrogen bonding (Asp, Asn and usually Gly or Arg) are highly conserved and participate in four key hydrogen bonds with the monosaccharide. Also, the sugar molecule is stabilised by stacking interactions with a hydrophobic residue, Phe, Tyr, Trp, or Leu (see further). All of these residues have an identical spatial disposition which is attributable to the constraints imposed by the indispensable divalent cations. The same core of residues in legume lectins provides the infrastructure to fit a diverse array of monosaccharides.

In the mannose/glucose specific legume lectins (e.g. LOL1, ConA, PSL, favin and LenL), the hydrogen bonding occurs mainly with the C4 and C6 hydroxyl groups of

the hexose. However, in EcorL, PNA, DBL or SBA that bind galactose or N-acetylgalactosamine, these interactions are mainly with the C3 and C4 of the sugar (Figure 1.1). In other words, the binding site is designed to allow an equatorial OH (C4-OH in mannose and glucose, C3-OH in galactose or GalNAc) to establish contact with three conserved binding residues (Asp in loop A, Gly or Arg in loop B, and Asn in loop C) and an axial OH (C6-OH in mannose and glucose, C4-OH in galactose or GalNAc) hydrogen bond with Asp in loop A (Sharma & Surolia, 1997). These limitations in sugar orientation are not sufficient to discriminate the different monosaccharide specificities. Specificity arises by additional variation (in sequence and in size) in the loops. Difference in primary specificity is dictated by the conformation of loop D (Shaanan *et al.*, 1991). Alignment of 26 legume lectins showed a direct correlation with the length of loop D and the monosaccharide specificity of the lectins (Sharma & Surolia, 1997). Binding of EcorL exclusively to galactose shows that the carbohydrate ring has rotated in the binding site so that more bonds are focussed on the C3 and C4 hydroxyls. The residues in EcorL involved in hydrogen bonding with C4-OH are located in an entirely different loop (loop D) as those found in LOL1 (loop C). The variable residues adjacent to the conserved ones are thought to be the reason why the orientation of the pyranose ring with respect to the framework for ligand binding differs (Elgavish & Shaanan, 1997).

Mutational analysis of pea lectin indicated that replacement of Asn125 by a conservative Asp (Van Eijsden *et al.*, 1992) or non-conservative Ala (Van Eijsden *et al.*, 1994) completely abolished mannose/glucose binding. Surprisingly, this demonstrates that the presence of other conserved amino acid residues involved in sugar binding (haemagglutination) is insufficient for sugar binding. Similar results

residue (Tyr73), with the sugar ring (Tyr66) and with the glycerol tail (Tyr64) of sialic acid (Wright, 1990), underlining the importance of the contribution made by non-polar contacts in stabilising the lectin-sugar complex. WGA also recognizes GlcNAc. The essential specificity determinant for the two monosaccharides is the N-acetyl group and the adjacent 4-OH, which provide a cluster of three spatially close hydrogen bonds and a hydrophobic contact (acetamido-CH₃ with the aromatic ring of Tyr73) in the least exposed part of the binding cavity, where the conformation of the protein is most stable (Sharon, 1993).

Extended Sites and Secondary Sites

Monosaccharide interactions with lectins are very weak, with dissociation constants in the 0.1 - 1.0 mM range. The selectivity towards a particular target is augmented by several orders of magnitude (in the μ M range) through additional binding in extended and secondary sites, often referred to as subsite multivalency (Rini, 1995).

Some lectins bind to single saccharide residues only, whether they are in free form or exist at the terminal end of an oligosaccharide. The X-ray crystal structure of the EcorL-lactose (lactose is Gal β 1,4Glc) complex shows that binding is mediated exclusively through the galactose moiety (Shaanan *et al.*, 1991). This observation is further supported by Elgavish and Shaanan (1998) who reported that binding of disaccharides (lactose and N-acetyllactosamine) by EcorL is carried out by the same set of residues that are involved in binding monosaccharides at the primary site. This feature is also shared by the PNA-lactose complex (Banerjee *et al.*, 1996).

Other lectins are capable of interacting with the additional monosaccharide residues along the carbohydrate chain, even though they may not recognize the second sugar as a monosaccharide. These lectins have extended sites allowing for increased interaction and therefore increased affinity and consequently increased selectivity. PNA has a 20-fold higher binding affinity to the T-antigen disaccharide (Gal β 1-3-GalNAc) compared to lactose (Ravishankar *et al.*, 1997). This specificity is entirely due to additional water-mediated protein-carbohydrate interactions; the amount of direct protein-carbohydrate hydrogen bonds and non-polar contacts being identical (Loris *et al.*, 1998). The mannose/glucose-specific ConA and the *Viciaeae* lectins LOLI, pea and lentil lectins all bind mannose and glucose in essentially the same way. However, ConA has a high affinity for the trimannosyl core Man α 1-3[Man α 1-6]Man found in N-glycans (Debray *et al.*, 1981), whereas no significantly enhanced affinity for the trimannose core over α -D-mannopyranoside is observed with the *Viciaeae* lectins (Kornfeld *et al.*, 1981). The reason is that the latter lectins lack two residues that in ConA are responsible for its specificity for the trimannose core; Tyr12 and Tyr100 in ConA are substituted by Phe123 β and Glu31 α in LOL, respectively, which result in a difference in orientation of the saccharide leading to interaction reduction (Naismith & Field, 1996). Notwithstanding, the *Viciaeae* lectins display higher affinity for N-linked oligosaccharides containing α (1,6)-linked Fuc than for those lacking Fuc, whereas ConA does not (Debray *et al.*, 1981; Kornfeld *et al.*, 1981). Crystal structures of LOL complexed with fragment N2 of human lactotransferrin and with an isolated biantennary glycopeptide showed that the fucose residue interacts with the subsite residues Phe and Glu (Bourne *et al.*, 1994; Loris *et al.*, 1998).

Some lectins show little or no measurable affinity for monosaccharides, but instead bind specifically to larger oligosaccharides. The *Griffonia simplicifolia* lectin (GS4) recognizes galactose, not on its own, but only when it is part of a saccharide complex, as found on the Lewis b blood group substance (Delbaere *et al.*, 1993). In this case, the typical legume lectin loop D structure, which is important for monosaccharide binding (cfr. EcorL-galactose complex), is truncated so as to completely eliminate this side of the binding site. The $\alpha(1,4)$ -linked fucose moiety of the bound saccharide occupies the space provided and is stabilised by hydrogen-bond and van der Waals interactions. Clearly, specificity towards carbohydrates requires extended sites.

Lectins possessing two or more separate and independent carbohydrate-binding sites (secondary sites) per protomer may increase their binding affinity towards complex sugars, especially multivalent ligands. The dimeric WGA has a hierarchy of binding sites (4 sites per protomer), but despite contributions from helper domains located on the other protomer, the carbohydrate-binding sites can only recognize terminal sialic acid or N-acetylglucosamine (Rini, 1995). Thus, carbohydrate specificity is based on the difference in monosaccharide affinities and on the multitude of carbohydrate-binding sites. GNA, too, has multiple binding (3) sites per protomer. All sites recognize mannose monosaccharides (Hester *et al.*, 1995), however, one of the sites is also an extended binding region complementary to α -1,3 linked mannosides (Hester & Wright, 1996), giving this plant lectin the ability to bind to complex oligo-mannan receptors.

Quaternary Associations

As described above, binding of an individual lectin site (monovalent binding) to monosaccharides is very weak, with dissociation constants (K_D) typically in the 0.1 – 1.0 mM range (Rini, 1995; 1996). Many lectins achieve much higher affinities (in the nM range) by clustering several similar or identical binding sites by formation of oligomers i.e. quaternary structures. The free energy of binding to a multivalent ligand to multiple sites on an oligomeric lectin can be as large as the sum of the free energies of the individual binding interactions. However, in reality, considerations in geometry in oligosaccharide recognition have to also be taken into account (Weis & Drickamer, 1996). Clustering so that the spacing and orientation of the binding sites are in a particular order, i.e. projection in one direction, and location at opposite ends (Drickamer, 1995) can have important effects on the selectivity with which multivalent ligands are recognized.

Proteins in which the sugar binding sites project in one direction are able to recognize multivalent sugar-bearing surfaces as the optimal ligand for these multivalent lectins. Many animal and viral lectins are known to recognize and bind avidly to host cell-surface glyconjugates (Weis, 1997; Drickamer, 1999; Rini & Lobsanov, 1999). On the other hand, only a few plant lectins to date show this geometrical feature e.g. the plant toxin ricin (Drickamer, 1995).

Many plant lectins which are dimeric harbour binding sites at opposite ends of the dimer. This allows linear assemblies of lectin (e.g. ConA) and carbohydrate. Additional numbers of binding sites on an oligomeric lectin, such as the tetrameric SBA, allow for 3D lattices with a synthetic pentasaccharide (Dessen *et al.*, 1995). The

bridging function between plant lectins and cell-surface glycoproteins has become a widely exploited property, for instance, in *in vitro* agglutination assays.

Several plant lectins display more complex arrangements of multiple binding sites. GNA (Hester *et al.*, 1995) and WGA (Wright & Jaeger, 1993) molecules are practically studded with binding sites, with the GNA tetramer representing the highest density lectin characterised to date. Tetrameric GNA contains 12 sites (3 sites per subunit) all over its surface, whereas dimeric WGA has a total of 8 sites. Both GNA (Wright & Hester, 1996) and WGA (Wright, 1992) have a hierarchy of binding sites, with some of the sites dominating the interaction with complex ligands. Availability of such a large number of sites allows formation of a wide range of multi-dimensional lattices with a high degree of specificity.

Quaternary association and multivalency that results as a consequence are important for the functional properties of proteins. The much studied legume lectins, whose individual subunits each fold in essentially the same manner, are known to exhibit a number of different quaternary structures. Thus, large variations in quaternary structure is possible due to small alterations in essentially the same tertiary structure. Rapid evolution in quaternary structure may be necessary to circumvent the problem of high affinity binding of carbohydrate-binding proteins to the intrinsic flexible carbohydrates.

Oligomerisation in all legume lectins basically involves interactions of the six-stranded back β -sheet. In the formation of a dimer, two β -sheet arrangements are found: (i) side-by-side, where the two sheets form a contiguous 12-stranded β -sheet,

and (ii) back-to-back, where the two sheets may form a handshake mode. The side-by-side arrangement is often known as the canonical mode of legume lectin dimerisation. ConA, pea lectin, lentil lectin and *Lathyrus* lectin all assume this structure. *E. corallodendron* lectin cannot form the canonical dimer mode due to sterical hindrance from a carbohydrate covalently bound to Asn17 at the interface (Shaanan *et al.*, 1991). GS4 cannot assume the standard dimeric mode due to carbohydrate interference and burial of a charged residue (Glu58) in the subunit interface (Delbaere *et al.*, 1993). WBAI is homologous in sequence to EcorL, but the glycosylation sites in WBAI are far from the monomer-monomer interface in a canonical dimer. Nevertheless, in WBAI a canonical dimer is not formed (Prabu *et al.*, 1998). It seems more likely that the mode of dimerisation is primarily dictated by factors intrinsic to the protein itself (Prabu *et al.*, 1998).

Dimer-dimer associations also differ within the legume lectin family. The tetrameric formation of ConA, SBA and PHA-L can be described as the back-to-back arrangement of the canonical dimers. PNA formation, however, involves the association of two back-to-back dimers resulting in the only known tetrameric protein that does not contain 4-fold or 222 symmetry (Banerjee *et al.*, 1996).

Novel quaternary structures of legume lectins have been reviewed recently by Bouckaert *et al.* (1999). Lectins from *Dolichos biflorus* exhibit a surprising quaternary structure involving an α helix sandwiched between two β -sheets. This unusual interface hosts the binding site for adenine and plant hormones (Hamelryck *et al.*, 1999). The tetrameric structure of a chitobiose-specific lectin from *Ulex*

europaeus (UEA-II) (Dao-Thi *et al.*, 1998), resembling that of PHA-L and SBA, is stabilised by inter-protomer disulfide bridges.

High-affinity interactions between oligomeric lectins and multivalent ligands may be the result of the combined effects of subsite and subunit multivalency, or as a consequence of subunit multivalency alone. GS1 exists as a heterotetramer composed of subunits A and B in the five possible combinations, giving isolectins A₄, A₃B, A₂B₂, AB₃, and B₄. The first four are tetravalent for human type A erythrocytes. Knibbs *et al.* (1998) showed that the association constant of the GS1 isolectins for human type A blood cells increased with increasing valency of the isolectin. This increase was solely dependent on subunit multivalency, not on an extended site.

Specificity generation for different complex carbohydrate systems by oligomerisation of sugar-binding proteins is exemplified by the monocot mannose-specific lectin family. The snowdrop (Hester *et al.*, 1995), daffodil (Sauerborn *et al.*, 1999), and bluebell (Wood *et al.*, 1999) lectins are all tetramers, whereas the garlic lectin (Chandra *et al.*, 1999) is a dimer. Although the mannose-binding sites and structure are similar in the subunits of these lectins, their specificities to complex glycoproteins such as GP120, the major surface glycoprotein of the human immunodeficiency virus, vary considerably (Balzarini *et al.*, 1991); the tetramers bind GP120 with a high degree of affinity, whereas the dimeric garlic lectin does not. Comparison of the snowdrop lectin with the garlic lectin elucidates the binding affinity differences. Two distinct modes for branched mannopentaose binding exist for snowdrop lectin (Wright & Hester, 1996), one cross-links the two monomers in the dimer, and the other cross-links the two dimers in the tetramer. The latter mode is consistent with the

specificity of the snowdrop lectin for α -1,3 linkages (Hester & Wright, 1996) and is therefore considered to be the biologically relevant mode. The dimeric garlic lectin cannot bind the mannopentose in the same fashion and this appears to be a possible cause of the loss of its antiretroviral activity (Vijayan & Chandra, 1999). Also, by virtue of this multivalency, the snowdrop lectin is a more potent insecticidal lectin than garlic lectin (Powell *et al.*, 1995b).

1.5 The Snowdrop Lectin

The snowdrop lectin (*Galanthus nivalis* agglutinin; GNA) is an unglycosylated homotetrameric (50 kDa) protein that belongs to the monocot mannose-specific lectin family. GNA specifically recognizes mannose structures and has a high affinity for α -1,3 linked mannosides (Shibuya *et al.*, 1988). However, unlike the less discriminating mannose/glucose-specific legume lectins, GNA completely lacks specificity for glucose (Kaku & Goldstein, 1989). Binding affinities for GNA were much weaker than for legume lectins; GNA has an association constant of between 1 and 5 mM, whereas a 100-fold higher association constant is found with ConA (Kaku & Goldstein, 1992; Chervenak & Toone, 1995).

GNA is a robust molecule resisting heat (it withstands heating at 70 °C for 10 minutes) and acid/alkali denaturation (it is stable within the pH range of between 3 and 12) (Van Damme *et al.*, 1987). The GNA lectin is present in almost all snowdrop tissues, but found mainly in the bulbs (Van Damme & Peumans, 1990). Six different



Figure 1.2 Three-dimensional representation of the tetrameric GNA structure (Hester *et al.*, 1995). Each subunit is represented by a different colour: A, green; B, orange; C, yellow; D, purple. Methyl-mannose is shown interacting with one (site I) of its three binding domains. C-terminal strand exchange involves residues 99-109.

isolectins have been identified (five have been published; (Van Damme *et al.*, 1991a)) showing sequence variability in the region of the C-terminus.

Sequence analysis of a cDNA clone revealed that GNA is synthesized as a preproprotein, a polypeptide precursor of 157 amino acid residues composed of a 23-residue N-terminal signal sequence and a 29-residue carboxy-terminal extension, which is post-translationally cleaved to yield a mature 105 amino acid residue protein (Van Damme *et al.*, 1991b).

X-ray crystallographic studies of GNA in complex with methyl α -D-mannose (Fig. 1.2) revealed a novel tertiary fold exhibiting local threefold symmetry generated by three antiparallel four-stranded β sheets (Hester *et al.*, 1995). However, the polypeptide can accommodate four extra residues at the C-terminal end yielding a total length of 109 amino acids. The independently folded β sheets each harbour a conserved carbohydrate-binding site (site I, II, and III). The tetramer consists of two tightly associated dimers (A-D and B-C), which are stabilized through C-terminal strand exchange and β sheet association. The interfaces between these two dimers (A-B and C-D) consist of hydrophobic contacts between loops.

Further crystallographic studies of GNA in complex with a monomannoside (methyl α -D-mannoside), a dimannoside (Man- α 1,3-D-Man-OMe) and a branched 3,6 core mannopentaose (Man α 1,6-(α 1,3-Man)Man- α 1,6-(α 1,3-Man)Man) revealed that the snowdrop lectin exhibits two binding modes (Hester & Wright, 1996; Wright & Hester, 1996). Binding of the monosaccharide or terminal mannose was observed in each independent subunit at all three of the conserved monosaccharide binding sites

(12 sites/tetramer), although site I had highest occupancy since the residue bound in the specificity pocket is further stabilised through numerous van der Waals contacts from the C-terminal arm of the contacting subunit (Hester & Wright, 1996). Substitutions of residues of the C-terminal arm may explain the stronger binding affinity of daffodil (*Narcissus pseudonarcissus*, a member of the Amaryllidaceae family) lectin (NPA) to methyl-mannose when compared to GNA; there is more extensive stacking contact with His83 & Tyr107 relative to GNA where these residues are Asn & His (Wright & Hester, 1996). Both terminal mannoses of the 3,6 tri-Man arm of the branched mannoside cross-linked the two-fold related GNA dimers utilizing the conserved monosaccharide pocket of site I (Hester & Wright, 1996; Wright & Hester, 1996). Site III is the only site with an extended binding region complementary to α -1,3 linked mannosides. Both the disaccharide and trimannoside of the mannopentaose filled into the binding pocket with interactions of the same residues, including the same subsidiary contacts from a second subunit, but the latter saccharide had extra contacts from within the same subunit (Hester & Wright, 1996; Wright & Hester, 1996).

Mannose-containing glycoconjugates are abundant on cell surfaces in higher organisms and viruses, and thus present numerous possible targets for GNA binding. GNA binds to the GP120 glycoprotein of retroviruses and as a consequence has the ability to inhibit retroviral activity (Balzarini *et al.*, 1991). In agriculture, GNA has been shown to be toxic towards sucking and chewing insects when transgenically expressed (Hilder *et al.*, 1995).

1.6 Phytohaemagglutinin

Phytohaemagglutinin (PHA) is a tetrameric glycosylated lectin found in the seeds of the common bean (*Phaseolus vulgaris*). Two tandemly linked genes *dlec1* and *dlec2* encode two different (>80 % homologous) polypeptides, namely, PHA-E and PHA-L, respectively. The PHA tetramer is built up of these two forms in all possible combinations (i.e. E₄, E₃L, E₂L₂, EL₃, and L₄). Both subunit types are glycosylated; PHA-E contains 3 glycosylation sites, whereas PHA-L has two (Hoffman & Donaldson, 1985). PHA-L is N-glycosylated at two different sites with consensus sequence Asn-X-Ser/Thr: (i) high mannose type glycan at Asn12, and (ii) a complex type glycan at Asn60 (Sturm & Chrispeels, 1986). PHA-E has an extra glycosylation site at Asn80. PHA-E confers erythroagglutinating activity and PHA-L confers leucoagglutinating and mitogenic activity (Miller *et al.*, 1975).

The PHA-E precursor contains a 21-residue hydrophobic signal peptide and a 254 amino acid mature protein, whereas the PHA-L precursor consists of a 20-residue signal peptide and a 252-residue mature lectin (Hoffman & Donaldson, 1985). The observed difference in migration pattern on an SDS-polyacrylamide gel system may be due to the difference in the number of covalently bound carbohydrate side chains.

PHA binds to complex sugars; PHA-L binds with high affinity to the pentasaccharide Galβ1,4GlcNAcβ1,2(Galβ1,4GlcNAcβ1,6)Man (Hammarstrom *et al.*, 1982). Two bound metal ions (Ca²⁺ and Mn²⁺) present in the vicinity of the sugar-binding sites are vital for the sugar binding capabilities.

Crystallographic structure determination of the PHA-L tetramer revealed that it consists of two canonical dimers packed together through the formation of a close contact between the two outmost strands, creating a large channel in the middle of the tetramer (Hamelryck *et al.*, 1996). This channel is mainly lined with Ser, Thr, and small, apolar residues (Val, Ala, Ile, and Leu) creating a hydrophobic site which recognizes adenine and certain adenine-related cytokinins (Hamelryck *et al.*, 1996).

1.7 Secretory Pathways in Plants and Yeast (*Pichia pastoris*)

The eukaryotic secretory pathway is a complex multi-organelle system which provides for the folding, assembly, post-translational modifications, and transport of newly translated polypeptides. The major compartments of the system are the endoplasmatic reticulum (ER), the Golgi apparatus and the vacuoles, although intermediate compartments exist as well. Figure 1.3 shows a general representation of the secretory pathway in yeast. A hydrophobic signal peptide at the amino-terminus of newly synthesized secretory proteins is necessary for targeting of the polypeptide to the ER; this peptide is removed co-translationally by a signal peptidase as the secretory protein is translocated across the endoplasmatic reticulum membrane (von Heijne, 1994). In the ER, the polypeptides fold, assemble and may undergo additional processing. As this environment contains very high concentrations of unfolded polypeptides, chaperones, foldases, and other enzymes are necessary to assist folding and limit aggregation of the polypeptide (Gething & Sambrook, 1992). In addition to folding and assembly, polypeptides may be subject to other post-translational modifications in the ER, including the addition of carbohydrate groups.

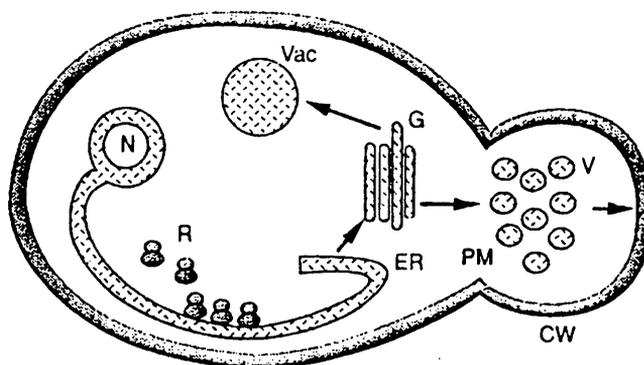


Figure 1.3 Yeast secretory pathway. Arrows indicate the routes taken by the secretory proteins from the ER to the vacuole or plasmamembrane (Romanos *et al.*, 1992). N: nucleus; R: ribosomes; ER: endoplasmatic reticulum; G: Golgi apparatus; Vac: vacuole; V: secretory vesicles; PM: plasmamembrane; CW: cell wall.

In all eukaryotes, *N*-glycosylation begins in the ER with the transfer of a lipid-linked oligosaccharide unit, $\text{Glc}_3\text{Man}_9\text{GlcNAc}_2$, to Asn in the sequence Asn-X-Ser/Thr, where X can be any residue except proline. This oligosaccharide is then trimmed to $\text{Man}_8\text{GlcNAc}_2$. At this point, glycosylation patterns of lower and higher eukaryotes begin to differ. In plants, as the glycoprotein is transported in the secretory pathway from the ER to the Golgi apparatus, the precursor oligosaccharide is processed into high-mannose-type, paucimannosidic-type, hybrid-type and/or complex-type *N*-glycans (Lerouge *et al.*, 1998). These sugar moieties are necessary for efficient secretion of plant glycoproteins and have no specific function in the transport of the glycoproteins into the plant vacuole (Lerouge *et al.*, 1998).

The carbohydrate structures added to secreted proteins in the yeast *Saccharomyces cerevisiae* and *Pichia pastoris* differ from those of plants; the structures are of the high-mannose type only. Comparison of invertase secreted by both yeasts has revealed that although the oligosaccharide chains are of the high-mannose type, the length of the chains of *P. pastoris*-secreted invertase is much shorter than those from *S. cerevisiae* ($\text{Man}_{8-9}\text{GlcNAc}_2$ as opposed to $\text{Man}_{50-150}\text{GlcNAc}_2$, respectively; (Tschopp *et al.*, 1987); (Grinna & Tschopp, 1989)). Also, glycans on *P. pastoris*-secreted invertase do not have the terminal $\alpha 1,3$ -linked mannose residues that are characteristic of *S. cerevisiae* core oligosaccharides (Cregg *et al.*, 1993); *S. cerevisiae* glycans are responsible for the highly antigenic nature of their glycoproteins and make their recombinant proteins unsuitable for human pharmaceutical use (Romanos *et al.*, 1992). Other foreign proteins secreted from *Pichia pastoris* are not subjected to

the extensive mannosylation (hyperglycosylation) that commonly occurs in proteins secreted from *S. cerevisiae* (Montesino *et al.*, 1998).

In *Pichia pastoris*, *O*-linked oligosaccharides, solely composed of mannose residues, are added at specific Ser and Thr residues which are not necessarily the same as found in the native host (Higgins & Cregg, 1998). In plants, *O*-linked glycosylation of secretory proteins is much more complex. For example, *O*-linked glycans of the precursor of the sweet potato sporamin contained galactose and arabinose as major sugar components (Matsuoka *et al.*, 1995).

The pathways for secretion and transport of proteins to the vacuole diverge at the Golgi apparatus, with secretion being the default pathway for soluble proteins. A targeting signal within the protein is required for further transport to the vacuole. With plants, three independent signals that direct protein to plant vacuoles exist: N-terminal pro-peptides, e.g. sweet potato sporamin (Neuhaus & Rogers, 1998); C-terminal pro-peptides, e.g. barley lectin (Neuhaus & Rogers, 1998); regions within mature proteins, e.g. castor bean ricin (Frigerio *et al.*, 1998). There is no homology between these sorting signals, which appear to be unique to plants, and a distinct receptor is implicated in the recognition of each signal (Vitale & Raikhel, 1999). Sequence comparison and mutagenesis analysis of the C-terminal pro-peptide of barley lectin (Dombrowski *et al.*, 1993) and tobacco chitinase A (Neuhaus *et al.*, 1994) suggest a common physicochemical property, rather than a primary sequence, provides the vacuolar targeting determinant for these proteins. In contrast, the N-terminal pro-peptides of sweet potato sporamin (Matsuoka & Nakamura, 1991) and barley aleurain (Holwerda *et al.*, 1992) contain a conserved amino acid sequence, NPIR (Nakamura & Matsuoka, 1993).

As with plants, targeting signals of yeast vacuolar proteins also reside in distinct protein domains (Rothman *et al.*, 1989). For example, vacuolar precursor proteins of carboxypeptidase Y (Valls *et al.*, 1987) and proteinase A (Ammerer *et al.*, 1986) contain distinct N-terminal pro-peptide domains which are cleaved upon delivery to the vacuole. However, plant and yeast vacuolar sorting machineries appear to be different since plant precursor proteins expressed in yeast are targeted correctly to the yeast secretory pathway, but fail to be transported to the yeast vacuole and are secreted (Nagahora *et al.*, 1992; Chao & Etzler, 1994). On the other hand, PHA-L expressed in *Saccharomyces cerevisiae* accumulates mainly in the vacuole (Tague & Chrispeels, 1987) due to the presence of 'cryptic' vacuolar targeting determinants (Von Schauwen & Chrispeels, 1993).

In *Pichia pastoris*, secretion of recombinant protein is favoured over intracellular expression because this yeast secretes only low levels of endogenous proteins which makes purification easier. An extensive list of heterologous proteins expressed in *P. pastoris* is given by Cereghino & Cregg (2000). Several different secretion signal sequences have been used successfully to target recombinant proteins to the secretory pathway, including the native secretion signal present on some plant proteins, e.g. barley α -amylase (Juge *et al.*, 1996) and maize cytokinin oxidase (Morris *et al.*, 1999). the *P. pastoris* acid phosphatase (*PHO1*) and *S. cerevisiae* invertase (*SUC2*) signals have also been used to direct heterologous proteins for secretion. However, the most frequently used signal sequence in recombinant protein expression in *Pichia pastoris* is the α -factor prepro-peptide from *S. cerevisiae* (Cereghino & Cregg, 2000). This sequence consists of a hydrophobic 19-amino acid pre-sequence followed by a 66-residue pro-sequence containing three consensus N-linked

glycosylation sites (Kurjan & Herskowitz, 1982). In some cases, these pro-peptide *N*-linked oligosaccharides are believed to facilitate secretion in *S. cerevisiae* (Kjeldsen *et al.*, 1998). Processing of the α -factor prepro-peptide involves three steps: removal of the pre-signal by signal peptidase in the ER; cleavage of the pro-peptide by the dibasic Kex2 endopeptidase on the carboxyl side of Lys-Arg in the Golgi apparatus; removal of the remaining N-terminal Glu-Ala repeats by the Ste13 dipeptidyl aminopeptidase also in the Golgi apparatus (Brake *et al.*, 1984).

1.8 Research Objectives

The previous sections described plant lectins and the various strategies these proteins employed to specifically recognize and interact with carbohydrates. Also, the importance of plant lectins pertaining to their exploitable properties such as insect resistance was underlined. At the Department of Biological Sciences, University of Durham, different experiments are being carried out in exploiting the plant lectin insecticidal properties. Studies are being carried out in order to elucidate the mode of action of these lectins. This is done by analysing the effects of plant lectins (and chimeras) on insects fed on artificial diet and transgenic crops, by performing binding studies with immunogold labelled lectins, by identifying gut receptors (glycoprotein), and by carrying out mutagenesis on the carbohydrate-binding sites of the proteins.

The main objective of this thesis is to provide efficient heterologous expression systems for recombinant functional plant lectin production in order to complement the on-going mode of action studies. Many plant lectins occur as heterogenous mixtures of isoforms *in planta*. The difficulty of separating multiple lectin isoforms by conventional techniques has led to functional properties and biological activities

being defined for protein mixtures. Sometimes differences in sequence between isoforms cause significant effects on the biological activity of the molecule (e.g. PHA-E and PHA-L). Thus, there is a need to develop a system to produce a single, well-defined lectin species. In addition, such a system is a prerequisite for site-directed mutagenesis and creation of lectin chimeras.

Therefore, the specific objectives of this work are:

- (i) to produce functional snowdrop lectin (GNA) in *Escherichia coli* and to carry out mutagenesis of the sugar-binding site utilizing this system;
- (ii) to produce functional snowdrop lectin (GNA) and phytohaemagglutinin (PHA) in *Pichia pastoris* utilizing commercial and novel constructs;
- (iii) to produce a functional GNA-GFP chimeric protein in *Pichia pastoris*.

Chapter 2

Materials & Methods

2.1 Microorganisms

Plasmid DNAs were propagated in the following *Escherichia coli* strains: DH5 α (genotype: F⁻ ϕ 80dlacZ Δ M15, Δ (lacZYA-argF)U169, *recA1*, *endA1*, *gyrA96*, *thi-1*, *hsdR17*(r_K⁻, m_K⁺), *supE44*, λ ⁻, *relA1*, *deoR*, *phoA*), TOP10F' (genotype: F'[lacI^q, Tn10(*tet*^R)], *mcrA*, Δ (*mcr-hsdRMS-mcrBC*), ϕ 80lacZ Δ M15, Δ lacX74, *deoR*, *recA1*, *araD139*, Δ (*ara-leu*)7697, *galU*, *galK*, *rpsL*(*str*^R), *endA1*, *nupG*), and GM119 (genotype: F⁻, *fhuA2*, *lacY1*, *tsx-1*, *gluV44*(AS), *galT22*, *galK2*(Oc), LAM⁻, *dcm-6*, *dam-3*, *mtlA2*, *metB1*, *thi-1*). DH5 α , TOP10F,' and GM119 were obtained from Gibco BRL, Invitrogen, and Department of Biological Sciences (Durham), respectively.

The *Escherichia coli* strain BL21(DE3)pLysS (genotype: F⁻, *ompT*, *hsdS_B*(r_B⁻, m_B⁻), *dcm*, *gal*, λ (DE3), pLysS(cm^R)) from Stratagene was used for expression of wild-type GNA and the GNA variants. *Pichia pastoris* strains GS115 (genotype: *his4*) and KM71 (genotype: *arg4*, *his4*, *aox1::ARG4*) from Invitrogen were used for expression of wild-type GNA, GFP, GNA-GFP fusions, and wild-type PHA (E and L forms) after stable integration of the expression construct into the *AOX1* locus.

2.2 Chemicals and Reagents

All general chemicals and reagents of analytical grade were obtained from BDH or Sigma Chemical Company (Poole, Dorset, UK), unless otherwise stated.

2.3 Standard Molecular Techniques

All basic molecular techniques employed were standard practice at the Biological Sciences Department, University of Durham, and were based on protocols in *Molecular Cloning: A Laboratory Manual* (Sambrook *et al.*, 1989).

2.3.1 Bacterial Culture

For routine work, liquid bacterial cultures were grown in Luria-Bertani (LB; 10 g peptone (Becton Dickinson, Cowley, Oxon, UK), 10 g NaCl, 5 g yeast extract (Umpath Ltd., Basingstoke, Hampshire, UK) per litre) broth from a single colony picked from a fresh agar plate. Liquid cultures were grown at 37 °C on a rotary shaker (200rpm). For bacterial cultures on agar, a base of 1.5 % bacteriological agar in LB broth was used and inoculated plates were incubated at 37 °C. Where appropriate, antibiotics for selection were added to culture media. Where Zeocin (Invitrogen) was used, bacterial cultures were grown in low salt LB broth (as above, but containing 5 g NaCl per litre) at pH 7.0 - 7.5.

2.3.2 Competent Cells

Competent cells of *Escherichia coli* used for plasmid transformation were commercially obtained as follows: DH5 α (Gibco-BRL), Top10F' (Invitrogen), and

BL21(DE3)pLysS (Stratagene).

Competent GM119 was prepared by the one-step preparation method of Chung *et al.* (1988). Briefly, *E. coli* GM119 cells were grown in liquid LB broth to exponential phase ($OD_{600nm} = 0.6$) and pelleted by centrifugation at 1000 g for 10 minutes at 4 °C. The cell pellet was resuspended at one-tenth the original culture volume in ice-cold transformation and storage solution (LB with 10 % (w/v) PEG 8000, 5 % (v/v) dimethylsulfoxide (DMSO), and 25 mM $MgCl_2$; final pH 6.5). A 0.1 ml aliquot of cells was used for transformation.

2.3.3 Transformation of *E. coli*

Transformation of DNA into *E. coli* chemically competent cells was performed using standard procedures (Sambrook *et al.*, 1989). Tubes containing competent cells were removed from -80 °C and thawed on ice. DNA for transformation (1 - 5 μ l) was added to the cells and the tubes mixed very gently. After leaving the cells on ice for 30 minutes, the cells were heat-shocked by placing the tubes in a 42 °C water bath for 90 seconds. The tubes were then placed back on ice for 1 - 2 minutes. To each tube, LB broth (0.5 ml) was added, and the tubes were incubated at 37 °C for 45 minutes to 1 hour. The resulting cell suspension was selected for transformants by plating out aliquots (usually 100 and 400 μ l) of the suspension on LB-agar plates containing the appropriate antibiotic. When transforming cells with plasmids allowing blue-white colour selection, 40 μ g/ml 5-bromo-4-chloro-3-indolyl- β -D-galactoside (X-Gal) and 0.1 mM isopropyl- β -D-thiogalactoside (IPTG) were added to the agar plates. Plates were incubated at 37 °C for 16 hours.

2.3.4 Plasmid DNA Isolation

The method for recombinant plasmid DNA isolation was based on the alkaline lysis procedure (Birnboim & Doly, 1979). Single colonies of bacteria containing the appropriate plasmid were cultured overnight under the appropriate antibiotic selection in 5 ml (miniprep) or 50 ml (midiprep) of LB broth. The WizardTM Plus SV Minipreps DNA Purification System (Promega) was used for the minipreps, and when the plasmid was needed in greater quantity for *Pichia* transformation, the QIAfilter Plasmid Midi Kit (QIAGEN) was used for the midipreps. Protocols supplied by the manufacturers were followed.

2.3.5 DNA Concentration Determination

The concentration of purified plasmid DNA was determined spectrophotometrically by optical density measurement at 260 nm using a Beckmann DU 7500 Spectrophotometer (assuming $OD_{260\text{ nm}} = 1$ corresponds to a concentration of double stranded DNA = 50 $\mu\text{g/ml}$).

2.3.6 DNA Digestion with Restriction Enzymes

Restriction enzyme digests were performed using commercially available enzymes and buffers (Promega, New England Biolabs, Boehringer-Mannheim, MBI Fermentas). Typically, digests were performed using 0.5 - 2.0 μg of plasmid DNA and 2 - 10 units (where 1 unit will completely digest 1 μg of DNA in a total volume of 50 μl in 1 hour under optimal conditions) of each restriction enzyme. Digests were

incubated at 37 °C for 1 - 3 hours before separation of fragments by agarose gel electrophoresis.

2.3.7 Phenol Extraction and Ethanol Precipitation of DNA

Extraction of DNA was carried out by adding an equal volume of phenol:chloroform:isoamylalcohol (25:24:1) to the DNA solution, vortexing for 1 minute and centrifuging at 13000 g for 5 minutes. The upper aqueous phase was transferred to a fresh tube and an equal volume of chloroform:isoamylalcohol (24:1) added. The sample was vortexed for 30 seconds, and centrifuged at 13000 g for 2 minutes. After transferring the upper aqueous phase to a fresh tube, the DNA was then precipitated by adding 0.1 volume of 3 M sodium acetate pH 5.2 and 2 volumes of 100 % ethanol. The sample was stored at -20 °C for 1 hour and then centrifuged at 13000 g at 4 °C for 10 minutes. The supernatant was removed and 200 μ l of 70 % ethanol added. After centrifugation at 13000 g at 4 °C for 5 minutes, the supernatant was then discarded and the DNA pellet dried briefly in a vacuum dessicator. The DNA was dissolved in 50 μ l H₂O.

2.3.8 DNA Ligation

Ligation of DNA fragments were usually performed in 10 μ l reactions, using commercially available T4 DNA ligase and appropriate buffers (Boehringer-Mannheim). Digested, purified vector and insert(s) were added in an approximate ratio of 1:3, and the reaction was made up to 10 μ l with distilled water. Ligations were incubated at 14 °C for 16 hours prior to transformation into *E. coli*.

2.3.9 Oligonucleotides

Oligonucleotide primers were synthesized by either Perkin-Elmer or MWG Biotech (Ebersberg, Germany; URL:

http://www.mwgdna.com/services/orders/order_oligo/index.htm).

2.3.10 Standard DNA Amplification using PCR (Polymerase Chain Reaction)

PCR was performed using standard conditions. Each reaction (25 - 100 μ l) contained reaction buffer consisting of 2.5 mM MgCl₂, 0.2 mM each of dATP, dCTP, dGTP and dTTP, 2 oligonucleotide primers (each 1 μ M), DNA template (50 - 80 ng) and 1 unit of *Pfu* (Stratgene) or *Tli* DNA polymerase (for expression constructs) or *Taq* DNA polymerase (for PCR-screening; colony PCR), where 1 unit catalyses the incorporation of 10 nmol of dNTP into acid-insoluble form in 30 minutes at 74 °C. Typically, for amplification using plasmid templates, PCR amplifications consisted of a denaturation at 94 °C for 5 minutes, 30 cycles of the conditions: denaturation at 94 °C for 30 seconds, annealing at 55 °C for 1 minute, extension at 72 °C for 1 minute, and a further extension at 72 °C for 9 minutes. Reactions were carried out using a Perkin Elmer 2400 thermal cycler.

2.3.11 Colony PCR

A rapid method to identify recombinant plasmids in *E. coli* was by PCR-screening of ten to twenty colony transformants. From a selected colony, bacterial cells were picked and added to 50 μ l of colony lysis buffer (20 mM Tris-HCl at pH 8.3, 2 mM EDTA, 1 % Triton X-100) in a 0.5 ml microcentrifuge tube. The cell suspension was

mixed by vigorous vortexing for 30 seconds and then incubated at 94 °C for 15 minutes. After centrifugation, 5 μ l of the supernatant was used in a standard PCR.

2.3.12 Agarose Gel Electrophoresis of DNA

The separation of DNA by electrophoresis in agarose gel was carried out as described by Sambrook *et al.* (1989). Briefly, DNA was size-fractionated in gels containing 0.7 - 1.0 % (w/v) agarose (Gibco BRL) in TAE buffer (40 mM Tris-acetic acid at pH 7.7, 1 mM EDTA) containing 0.5 μ g/ml ethidium bromide for visualisation of DNA under UV light. Electrophoresis was carried out in a Pharmacia GNA-100 mini-gel system or in a NBL medium-size gel apparatus. One-fifth of sample volume of 6 x gel loading dye mix (10 mM Tris-HCl at pH 8.0, 10 mM EDTA, 30 % (w/v) glycerol, 0.1 % (v/v) fast orange G) was added to the DNA samples prior to loading. The gels were run horizontally at 50 - 100 V at room temperature in TAE buffer containing 0.5 μ g/ml ethidium bromide. *HindIII*-, *PstI*- or *Eco47I*-digested lambda DNA (Northumbria Biochemicals Ltd., Cramlington, Northumberland, UK) was used as a molecular size marker. Gels were photographed under UV illumination and recorded using a red-orange filter (Kodak 23A Wratten) and Polaroid Film type 667-3000.

2.3.13 DNA Purification from Agarose Gel

DNA fragments were isolated from low melt agarose gels using silica fines (Sigma, 0.5 - 1.0 μ m particle size). The DNA band of interest was excised from the gel in as small a volume of gel as possible and transferred to a 1.5 ml polypropylene tube. To the tube, 1 ml of 1M NaI was added and the sample incubated at 70 °C for 10 minutes. Twenty μ l of silica fines was added and after incubation at room temperature

for 20 minutes with occasional shaking, the bound DNA was pelleted by spinning for 1 minute in a benchtop microcentrifuge. The pellet was washed with 1 ml of 70 % (v/v) ethanol and air-dried. The DNA was eluted by incubating the pellet in 20 μ l of sterile water for 5 minutes at room temperature. After centrifugation to remove the silica, the supernatant containing the DNA was recovered and stored at -20 °C.

2.3.14 DNA Sequencing

Sequencing of plasmid template DNAs was carried out by a modification of the dideoxy chain termination method (Sanger *et al.*, 1977) using fluorescent-labelled dye terminators and an AmpliTaq cycle sequencing kit (Perkin-Elmer Applied Biosystems, Warrington, Cheshire, UK). Reaction products were analysed on a Perkin-Elmer ABI 373 stretch or 377 DNA sequencer provided by the DNA sequencing services at the Department of Biological Sciences, University of Durham, Durham, UK. All final constructs described were sequenced using standard vector primers, or primers used for creating the constructs, or a combination of the two, to verify that no base misincorporations had occurred during PCR amplification steps.

2.3.15 Glycerol Stocks

Single colonies of *E. coli* or *P. pastoris* harbouring the desired recombinant DNA were inoculated into the appropriate liquid growth medium (LB for *E. coli* and YPD for *P. pastoris*) containing the appropriate antibiotic (ampicillin, chloramphenicol, and/or Zeocin) and grown overnight at 37 °C. An equal volume of liquid culture was added to 500 μ l of sterile 80 % (v/v) glycerol in 1 ml glass vials (BDH), resuspended by vortexing, and stored at -80 °C.

2.4 Standard Protein Analysis

2.4.1 SDS-PAGE

Protein samples were analysed by sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) on a 12.5 % or 15 % resolving gel (12.5 % or 15 % (w/v) acrylamide, 0.333 % or 0.4 % (w/v) bisacrylamide, 0.375 M Tris/HCl (pH 8.8), 0.1 % (w/v) SDS, 0.075 % (w/v) ammonium persulphate, 0.05 % (v/v) *N,N,N',N'*-tetramethylethylenediamine) and 2.5 % stacking gel (2.5 % acrylamide, 0.1 % (w/v) bisacrylamide, 0.125 M Tris/HCl (pH 6.8), 0.1 % (w/v) SDS, 0.1 % (w/v) ammonium persulphate, 0.075 % (v/v) *N,N,N',N'*-tetramethylethylenediamine) according to Laemmli (1970). Minigels (8 x 10 cm) were prepared as described by Hames and Rickwood (1990) and run in electrophoresis buffer (0.025 M Tris/HCl, 0.192 M glycine, 0.1 % (w/v) SDS; pH 8.3) at constant voltage (100 - 150 V) using an ATTO AE-6450 gel tank apparatus (Genetic Research Instrumentation Ltd., Dunmow, Essex, UK). The samples were prepared by adding an equal volume of 2 x SDS sample buffer (0.2 M Tris/HCl, 20 % (v/v) glycerol, 2 % (w/v) SDS, 0.002 % (w/v) bromophenol blue; pH 6.8), treated with 10 % β -mercaptoethanol and heated in boiling water for 10 minutes before loading onto gel. Protein bands were visualized by either Kenacid Blue staining (0.05 % (w/v) Kenacid Blue, 40 % methanol, 7 % glacial acetic acid) or silver staining (BioRad). A molecular weight marker (SDS7; Sigma) was used to calibrate the gels. This marker set contained bovine albumin (66 kDa), ovalbumin (45 kDa), glyceraldehyde-3-phosphate dehydrogenase (36 kDa), carbonic anhydrase (29 kDa), trypsinogen (bovine pancreas; 24 kDa), trypsin inhibitor (soybean; 20 kDa), α -lactalbumin (14.2 kDa).

2.4.2 Immunoblotting

Electrophoretic transfer of proteins from the protein gel to the nitrocellulose membrane (grade BA85, Schleicher and Schuell Inc., Anderman & Co. Ltd., Kingston-upon-Thames, Surrey, UK) was carried out by the semi-dry blotting technique of Khyse-Andersen (1974). The membrane and 3MM paper (Whatman Ltd., Maidstone, Kent, UK) were cut to the same size as the gel. After electrophoresis, the gel was assembled onto the ATTO blotting apparatus (Genetic Research Instrumentation Ltd., Dunmow, Essex, UK) in the following order: ANODE; 2 sheets of 3MM paper soaked in anode buffer 1 (0.3 M Tris/HCl, 20 % (v/v) methanol; pH 10.4); 1 sheet of 3MM paper soaked in anode buffer 2 (0.025 M Tris/HCl, 20 % (v/v) methanol; pH 10.4); 1 sheet of nitrocellulose membrane soaked in distilled water; the gel; 3 sheets of 3MM paper soaked in cathode buffer (0.025 M Tris/HCl, 40 mM 6-aminohexanoic acid, 20 % (v/v) methanol; pH 9.4); CATHODE. Electroblothing was conducted at constant current at between 125 - 150 mA (~ 2.5 mA/cm²) for 45 minutes. The nitrocellulose membrane was stored between 3MM paper at 4 °C.

For immunodetection, the membrane was incubated in phosphate buffered saline (PBS: 8 g NaCl, 0.2 g KCl, 1.15 g Na₂HPO₄, 0.2 g KH₂PO₄ per litre) pH 7.4, containing 5 % nonfat milk powder (Marvel) and 1 % Tween 20 at room temperature for 1 hour with gentle shaking. Polyclonal rabbit anti-GNA antiserum (produced by Drs R.R.D. Croy and L.N. Gatehouse, Department of Biological Sciences, University of Durham), rabbit anti-PHA antiserum (Vector Laboratories) or mouse anti-GFP antiserum (Clontech) were diluted (1:10,000, 1:20,000 or 1:10,000, respectively) with PBS containing 5 % nonfat milk powder and 0.1 % Tween 20. The membrane was

incubated in primary antibody solution at room temperature for 1.5 hours with gentle shaking. The membrane was then washed in antibody dilution buffer for 3 x 5 minutes at room temperature. Goat anti-rabbit or -mouse (for anti-GFP detection) IgG horseradish peroxidase conjugate (BioRad) was used as secondary antibody (1:10,000 dilution, as above) to treat the membrane at room temperature for 1.5 hours followed by 3 x 5 minutes wash in antibody dilution buffer and a rinse in distilled water. Enhanced chemiluminescence (ECL) reagents (Amersham) were used to detect the specifically bound secondary antibody as instructed by the manufacturer, and bands were visualised by exposure to X-ray film (Fuji-RX; Fuji Photo Film Ltd., London, UK). Autoradiographs were developed either manually or with an automatic developer (X-ograph Imaging Systems Compact X4, Malmesbury, Wiltshire, UK).

2.4.3 Electroblotting of Proteins to PVDF Membranes

Prior to running samples on an SDS-PAGE gel, thioglycolic acid (2 μ M) was added to the electrophoresis anode buffer and the gel pre-run for 30 minutes to remove any chemicals which could potentially N-terminally block polypeptides. Proteins separated by SDS-PAGE were electrophoretically transferred from gels to polyvinylidene fluoride (PVDF; ProBlott; Applied Biosystems) membranes according to LeGendre and Matsudaira (1988) using the ATTO semi-dry electroblotting apparatus. The PVDF membrane was briefly prewetted in 100% methanol, rinsed with distilled water and equilibrated in transfer buffer (CAPS; 10 mM 3-(cyclohexylamino)-1-propanesulfonic acid, 10 % methanol; pH 11) for a minimum of 15 minutes. The blotting assembly was set up as described above (see section 2.4.2), except that the gel was rinsed in the transfer buffer for 5 minutes and all sheets of 3MM paper were soaked in the same buffer prior to electroblotting. After blotting, to

check for successful transfer, the ProBlott membrane was rinsed in distilled water, stained for 1 minute in staining solution (0.1 % Kenacid Blue, 50 % methanol), destained with several changes of destaining solution (50 % methanol, 10 % acetic acid), rinsed again with several changes of distilled water and air-dried. The bands of interest were excised from the membrane and used for protein sequencing.

2.4.4 N-Terminal Sequencing

N-terminal sequencing was carried out on affinity-purified proteins or protein bands blotted onto Problott PVDF membrane after separation by SDS-PAGE, as described in section 2.4.3, using an Applied Biosystems model 477 amino acid sequencer. A standard procedure for identifying N-terminal sequences was followed, where the machine was run for 6 cycles and the resulting sequence data was analysed for agreement with the amino acid sequences predicted by the nucleotide sequences of the expression constructs. Any ambiguities were resolved by extending the number of sequencing cycles.

2.4.5 Gel Filtration

The state of oligomerisation of the recombinant lectin was analysed by gel filtration using a Superose 12 10/30 column (Pharmacia). The running buffer used was PBS containing 0.2 M mannose at a flow rate of 0.3 ml/min.

2.4.6 Deglycosylation of Protein

Recombinant N-glycosidase F (PGNase-F; Boehringer-Mannheim) hydrolyses all types of asparagine-bound N-glycan chains from glycoproteins provided that the amino and carboxyl group of the asparagine are present in peptide linkages. Protein

samples were denatured by boiling for 10 min in the presence of 1 % (w/v) SDS and 8 % β -mercaptoethanol. Samples were diluted to 0.1 % SDS, 0.8 % β -mercaptoethanol, 20 mM sodium phosphate (pH 7.4), 25 mM EDTA and 2 % Triton X-100. PGNase-F (0.6 units) was added and samples were incubated overnight at 37 °C. Control reactions omitted the enzyme. Following SDS-PAGE, samples were analysed by immunoblotting (see section 2.4.2).

2.4.7 Protein Concentration Determination

Protein content of solutions was determined using a commercial dye-binding assay (BioRad) based on the Bradford (1976) assay, using bovine serum albumin as a protein standard. For determining the concentration of recombinant GNA-GFP nGNA was used as standard. Dilutions of BSA (or nGNA) were prepared in the same buffer as that used for sample preparation. In a microtitre plate, protein samples were diluted to a volume of 160 μ l with distilled water. Forty μ l of the Bradford reagent was added to each sample-containing well. The plate was shaken briefly at 900 rpm for 15 seconds on a microtitre plate shaker and the samples measured at 570 nm using a Dynatech MT 5000 microtitre plate reader. Empty wells were used as blanks.

The concentration of the rGNA was estimated by spectrophotometric analysis using a Beckmann DU 7500 Spectrophotometer. The extinction coefficient for GNA of A_{280} 1 mg/ml = 2.05 was determined empirically.

2.5 Construction of Expression Vectors

Different strategies were employed to create the expression constructs. These are described below. The inserts of all expression vectors were sequenced on both strands

(5' → 3' and 3' → 5') to confirm that no unexpected mutations had been generated during oligonucleotide synthesis, PCR amplification or DNA cloning.

2.5.1 Expression Constructs for GNA in *Escherichia coli*

Plasmid pGNAH2, containing the complete cDNA sequence encoding LECGNA2 (GenBank/EMBL Data Library accession number M55556), one of several GNA isoforms (Van Damme *et al.*, 1991a), has been described previously (Shi *et al.*, 1994). Two expression vectors were constructed. To produce mature LECGNA2 polypeptides containing 105 amino acids, sense (5' CGA TCC ATG GAC AAT ATT TTG TAC TCC GG 3') and antisense (5' ATT AGG ATC CTC ATC CAG TAG CCC AAC G 3') oligonucleotides were synthesized to generate a *NcoI* site in the start codon of the open reading frame and a stop codon after corresponding amino acid 105 of GNA followed by a *BamHI* site, respectively. For the second construct, the same sense but a different antisense primer (5' ATT AGG ATC CTC ATC CGG TGT GAG TTC CAG 3') was used to generate a longer fragment that encodes the four extra C-terminal amino acids (residues 106 - 109; Thr-His-Thr-Gly). The corresponding region from GNA was amplified by PCR from plasmid pGNAH2. The fragments obtained were subcloned into pET11d (Novagen) previously digested with *NcoI* and *BamHI*, resulting in plasmid pGNA105 or pGNA109, depending on the length of the sequence.

2.5.2 Site-Directed Mutagenesis

Site-directed mutagenesis was performed on plasmid pGNA105 as template using the QuikChange™ Site-Directed Mutagenesis Kit (Stratagene) using protocols supplied by the manufacturer. In creating the mutation Tyr₉₇ to Phe₉₇ (Y97F), two mutagenic

oligonucleotide primers, each complementary to opposite strands of the GNA DNA insert, were synthesized as follows: sense (5' GTT GTG ATC TTC GGA ACT GAT 3') and antisense (5' ATC AGT TCC GAA GAT CAC AAC 3') with the nucleotide change shown in bold. The primers were extended during temperature cycling by *Pfu* DNA polymerase generating a mutated plasmid containing staggered nicks. After temperature cycling, the product was digested with *DpnI* to remove the *dam*-methylated parental DNA template. The nicked vector DNA incorporating the desired mutation was then transformed directly into *E. coli*.

2.5.3 Expression Constructs for GNA in *Pichia pastoris*

Construct GNA:GNA was created for the expression of the leader and mature GNA polypeptide (Table 1.1; Figure 2.1). PCR primers were developed based on the LECGNA2 cDNA (in pGNAH2) and corresponded to the first five N-terminal codons of the GNA signal peptide sequence, including a consensus ATG initiation codon (ANN**ATGG**; (Kozak, 1987;1990)), and the last six codons of the mature coding sequence, including the stop codon. The amplified product, flanked by *XhoI* and *XbaI* restriction sites, was inserted into the multiple cloning site of pPICZB (Invitrogen).

Construct GNA:GNA(2) was created for the expression of the leader, mature, and C-terminal extension of GNA (Table 1.1; Figure 2.1). The same protocol was used as for construct GNA:GNA (see above) with the exception that the C-terminal primer corresponds to the last six codons of the GNA C-terminal extension sequence, including the stop codon.

The mature coding sequence of GNA, which was PCR-amplified using the appropriate primers (Table 1.1; Figure 2.1), was cloned in frame and downstream of

Table 1.1 Oligonucleotide sequences of primers used to prepare *Pichia* expression constructs. The PHA-E:GNA and PHA-E:GFP constructs were produced by a two-step process; the product of amplification of the first pair of primers was used as a ‘Megaprimer’ in the second amplification step. M13RP1 = M13 reverse sequencing primer 1 (5’ CACACAGGAAACAGCTATGAC 3’); 5’ AOX1 = *Pichia* alcohol oxidase promoter primer (5’ GACTGGTTCCAATTGACAAGC 3’).

Construct	PCR Primers: N-terminal (5’ – 3’) C-terminal (5’ – 3’)	Vector	Cloning sites	Coding sequence(s)
PHA-E:PHA-E	GCGAATTCACCATGGCTTCTCCAACCTTACTC M13RP1	pPICZ B	<i>Eco</i> RI / <i>Xba</i> I	PHA-E signal peptide : Mature PHA-E
PHA-L:PHA-L	GCGAATTCACCATGGCTTCTCCAAGTTCTTC M13RP1	pPICZ B	<i>Eco</i> RI / <i>Xba</i> I	PHA-L signal peptide : Mature PHA-L
α :PHA-E	GCCTCGAGAAAAGAGCCAGCCAAACCTCCTTCAGC M13RP1	pPICZ α A	<i>Xho</i> I / <i>Xba</i> I	α -factor prepro- : Mature PHA-E
α :PHA-L	GCCTCGAGAAAAGAAGCAACGATATCTACTTCAAC M13RP1	pPICZ α A	<i>Xho</i> I / <i>Xba</i> I	α -factor prepro- : Mature PHA-L
GNA:GNA	ATTACTCGAGAAAATGGCTAAGGCAAGTC TAATTCTAGATTACTTTGCCGTCACAAGC	pPICZ B	<i>Xho</i> I / <i>Xba</i> I	GNA signal peptide : Mature GNA
GNA:GNA(2)	ATTACTCGAGAAAATGGCTAAGGCAAGTC TAATTCTAGATTACTTTGCCGTCACAAGC	pPICZ B	<i>Xho</i> I / <i>Xba</i> I	GNA signal peptide : Mature GNA plus C-terminal extension
α :GNA	ATTACTCGAGAAAAGAGACAATATTTTGTAC TAATTCTAGATCATCCGGTGTGAGTTCC	pPICZ α A	<i>Xho</i> I / <i>Xba</i> I	α -factor prepro- : Mature GNA
α EA:GNA	ATTACTCGAGAAAAGAGAGGCTGAAGCTGACAATATTTGTACTCC TAATTCTAGATCATCCGGTGTGAGTTCC	pPICZ α A	<i>Xho</i> I / <i>Xba</i> I	α -factor prepro- / EAEA repeats : Mature GNA
PHA-E:GNA	1. 5’ AOX1 GGAGTACAAAATATTGTCTGAGTTTGGTGGGTGAG 2. Megaprimer from above TAATTCTAGATCATCCGGTGTGAGTTCC	pPICZ B	<i>Eco</i> RI / <i>Xba</i> I	PHA-E signal peptide : Mature GNA
α EA:GFP	ATTACTCGACGAAGTAAAGGAGAAGAACTTTTTC TAATTCTAGAATTCATTATTTGTAGAGCTCATC	pPICZ α B	<i>Pst</i> I / <i>Xba</i> I	α -factor prepro- / EAEA repeats : GFP
PHA-E:GFP	1. 5’ AOX1 CCAGTGAAAAGTTCTTCTCCTTACTTGAGTTTGGTGGGTGAG 2. Megaprimer from above TAATTCTAGAATTCATTATTTGTAGAGCTCATC	pPICZ B	<i>Eco</i> RI / <i>Xba</i> I	PHA-E signal peptide : GFP

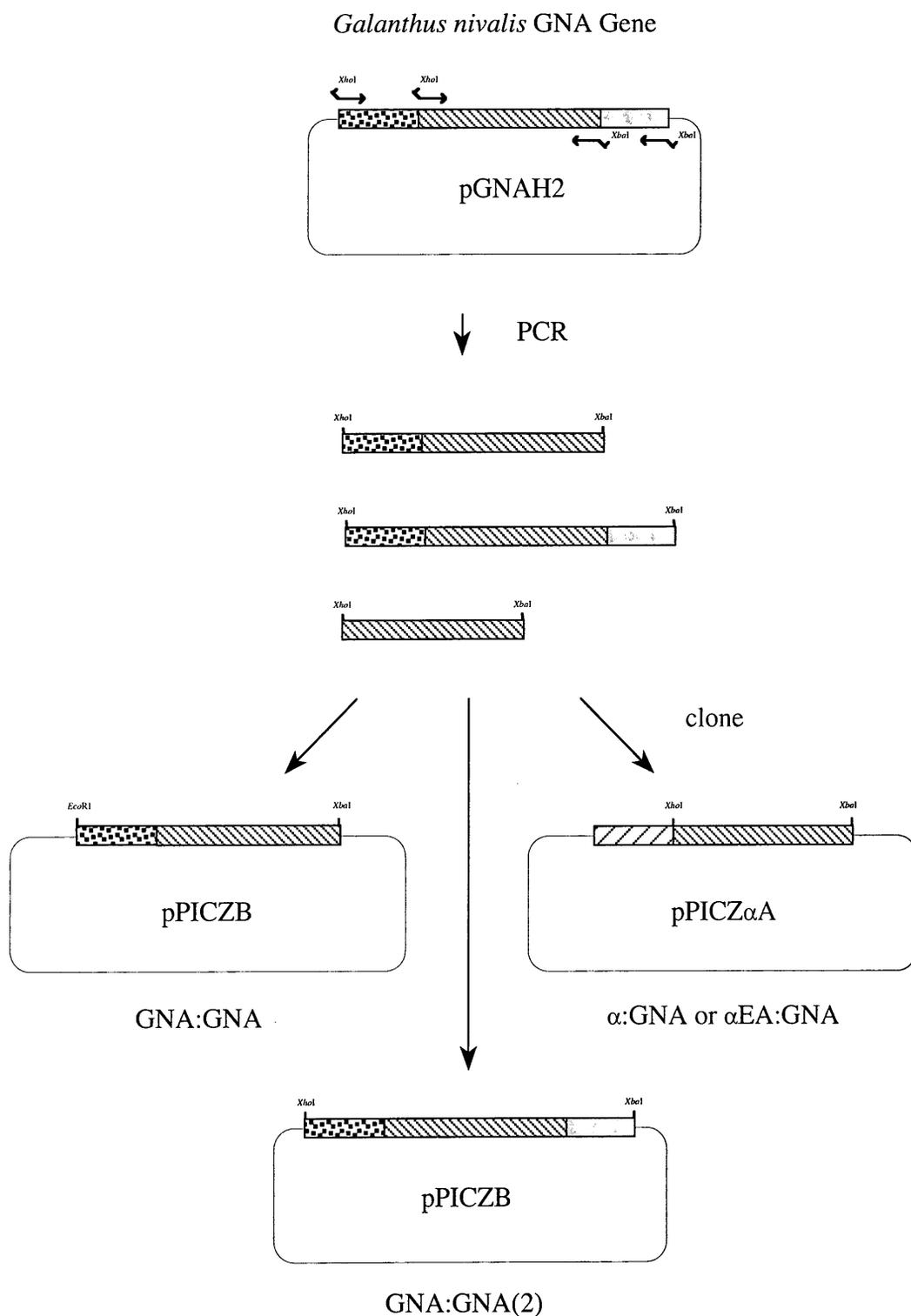


Figure 2.1 Development of expression constructs for GNA. The GNA coding sequence, with or without the leader sequence and/or the C-terminal extension, was amplified by PCR from pGNAH2 and cloned into pPICZB or pPICZ α A expression vector. : GNA leader sequence; : GNA mature coding sequence; : GNA C-terminal extension sequence; : α -factor prepro sequence; : primer (see Table 1.1 for details).

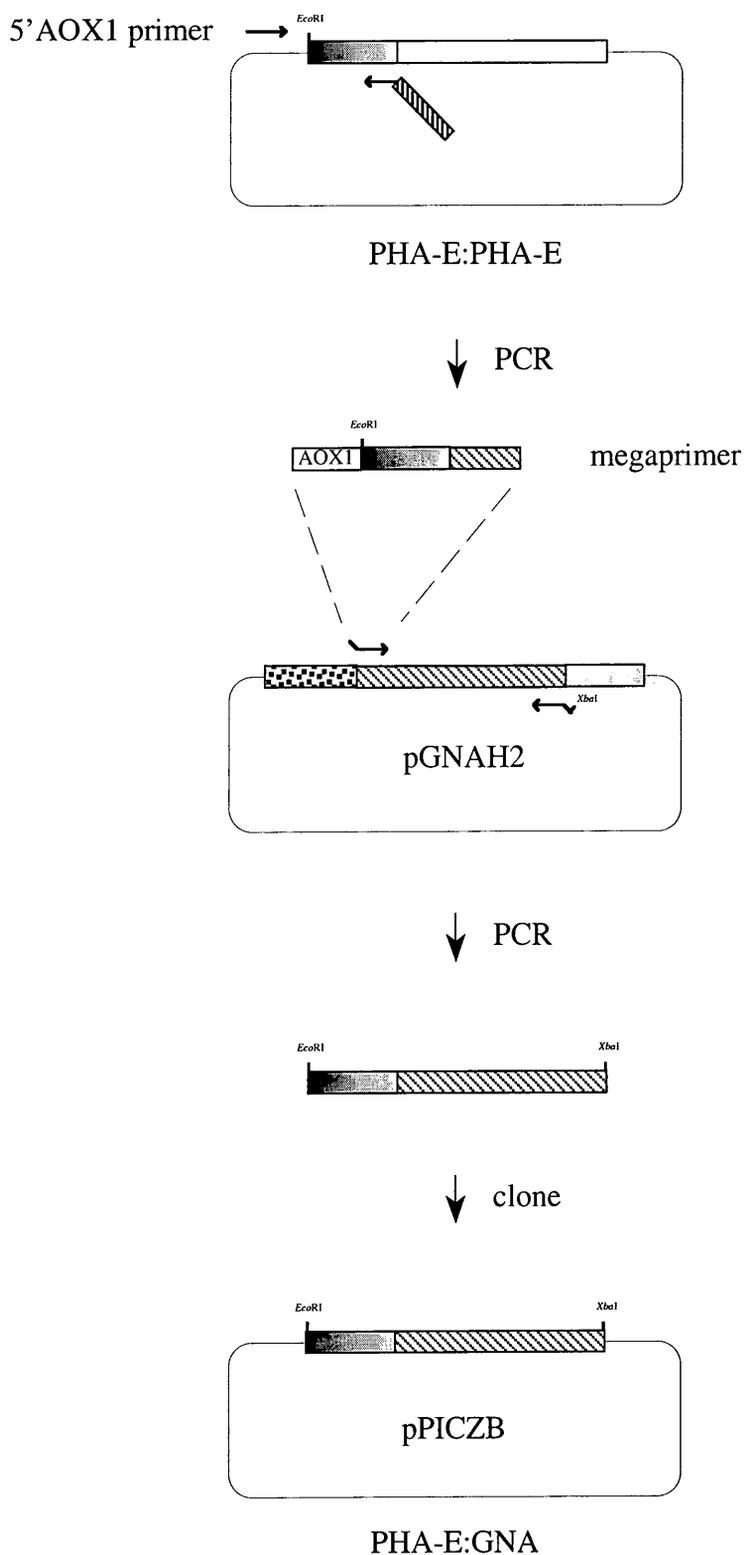


Figure 2.2 Development of expression constructs for GNA *continued*. The PHA-E leader sequence was amplified by PCR from the PHA-E:PHA-E construct. The PCR product was subsequently used as a (mega)primer for the amplification of the mature GNA coding sequence from pGNAH2 and cloned into pPICZB expression vector. : PHA-E leader sequence; : PHA-E mature coding sequence; : GNA leader sequence; : GNA mature coding sequence; : GNA C-terminal extension sequence; : primer (see Table 1.1 for details).

the α -factor prepro--sequence flush with the Kex2 protease cleavage site to yield construct α :GNA. The initiation codon ATG in the α -factor signal sequence in pPICZ α A corresponded to the native initiation codon of the *AOX1* gene. A second α -factor construct (α EA:GNA) contained two Glu-Ala repeats (EAEA) between the Kex2 protease cleavage site and mature GNA.

For construct PHA-E:GNA, mature GNA fused to the PHA-E signal sequence was created by the megaprimer method (Sarkar & Sommer, 1990) using two sequential polymerase chain reactions. Firstly, the PHA-E:PHA-E construct (in pPICZB; see below) was used as template for amplification of the PHA-E native leader sequence using the 5' *AOX1* primer (Invitrogen) and a C-terminal primer which introduced the first six codons of the mature coding sequence of GNA (omitting the methionine initiation codon) immediately after the last six codons of the PHA-E signal peptide coding sequence. The PCR parameters were as follows: 94 °C for 5 minutes; 30 cycles of 94 °C for 30 seconds, 50 °C for 30 seconds, 74 °C for 30 seconds; 74 °C for 5 minutes. The resulting PCR product was purified and used as a megaprimer in combination with a GNA C-terminal primer to produce the PHA-E:GNA fusion using the LECGNA2 cDNA (in pGNAH2) sequence as template. The reaction parameters were: 94 °C for 5 minutes; 30 cycles of 94 °C for 30 seconds, 45 °C for 30 seconds, 74 °C for 30 seconds; 74 °C for 5 minutes. In all cases above, PCR products were cloned directly into the appropriate pPIC vector (Table 1.1; Figure 2.2).

2.5.4 Expression Constructs for PHA-E and PHA-L in *Pichia pastoris*

Genomic DNA isolated from *Phaseolus vulgaris* cv. Tendergreen was used as template for the PCR amplification of PHA coding sequences. Primer sequences were

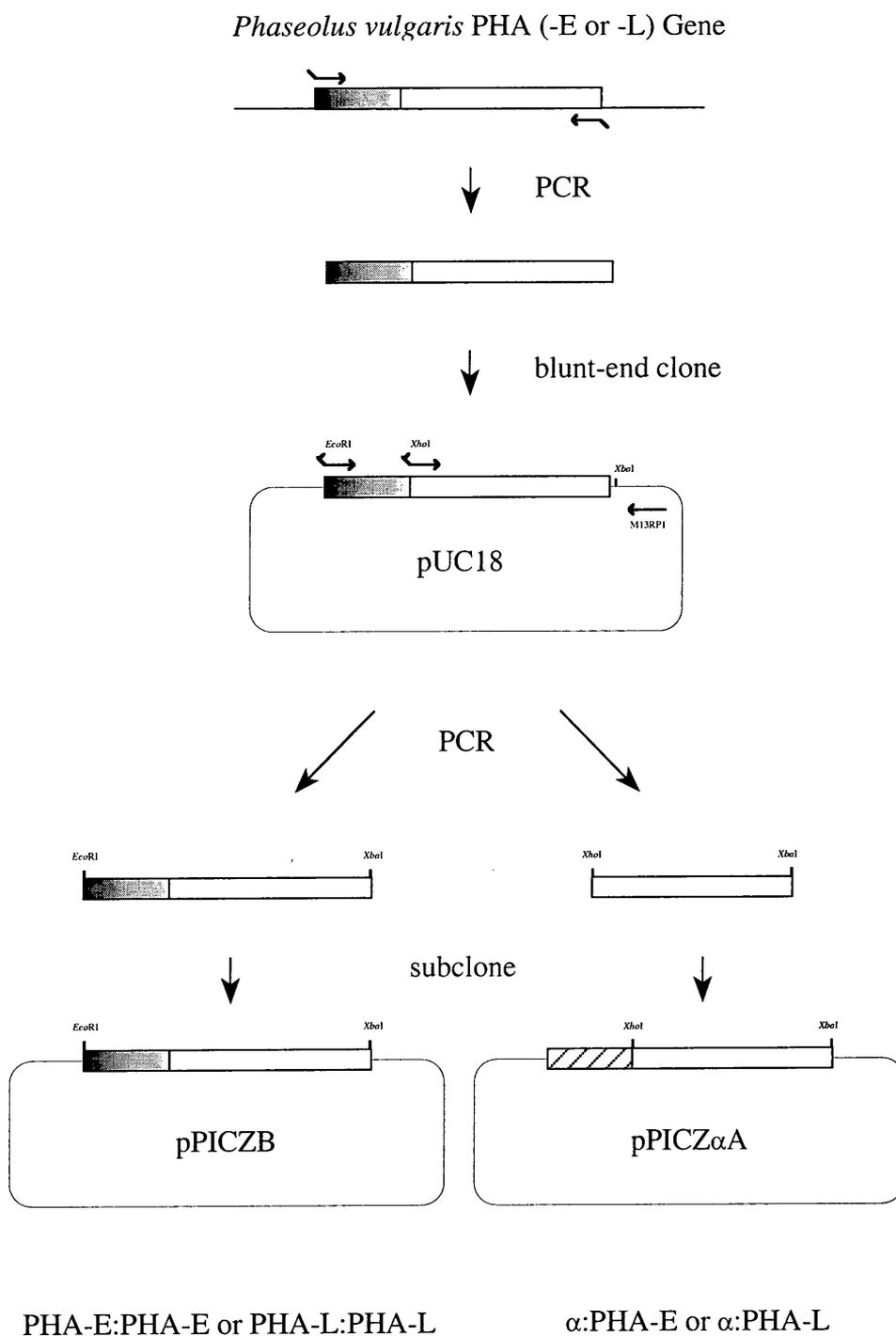


Figure 2.3 Development of expression constructs for PHA-E and PHA-L. The full-length PHA coding sequence was amplified by PCR from whole genomic *Phaseolus vulgaris* cv. Tendergreen DNA and cloned into pUC18. The PHA gene, with or without its native leader sequence, was amplified by PCR and cloned into pCR-Script (not shown) before subcloned into pPICZB or pPICZ α A expression vector, respectively. : PHA (-E or -L) leader sequence; : PHA (-E or -L) mature coding sequence; : α -factor prepro sequence; : primer (see Table 1.1 for details).

designed using the published sequences of PHA-E and PHA-L (GeneBank/EMBL Data Library accession numbers X02408 and X02409, respectively; (Hoffman & Donaldson, 1985)). PCR primers corresponded to the first six N-terminal codons of the PHA signal peptide sequence and the last six codons (including the stop codon) of the mature coding sequence and a further 5 bp of the 3' UTR. A *Bam*HI site was included at the 5' end of each primer. Primer sequences were: PHA-E N-terminal primer, 5' CGG ATC CAT GGC TTC CTC CAA CTT AC 3'; PHA-E C-terminal primer, 5' CGG ATC CTG GAG TCT AGA GGA TTT GGT TGA G 3'; PHA-L N-terminal primer, 5' CGG ATC CCA TGG CTT CCT CCA AGT TC 3'; PHA-E C-terminal primer, 5' CGG ATC CTG GAG TCT AGA GGA TTT TGT TGA G 3'.

After amplification, PCR products were blunt-end cloned into pUC18.

To produce *Pichia* expression constructs (see Table 1.1; Figure 2.3), modified N-terminal oligonucleotides were used to amplify PHA template DNA, in combination with a generic M13 sequencing primer (which amplifies across the *Xba* I site containing a stop codon in the PHA coding sequence). Primers for expression using the native signal peptide sequence contained a consensus ATG initiation codon. Primers for expression using the α -factor prepro--sequence needed no initiation codon and the mature coding sequence was fused flush to the Kex2 protease cleavage site. PCR products were cloned into pCR-Script (Stratagene) and subsequently subcloned into the respective pPIC plasmid vector using restriction sites incorporated into the primer sequences (Table 1.1).

2.5.5 Expression Constructs for GFP in *Pichia pastoris*

Primers used for the construction of GFP expression constructs are shown in Table 1.1. For construct α EA:GFP (fusion of GFP to α -factor containing Glu-Ala repeats) a modified GFP coding sequence was amplified by PCR from pGFPuv (Clontech) (Figure 2.4). Construct PHA-E:GFP (fusion of PHA-E signal peptide to GFP) was constructed by the megaprimer method (Figure 2.5). The first round of amplification used PHA-E:PHA-E in pPICZB as template with the 5'*AOX1* vector primer and a primer corresponding to the last six residues of the PHA-E signal peptide and the first six residues of GFP (omitting the methionine initiation codon). The second amplification involved pGFPuv as template, the purified megaprimer obtained above and the GFP C-terminal primer. The parameters used in the two sequential PCRs were the same as described above.

2.5.6 Expression Constructs for GNA-GFP in *Pichia pastoris*

For construct α :GNA-GFP, the *Pst*I/*Xba*I-restricted PCR product, resulting from the amplification of GNA from plasmid pGNAH2 with N-terminal primer, 5' ATT ACT GCA GAC AAT ATT TTG TAC TCC GGT 3', and C-terminal primer, 5' TAA TTC TAG AGT TCC GGT GTG AGT TCC AGT 3', was fused to purified, *Xba*I/*Eco*RI-restricted GFP (from pGFPuv) and *Pst*I/*Eco*RI-restricted pPICZ α B, simultaneously, in a triple ligation reaction (see section 2.3.8). For construct PHA-E:GNA-GFP, the megaprimer PCR method was employed. Here, in a first PCR, the PHA-E native signal sequence was amplified from construct PHA-E:PHA-E to produce the megaprimer (see section 2.5.4) using the same primers and conditions described for construct PHA-E:GNA (see section 2.5.3). In a second PCR, GNA was amplified

Green Fluorescent Protein GFP Gene

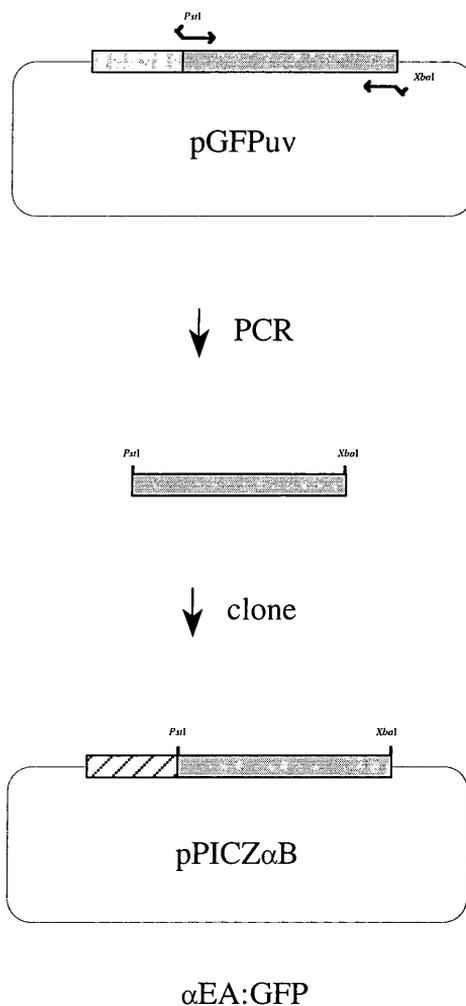


Figure 2.4 Development of expression constructs for GFP. A modified GFP (GFPuv) coding sequence without its leader sequence was amplified by PCR from pGFPuv (Clontech) and cloned into pPICZ α B expression vector. : GFPuv leader sequence; : GFPuv mature coding sequence; : α -factor prepro sequence; \leftarrow : primer (see Table 1.1 for details).

from pGNAH2 using the megaprimer and the C-terminal primer used for the construct α :GNA-GFP. After restriction digest with *EcoRI* and *XbaI*, the product was then fused in a triple ligation reaction to *XbaI/EcoRI*-restricted GFP and *EcoRI*-restricted pPICZB.

2.6 Expression of Recombinant Protein in *E. coli* and *P. pastoris*

2.6.1 Induction of Expression of Recombinant GNA in *Escherichia coli*

The bacterial strain *Escherichia coli* BL21(DE3)pLysS (Studier & Moffatt, 1986) was used for the expression of recombinant GNA. This strain is a lysogen containing a single T7 RNA polymerase gene under control of the *lacUV5* promoter, which is inducible by isopropyl- β -D-thiogalactopyranoside (IPTG). Plasmid pLysS codes for T7 lysozyme, a natural inhibitor of T7 RNA polymerase for minimal background expression levels. The bacteria containing the recombinant plasmid were grown at 37 °C with shaking in 50 ml LB medium containing 100 μ g/ml ampicillin until mid-log phase ($OD_{600nm} = 0.6$), then induced by the addition of IPTG to a final concentration of 1 mM and by continuing the incubation at 37 °C for another 3 hours. Aliquots (1 ml) of bacterial culture were removed after pre-determined periods of induction. The cells were collected by centrifugation and suspended in 50 μ l H₂O and 50 μ l of 2 x SDS-PAGE sample buffer (0.2 M Tris-HCl, pH 6.8, 2 % sodium dodecyl sulphate, 20 % glycerol and 0.002 % bromophenol blue). After SDS-PAGE, detection of the proteins in the bacterial lysates was done by staining with Coomassie Brilliant Blue R250, and the presence of recombinant GNA was confirmed by immunoblotting (see section 2.4.2).

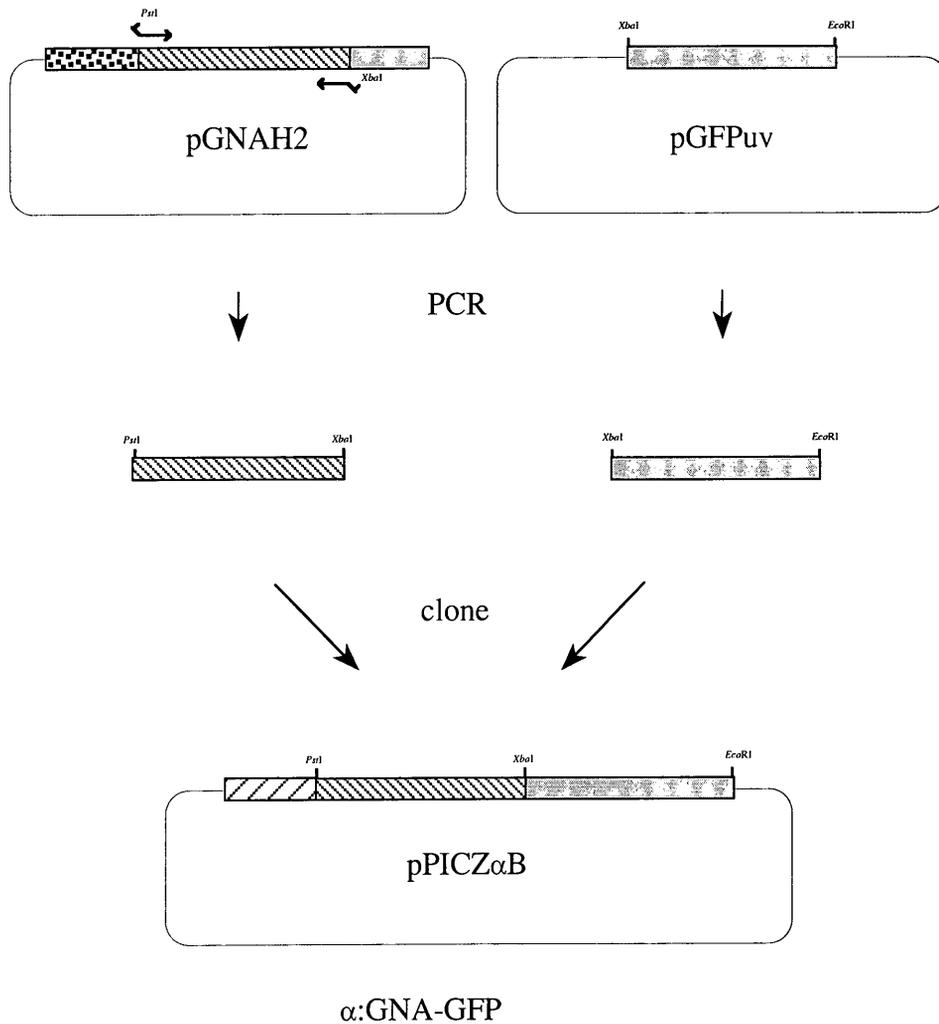


Figure 2.6 Development of expression constructs for GNA-GFP. The GNA and GFP coding sequences were amplified by PCR from pGNAH2 and pGFPuv, respectively, and cloned into pPICZ α B expression vector. : GNA leader sequence; : GNA mature coding sequence; : GNA C-terminal extension sequence; : GFPuv mature coding sequence; : α -factor prepro sequence; : primer (see Section 2.5.6 for details).

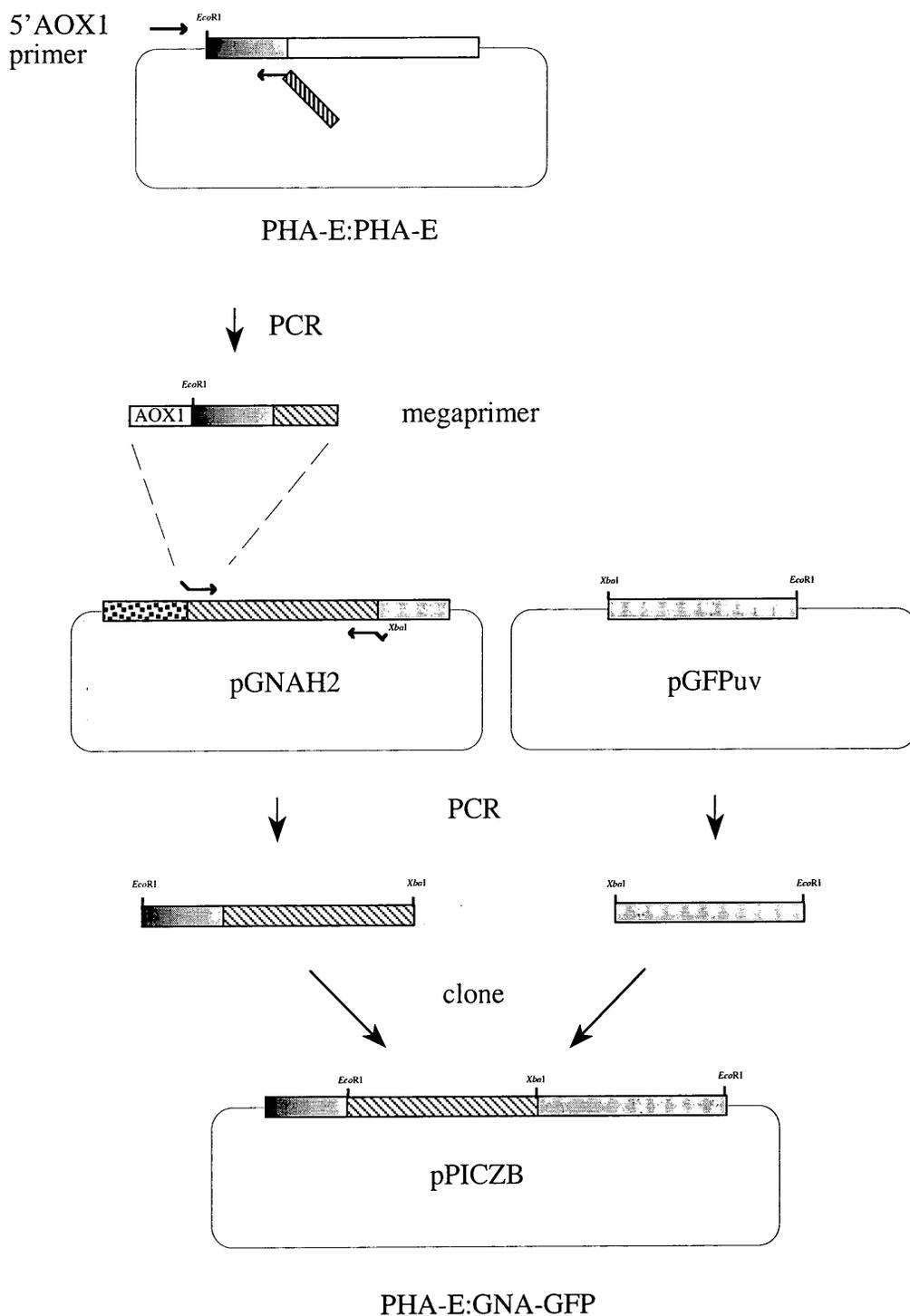


Figure 2.7 Development of expression constructs for GNA-GFP *continued*. The PHA-E leader sequence was amplified by PCR from the PHA-E:PHA-E construct. The PCR product was subsequently used as a (mega)primer for the amplification of the mature GNA coding sequence from pGNAH2. The GFPuv coding sequence was amplified by PCR from pGFPuv. The PCR products were cloned into pPICZB expression vector. : GNA leader sequence; : GNA mature coding sequence; : GNA C-terminal extension sequence; : GFPuv mature coding sequence; : PHA-E leader sequence; : PHA-E mature coding sequence; : primer (see Section 2.5.6 for details).

To obtain larger amounts of recombinant GNA, 2 l LB medium was inoculated with a 20 ml culture and incubated at 37 °C with shaking until the OD_{600nm} reached 0.6. This culture was then induced with IPTG and further incubated overnight. Cells were collected by centrifugation and suspended in 200 ml of TE buffer (50 mM Tris-HCl, 2 mM EDTA, pH 8.0) containing 20 µg/ml DNase I. Lysis of the bacterial cells was performed on a cell disruption apparatus (Constant Systems Ltd., UK) at 25000 psi at 4 °C and repeated twice. The insoluble material was then collected by centrifugation at 18000 g at 4 °C for 25 minutes and stored at -20 °C until used for solubilization. The inclusion bodies obtained from the 2 l bacterial culture containing recombinant GNA were solubilized in 100 ml 6M urea and 0.1% β-mercaptoethanol and subsequently heated in boiling water for 10 minutes. The denatured solution was then clarified by centrifugation at 18000 g for 30 minutes. Refolding was done by dialysing against 4 l of 50 mM Tris-HCl pH 8.0 at room temperature for at least 1.5 hours. This last step was repeated before storage of the lectin solution at 4 °C.

2.6.2 Transformation of *Pichia*

Transformation of *P. pastoris* was based on a modified version of the procedure described for *Saccharomyces cerevisiae* (Gietz & Schiestl, 1995). Briefly, 50 ml of an overnight culture of *P. pastoris* in rich medium (YPD; 1 % yeast extract, 2 % peptone, 2 % glucose) was centrifuged at 1500 g at room temperature for 10 minutes. After washing the cells once with sterile distilled water, the cells were resuspended in 1 ml of 100 mM LiCl (in contrast to *S. cerevisiae*, lithium acetate does not work for *P. pastoris*), harvested by centrifugation at 13000 g for 15 seconds, resuspended in 400 µl of 100 mM LiCl and then dispensed in 50 µl aliquots in 1.5 ml microcentrifuge

tubes. Per transformation, an aliquot was centrifuged and to the pelleted cells, 240 μ l of 50 % PEG-3350, 36 μ l of 1 M LiCl, 25 μ l of 2 mg/ml single stranded salmon sperm carrier DNA and 30 μ g of *Bst*X I-restricted transformation plasmid in 50 μ l sterile water were added in this order followed by vigorous vortexing. After incubating the mixture at 30 °C for 30 minutes, 35 μ l of dimethylsulfoxide was added. The mixture was heat shocked in a water bath at 42 °C for 25 minutes. The cells were then harvested by centrifugation at 6000 g and resuspended in 1 ml of YPD. The cell suspension was incubated at 30 °C for 3 - 4 hours. Selection of transformants was done by spreading 200 and 800 μ l aliquots on YPD plates (YPD, 1.5% agar) containing 100 μ g/ml Zeocin and incubating at 30 °C for 2 - 3 days. Transformed colonies were transferred to fresh selection medium.

2.6.3 Induction of Expression of Recombinant Proteins in *Pichia pastoris*

Single colonies of transformed *Pichia* were grown overnight in 10 ml BMGY medium (1 % yeast extract, 2 % peptone, 100 mM potassium phosphate, pH 6.0, 1.34 % yeast nitrogen base (Invitrogen), 1 % glycerol, 4 x 10⁻⁵ % biotin, 0.004 % L-histidine), centrifuged, resuspended in 1.0 ml sterile distilled water and inoculated into 50 ml of fresh BMMY medium (BMGY, but glycerol was replaced by methanol). Cultures were shaken at 300 rpm and fresh methanol was added daily for 6 - 7 days to 0.5 % (v/v) at 9 am and 6 pm during the course of induction. In scaling up the production of recombinant protein the same procedure was used as described above except that an 50 ml BMGY culture was used to inoculate 500 ml of BMMY medium in a 2 l baffled flask.

2.6.4 Preparation of Protein Extracts from *Pichia* Cells

The contents of cells harvested after growth and induction of expression were analysed for the presence of recombinant protein. The method of cell disruption was based on mechanical breakage of cells by vortexing in the presence of glass beads and was based on a modified procedure described by Ausubel *et al.* (1993). Briefly, transformed *Pichia* cells, harvested from a 50 ml culture, were suspended in 500 μ l ice-chilled disruption buffer (20 mM Tris-HCl, pH 7.5, 0.1 mM EDTA, 10 % glycerol, 100 mM KCl, 1 mM dithiothreitol, 0.1 μ g/ml chymostatin, 1.0 μ g/ml pepstatin A, 7.2 μ g/ml E-64, 0.5 μ g/ml leupeptin, 1 mM phenylmethylsulfonylfluoride). An equal volume of chilled, acid-washed glass beads (0.45 - 0.55 mm) was added to the suspension. The cell suspension and beads were vortexed vigorously at 4 °C until the amount of cell breakage, visually checked under a microscope, represented more than 75 % of the original cell population. After centrifugation, the supernatant was stored at -20 °C until further required for immunoblotting analysis.

2.7 Purification of Recombinant Lectin

2.7.1 Purification of Recombinant Wild-Type GNA from *E. coli* by Affinity

Chromatography

Solid $(\text{NH}_4)_2\text{SO}_4$ was added to 10 ml of refolded protein solution to a final concentration of 1 M. The pH of the solution was lowered to 4.5 by the addition of 1 M HCl. The precipitate formed was removed by centrifugation and the supernatant loaded onto a column (1.6 cm diameter) containing 10 ml of D-mannose immobilized onto 4 % beaded agarose (Sigma). The column was washed with 1 M $(\text{NH}_4)_2\text{SO}_4$ to

remove any unbound proteins until the absorbance reached background level. The bound fraction was subsequently eluted isocratically with 20 mM diaminopropane (DAP) at a flow rate of 0.5 ml/min. Proteins were monitored by online measurement of absorbance at 280 nm. Fractions of 1.5 ml were collected and analysed by SDS-PAGE. Those containing recombinant GNA, as deduced from SDS-PAGE, were pooled and analysed spectrophotometrically.

2.7.2 Purification of Recombinant Wild-Type and Altered GNA from *E. coli* by Anion-Exchange Chromatography

The denatured/refolded recombinant GNA was dialysed overnight against 20 mM diaminopropane-HCl (DAP-HCl) buffer pH 9.0 and then passed over an anion-exchange chromatography column (30 ml, Q-Sepharose Fast Flow; Pharmacia) equilibrated in 20 mM DAP-HCl pH 9.0. The bound fraction was eluted with a 180 ml linear gradient of 0 - 1 M NaCl in DAP-HCl. Fractions containing the desired recombinant lectin were identified by immunoblotting (see section 2.4.2).

2.7.3 Affinity-Purification of PHA and GNA from *Pichia* Culture Supernatant.

Recombinant PHA isoforms expressed in *Pichia* were affinity-purified on thyroglobulin-agarose essentially as described (Osborn *et al.*, 1984). Briefly, proteins from *Pichia* culture supernatants (50 ml) were precipitated by the addition of solid $(\text{NH}_4)_2\text{SO}_4$ to 80 % saturation and incubated overnight at 4 °C. Precipitates were collected by centrifugation at 3000 g for 40 minutes, resuspended in 2 ml PBS and dialysed against 4 l PBS overnight at 4 °C. Dialysed samples were loaded onto a thyroglobulin-agarose column (5 ml) followed by washing in PBS until all non-bound material had been removed. The column was then washed with non-buffered saline (1

column volume), and eluted in 0.5 M NaCl, 50 mM glycine (pH 3.0). The pH of eluted samples was brought to 7.4 by the addition of Tris-HCl to 0.1 M and samples were dialysed against PBS overnight at 4 °C.

Recombinant GNA was purified from *Pichia* cultures (50 ml) by binding to a mannan-agarose (Sigma) affinity column. Briefly, proteins from culture supernatants were precipitated with $(\text{NH}_4)_2\text{SO}_4$ as described for the PHA purification. After resuspending precipitates in PBS, the soluble fraction was loaded directly onto a column (10 ml) containing mannan-agarose (Sigma). The column was washed with PBS to remove any unbound proteins until the absorbance reached background level. The bound fraction was subsequently eluted isocratically with 20 mM DAP and eluted proteins were monitored by online measurement of absorbance at 280 nm. Fractions containing recombinant GNA were pooled, dialysed against NH_4HCO_3 , lyophilised and dissolved in PBS.

Routinely, the ammonium sulphate method was not used. An alternative and more preferred method for the affinity purification of recombinant GNA involved loading the culture supernatant directly onto the column. This step eliminated the inefficient redissolving of the pellet after ammonium sulphate precipitation. The column was washed with PBS and the bound fraction eluted with DAP as described above.

2.8 Recombinant Protein Functional Assays

2.8.1 Plant Lectins and Green Fluorescent Protein

Native snowdrop lectin (GNA) was obtained from the laboratory of Professor W. Peumans of the Catholic University of Leuven, Belgium, and from Vector

Laboratories (Peterborough, Cambridgeshire, UK). Native PHA-E and PHA-L were purchased from Calbiochem (Novabiochem, Nottingham, UK).

Standard *E. coli*-expressed recombinant green fluorescent protein (rGFP) was obtained from Clontech (Basingstoke, Hampshire, UK).

2.8.2 Hemagglutination Assays

Hemagglutination assays were carried out in roundbottomed microtitre plates. A total volume of 100 μ l was used in each well: 50 μ l aliquots of serial twofold dilutions of the lectin in PBS and 50 μ l of a 2 % rabbit erythrocyte suspension in PBS. Agglutination assays were incubated at room temperature for 1 hour. The lowest concentration of lectin required to completely agglutinate the red blood cells was determined visually. The effects of mannose, methyl- α -D-mannoside and glucose on the hemagglutination by the lectin were studied by the addition of serial dilutions of these saccharides to the hemagglutination assay mixture containing the lectin at a final concentration of 15 μ g/ml. The inhibitory activity was expressed as the minimum concentration of the hapten that completely inhibited the activity of the lectin. All hemagglutination assays were carried out in duplicate.

2.8.3 GNA-Carboxypeptidase-Y-Binding Assay

Each well of an immunoplate (Immulon 4, Dynatech) was coated with 50 μ l of carboxypeptidase Y (2 μ g/ml in PBS) at room temperature for 2 hours with gentle shaking. The solution was discarded and the wells washed two times with PBST (PBS containing 0.01 % Tween20). The wells were then completely filled with 5 % (w/v) non-fat dry milk containing 0.1 % Tween20 in PBS and the plate was incubated at

room temperature for 2 hours with gentle shaking. The wells of the plate were washed three times with PBST. The GNA test sample (final concentration 20 $\mu\text{g/ml}$) with serial twofold dilutions of methyl- α -D-mannoside (initial final concentration 500 mM) were then added to a final volume of 50 μl . After incubation at room temperature for 1 hour, each well was emptied and washed three times, then 50 μl of antiGNA polyclonal antibodies (1:10,000 dilution) was added. The reaction was allowed to proceed at room temperature for 1 hour with gentle shaking. The plate was washed three times and 50 μl of horseradish peroxidase-labelled goat anti-rabbit antibodies (1:10,000 dilution) was added to each well. The reaction proceeded at room temperature for 1 hour with gentle shaking, then the wells were washed three times with PBST. Fifty μl of an equal volume of ABTS Peroxidase Substrate and Peroxidase Substrate Solution B (Kirkegaard and Perry Laboratories Inc.) was added to each well and incubation proceeded at room temperature for approximately ten minutes. Optical density measurements were taken with a Dynatech MR5000 microplate reader (Dynatech Laboratories Ltd., West Sussex, UK) at a wavelength of 405 nm. For each optical density reading, the negative control value with PBS in place of lectin solution was subtracted. All readings were taken in duplicate.

2.8.4 Direct ELISA of GNA Variants

Wells of a microtitre plate were coated with 50 μl of a serial twofold dilution of lectin (initial concentration of 10 $\mu\text{g/ml}$). Procedure for the detection of native and recombinant GNAs by anti-GNA antibodies was followed as described in section 2.8.3.

2.8.5 Probing of BPH Total Gut Protein Extract with Recombinant GNA and Detection by Immunoblotting

A stock culture of *Nilaparvata lugens* (rice brown planthopper, BPH) was reared as previously described (Powell *et al.*, 1993). Fifty brown planthopper adult female guts were isolated with the aid of a light microscope and dissolved in 100 μ l of 1 x SDS sample buffer. After incubation with shaking for 30 minutes the insoluble material was removed by centrifugation. The sample was treated with β -mercaptoethanol (5 %) and heated in a boiling water bath for 10 minutes. Ten μ l of the total gut extract was then loaded onto an SDS-PAGE (12.5 %) gel. CarboxypeptidaseY (3 μ g) was used as positive control. The proteins were blotted onto PVDF membrane (Boehringer-Mannheim) and the membrane blocked for 30 minutes at room temperature in 0.5 % (w/v) blocking reagent (Boehringer-Mannheim) in TBS (0.05 M TrisHCl, 0.15 M NaCl, pH 7.5). The membrane was washed 2 x in TBS for 10 minutes each, equilibrated in lectin probe buffer (TBS containing 1 mM MgCl₂, 1 mM MnCl₂, 1 mM CaCl₂, pH 7.5) for 5 minutes, and incubated with rGNA105 or rGNA105(Y97F) in 20 ml probe buffer at a concentration of 1 μ g/ml for 1 hour at room temperature. After two washes in TBS for 10 minutes each, the membrane was incubated in PBS blocking buffer containing 5 % nonfat milk powder and 0.1 % Tween20 for 1 hour at room temperature. Bound lectins were then detected by treatment with anti-lectin antibodies followed by ECL detection as described in section 2.4.2.

2.8.6 BPH Bioassays

Bioassays of BPH using artificial diets containing 0.05 % (w/v) GNA were carried out as previously described (Powell *et al.*, 1993). The corrected mortality values (Abbot, 1925) were based on the time taken for complete mortality of planthoppers exposed to a “no diet” treatment (6 days after start of trial) and was calculated as follows: $[(a-b)/a] \times 100 \%$, where a = number of surviving insects fed on diet only and b = number of insects fed on diet plus lectin. Survival curves for BPH in different treatments were compared by Survival Analysis using the Statview software package (Abacus Concepts Ltd., California, USA) on Apple Macintosh computers. Differences were assessed for significance using a logrank test.

2.8.7 Extraction of Haemolymph from *Lacanobia oleracea*

Lacanobia oleracea larvae were reared and maintained on potato leaf based artificial diet, as previously described (Fitches & Gatehouse, 1998). The larvae were chilled on ice for 10 minutes, the cuticle was swabbed with ethanol, dried and pierced with a sterile 21 gauge needle. Droplets of extruded haemolymph was collected using a sterile glass microcapillary tube and placed into a pre-chilled 0.5 ml eppendorfs dusted with phenylthiocarbamide-phenol oxidase inhibitor (PPO). Fifteen to 50 μ l haemolymph was extractable using this technique.

2.8.8 Preparation of Haemocyte Monolayers

The haemolymph from 6th instar *Lacanobia oleracea* larvae was diluted 1 in 8 in ice cold Tris buffered saline (TBS: 0.05 M Tris-HCl, pH 7.4, NaCl to osmolarity 309 mosmol/kg) in an Eppendorf tube and then applied to 8 mm diameter wells on a

microscopic slide (eight well slides, Hendly, Essex, UK). All incubations were performed in a moist chamber at room temperature. After 20 minutes, the monolayers were washed gently six times with TBS using a Pasteur pipette to remove plasma components. The haemocytes were then incubated with TBS for 10 minutes to attain a rounded configuration. The cells were overlaid with 50 μ l of TBS alone or containing 5 μ g/ml of recombinant GNA-GFP or GFP for 1 hour, washed six times with TBS and viewed by confocal microscopy.

2.8.9 Fluorescence Microscopy

A Nikon episcopic fluorescence microscope (OPTIPHOT-2) was used to detect GFP fluorescence in a suspension of *Pichia* cells or mannan-agarose beads. Cells taken directly from the culture were immobilised on a microscopic slide in 0.5 % (w/v) agarose. The light source was powered by a Nikon Super High Pressure Mercury Lamp Power Supply HB-10101 AF. The filter system used was as follows: a BV-2A filter (excitation wavelength 400 - 440 nm), a dichroic mirror DM 455 and a barrier filter BA 470. Exposure time of fluorescence to a Fujichrome (ASA 400) diafilm in a Nikon FX-35 camera was controlled by a Nikon AFX Photomicrographic attachment.

2.8.10 Confocal Microscopy

Haemocytes were viewed under epifluorescent illumination (excitation at 488 nm; emission filter HQ 518/40) with a Nikon Diaphot microscope using a BioRad MicroRadiance Confocal (MRC-1024) Scanning System. Images were captured on computer, and assembled using the manufacturer's LaserSharp imaging software.

2.8.11 Fluorometric Assays

Recombinant GFP fusion with GNA was used in fluorometric assays to confirm and measure GFP fluorescence. Fluorescence was measured using a Fluoroskan Ascent microtitre plate fluorimeter (Labsystems, Life Sciences International Ltd., Basingstoke, Hampshire, UK). An excitation filter of 355 nm and an emission filter of 538 nm were used. Quantitation of recombinant GNA-GFP levels was made possible by comparing the fluorescence intensity of the affinity-purified sample to that of a serial two-fold dilution of standard recombinant GFP_{uv}.



Chapter 3

Expression of Functional GNA in *E. coli*

3.1 Introduction

As described earlier, plant lectins have been proposed to play a role in the defence of plants against insect pests, and a number of plant lectins have been shown to exhibit antimetabolic or insecticidal effects on insects when fed in artificial diets. One of these lectins, the snowdrop lectin (GNA), has the added advantage of being non-toxic towards mammals (Pusztai *et al.*, 1990). Much interest has, therefore, been directed on utilizing GNA as a tool to promote insect resistance of crop plants.

The GNA gene used for plant transformation was derived from LECGNA2 cDNA, one of six isolectin clones obtained from a cDNA library constructed from total poly(A)+RNA isolated from snowdrop ovaries (Van Damme *et al.*, 1991b). Since GNA exists as a mixture of isoforms in its natural state, it is important to ensure that the isoform selected for transfer and expression in a genetically engineered plant is one which imparts the anticipated antimetabolic activity.

In this chapter, the expression, production, isolation and purification of recombinant GNA in *E. coli* using a coding sequence derived from LECGNA2 is described. Although the snowdrop lectin has previously been expressed in bacteria with several extra C-terminal His residues for purification purposes (Longstaff *et al.*, 1998), only the mature part of GNA was chosen to be expressed and a different purification strategy, that based on its inherent carbohydrate binding property, was employed.

There are conflicting reports in the literature concerning the actual length of the mature polypeptide. Van Damme *et al.* (1991b), using protein sequencing, determined the sequence of GNA isolated from the snowdrop bulbs as comprising of 105 amino acids. However, on the basis of X-ray structure determination on GNA also isolated from the snowdrop bulbs, Hester *et al.* (1995) suggested that the GNA polypeptide contains four extra C-terminal amino acid residues. To clarify the importance of this difference, both types of recombinant GNAs were produced in the current study. Various functional assays were employed to demonstrate the similarities and differences in activities of recombinantly expressed wild-type GNAs when compared to that of native GNA.

A correlation between binding of lectins to the gut, and toxicity towards insects, has been assumed since the initial reports of lectin toxicity. However, binding in itself is not sufficient to cause toxicity, since screening a number of different lectins for binding to insect guts showed that not all lectins which bound were toxic (Harper *et al.*, 1995). A dependence of toxicity on binding has recently been demonstrated for the lectin isolated from the leaves of *Griffonia simplicifolia* towards the cowpea bruchid *Callosobruchus maculatus*, since mutation of the sugar binding sites of the lectin to abolish binding resulted in loss of insecticidal activity (Zhu *et al.*, 1996; Zhu-Salzman *et al.*, 1998).

As part of a programme to elucidate the mode of action of GNA as a toxin towards the rice brown planthopper (*Nilaparvata lugens*), the structure-activity relationship of the lectin was analysed by site-directed mutagenesis of the high-affinity mannose-binding site (site I). Based on the crystal structure of GNA (Hester *et al.*, 1995), the conserved tyrosine (residue 97) in binding site I was altered to phenylalanine, thus

removing its potential to form a hydrogen bond with sugar residues in the binding site. Also, aspartic acid and asparagine (residues 91 and 93, respectively) were substituted by leucine. The effect of these changes on the activity of the lectin was studied using independent assay systems *in vitro*, as well as in insect bioassays.

3.2 Results

3.2.1 Design of GNA Constructs

Two types of mutations were designed. In the first type, GNA was produced with or without the four extra C-terminal residues predicted by the X-ray structure, but absent from the published sequence for the protein (rGNA109 and rGNA105, respectively; Fig. 3.1). The rGNA109 variant would show whether the extra four amino acids had any effect on the functional properties of the protein.

The second type mutations in GNA were designed to investigate the relationship between the carbohydrate-binding activity of GNA and its biological activity. The conserved residues of one of the mannose-binding sites, site I, of each subunit interact with methyl- α -D-mannoside by hydrogen bonding between the OH-groups of the sugar and the OH- or NH₂-groups of the amino acid residue side chains and also by hydrophobic contacts with the carbon ring atoms of the saccharide (Hester *et al.*, 1995); Fig. 3.2). The hydroxyl group of tyrosine at position 97 makes a hydrogen bond interaction with the OH-group of 4C of the saccharide. To abolish this specific interaction, tyrosine-97 was replaced by phenylalanine, which lacks the OH-group and thus cannot form an H-bond, to produce the mutant rGNA105(Y97F). Both aspartic acid and asparagine at positions 91 and 93, respectively, interact with the C2-OH group of methyl- α -D-mannoside by hydrogen bond formation. Leucine was

chosen as candidate to replace these two residues individually or simultaneously to give constructs rGNA105(D91L), rGNA105(N93L) and rGNA105(D91L/N93L). The rationale for choosing leucine was its hydrophobicity and similarity in molecular mass to their respective substituted amino acid residues so as to ensure minimal structural constraints within the GNA molecule, but at the same time not permitting hydrogen bond formation with the sugar.

3.2.2 Expression of Recombinant GNA

The recombinant vectors pGNA105 and pGNA109 were based on pET-11d, one of several pET-based expression vectors commercially available (Novagen). The pET vector is under control of strong T7 bacteriophage transcription signals when introduced in *E. coli* host strain BL21(DE3)pLysS; expression of the system is repressed unless induced by IPTG. A time course analysis of expression of rGNA105 from pGNA105 transformed into this host strain was carried out. SDS-PAGE of total bacterial protein extract revealed a prominent band migrating slightly faster than that of the native GNA monomer, which was apparent within 1 hour of IPTG induction (Fig. 3.3 (A)). Immunoblotting with polyclonal rabbit anti-nGNA antiserum confirmed the presence of recombinant GNA (Fig. 3.3 (B)). Background binding (less specificity of the primary antibody) did occur but this was similar for all samples. No basal expression (i.e. under non-inducing conditions) of rGNA105 in BL21(DE3)pLysS strain was observed. A strict control on expression was preferred so as to allow, with a degree of certainty, the bacterial culture to reach the desired optical density prior to induction. Leaky expression of rGNA105 was observed when a less stringent *E. coli* strain, namely BL21(DE3), was used. However, no difference in the amount of recombinant GNA expressed was observed between the two strains.

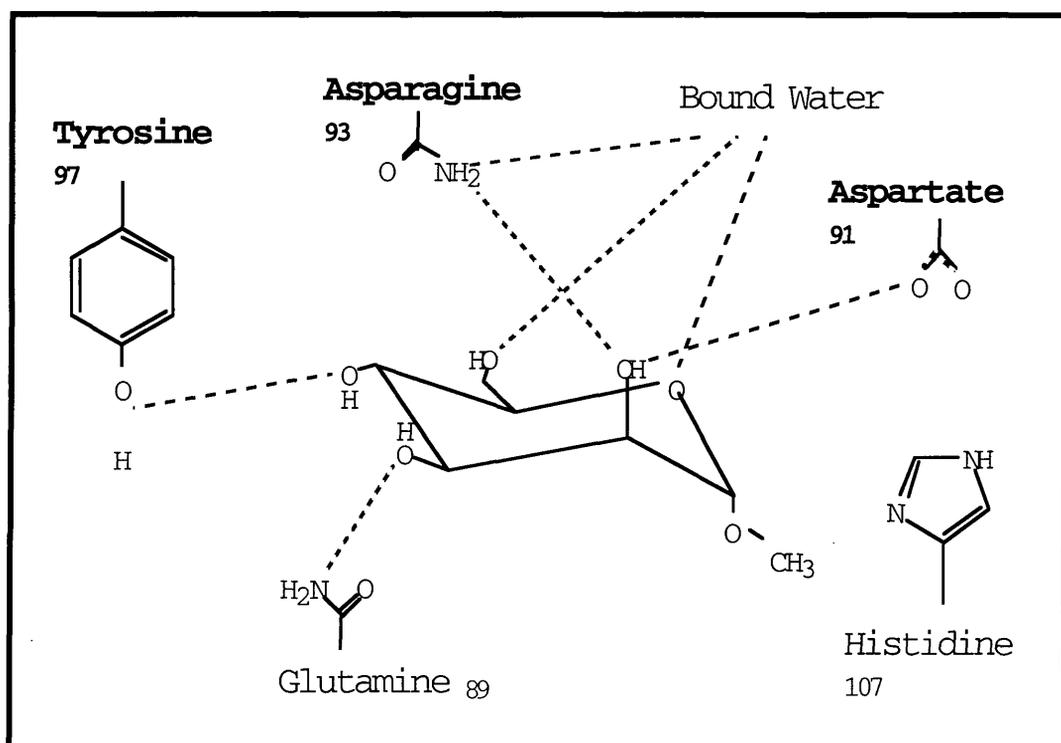


Figure 3.2 Region of high-affinity mannose-binding site I of the snowdrop lectin with bound methyl- α -D-mannoside based on the 3D-structure. Residues substituted in site-directed mutagenesis are shown in bold. Hydrogen bonds are indicated by dashed lines. (Figure modified from Hester *et al.*, 1995.)

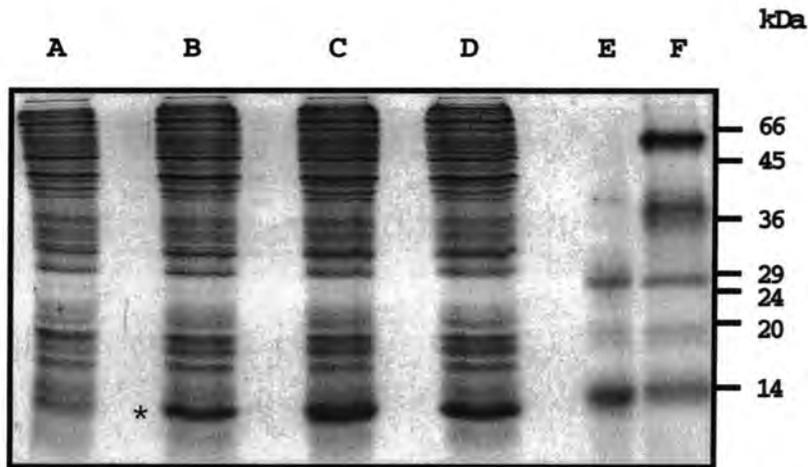


Figure 3.3 (A) SDS-PAGE analysis (12.5 % gel) by Coomassie staining of total (soluble + insoluble) protein extracts of pGNA105 (pET-11d + GNA DNA insert)-containing *E. coli* BL21(DE3)pLysS after induction. Lanes A - D: 0, 1, 2 and 3 hours, respectively, after addition of IPTG; lanes E and F: native GNA. All samples, except native GNA in lane F, were subjected to 10 % 2-mercaptoethanol and heat treatment in boiling water for 10 minutes. (* band represents rGNA105.)

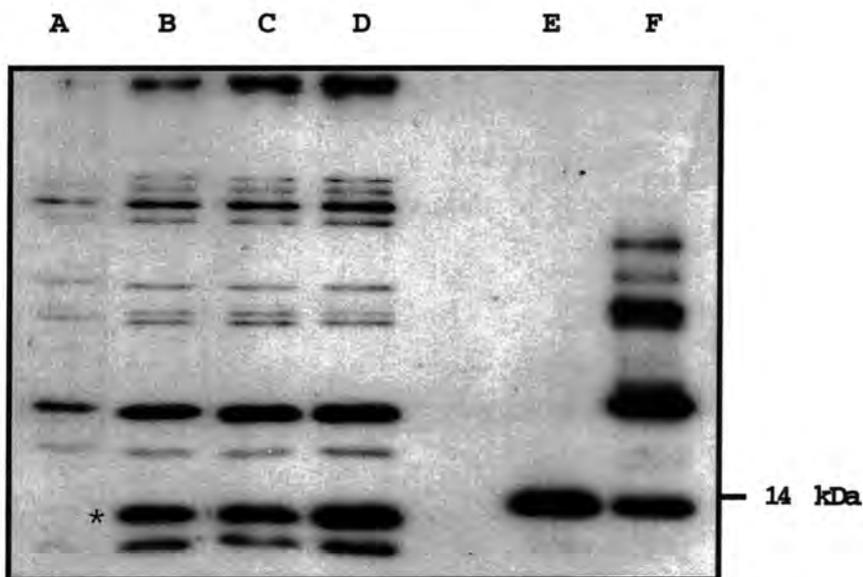


Figure 3.3 (B) Immunoblot analysis using polyclonal anti-nGNA antiserum of total (soluble + insoluble) protein extracts of pGNA105 (pET-11d + GNA DNA insert)-containing *E. coli* BL21(DE3)pLysS after induction. Lanes A - D: 0, 1, 2 and 3 hours, respectively, after addition of IPTG; lanes E and F: native GNA. All samples, except native GNA in lane F, were subjected to 10 % 2-mercaptoethanol and heat treatment in boiling water for 10 minutes. (* band represents rGNA105.)

As indicated by Figure 3.3 (B), subjecting native GNA to prolonged heat treatment (boiling > 10 minutes) in the presence of β -mercaptoethanol was necessary to completely resolve the high molecular weight bands observed in both native and recombinant GNA to the monomeric form. However, as shown in Figure 3.3 (A), if exhaustive denaturation and reduction was not carried out, the high molecular weight bands were not resolved into the monomeric form.

GNA has been expressed previously using the pET-21c expression vector in *E. coli* strain HMS174(DE3) as a His₆-tagged fusion protein (rGNA11; (Longstaff *et al.*, 1998)) and was found to occur predominantly as insoluble inclusion bodies. Recombinant GNA11 was also expressed in this work for comparative purposes. Similar observations were obtained for all the forms of rGNA where almost all the protein was present in the inclusion body fraction. Omission of the His₆-tag thus does not lead to soluble GNA being produced in *E. coli*. A time course on rGNA11 expression revealed a band migrating slower than that of the native GNA monomer, as confirmed by immunoblotting with polyclonal rabbit anti-nGNA antiserum (Fig. 3.4 (A) and (B)). Surprisingly, no “leaky” expression (i.e. low level expression when not induced) was noticeable when the non-stringent strain HMS174(DE3) was used.

3.2.3 Purification of Recombinant GNA

The inclusion bodies were released from the bacteria after cell disruption and collected by centrifugation. To isolate soluble and functional recombinant GNA forms rGNA105, rGNA109, and rGNA11, the aggregates were solubilized by treatment with 6 M urea / 0.1 % β -mercaptoethanol and heating in a boiling water bath for 5 minutes.

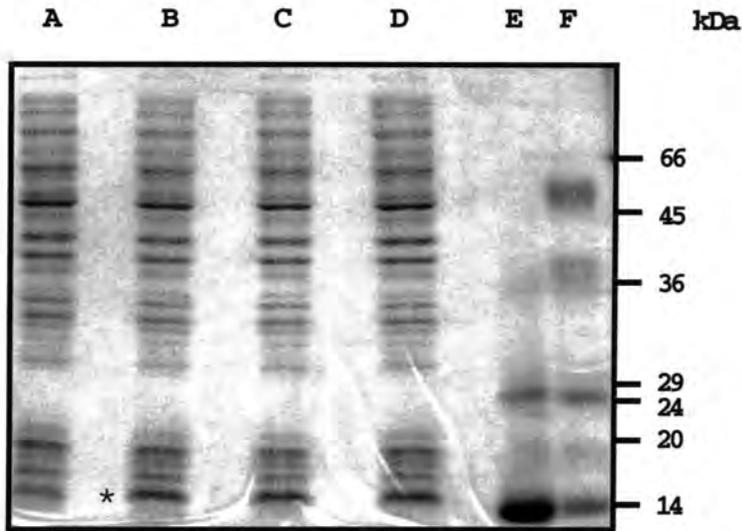


Figure 3.4 (A) SDS-PAGE analysis (12.5 % gel) by Coomassie staining of total (soluble + insoluble) protein extracts of pGNA11 (pET-21c + GNA DNA insert)-containing *E. coli* HMS174(DE3) after induction. Lanes A - D: 0, 1, 2 and 3 hours, respectively, after addition of IPTG; lanes E and F: native GNA. All samples, except native GNA in lane F, were subjected to 10 % 2-mercaptoethanol and heat treatment in boiling water for 10 minutes. (* band represents rGNA11.)

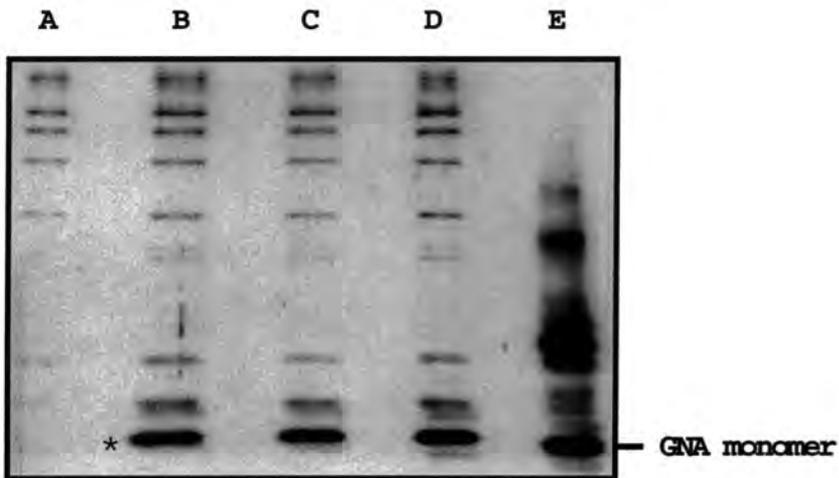


Figure 3.4 (B) Immunoblot analysis using polyclonal anti-nGNA antiserum of total (soluble + insoluble) protein extracts of pGNA11 (pET-21c + GNA DNA insert)-containing *E. coli* HMS174(DE3) after induction. Lanes A - D: 0, 1, 2 and 3 hours, respectively, after addition of IPTG; lane E: native GNA. All samples were subjected to 10 % 2-mercaptoethanol and heat treatment in boiling water for 10 minutes. (* band represents rGNA11.)

The heating was necessary to dissolve further the aggregates. Any remaining insoluble material was removed by centrifugation. Refolding of the recombinant lectin was allowed under very specific circumstances. The denaturant was removed by dialysis under slightly alkaline conditions (pH 8.0) at room temperature. Altering the temperature or the pH of the solution resulted in aggregate formation (results not shown) which decreased the yield substantially. The protein preparations were analysed by SDS-PAGE (Fig. 3.5(A)) which showed that recombinant GNA accounted for at least 65 % of the total protein content of the inclusion bodies. Losses due to protein aggregation during the transition from denaturing to renaturing conditions were kept to a minimum. Western blotting analysis (Fig. 3.5(B)) revealed that for rGNA105 only one band was present, with a molecular weight slightly lower than that for native GNA monomer (cf time course analysis shown above), whereas for rGNA11, two bands were noticeable, one major band with a molecular weight slightly larger than the native monomer and one minor band with an even larger molecular weight but smaller than 20 kDa.

The refolded protein was affinity-purified on a mannose-agarose column to isolate functional lectin. SDS-PAGE analysis of the fractions of rGNA105 collected after washing the affinity column with ammonium sulphate and eluting with 1,3-diaminopropane showed that the eluted fractions contained only rGNA105, whereas no detectable traces of the recombinant lectin were found in the unbound fractions (Fig. 3.6). rGNA109 produced similar results (data not shown). In the case of rGNA11, however, a band (14 kDa < Mw < 20 kDa) co-eluted with the major band (Fig. 3.7). Fractions eluted later after applying the elution buffer were consistently observed to contain more GNA, suggesting that the recombinant GNA was binding

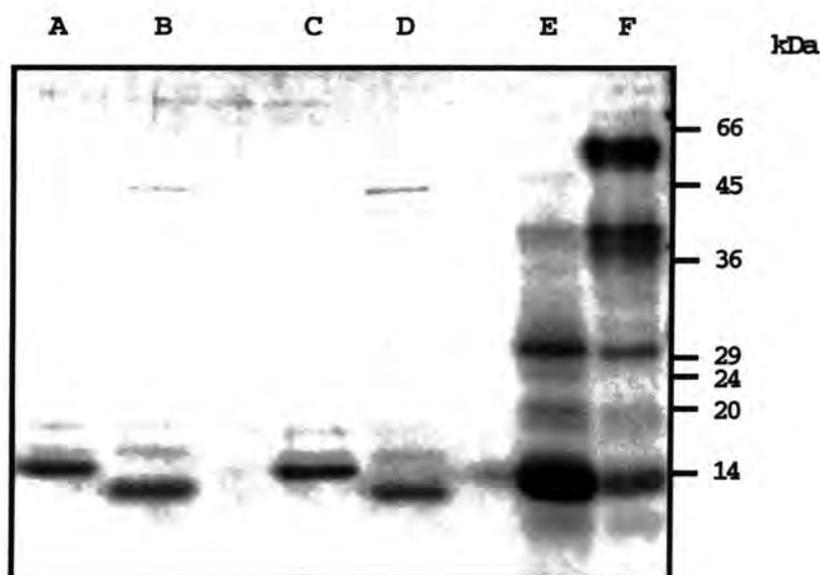


Figure 3.5 (A) SDS-PAGE analysis (15 % gel) by Coomassie staining of denatured/renatured rGNA11 and rGNA105. Solubilisation of insoluble pellets from 10 ml induced culture was carried out in 1 ml 6 M urea / 0.1 % 2-mercaptoethanol; lanes A and B, respectively. Refolding was done by dialysis against 50 mM Tris-HCl pH 8; lanes C and D, respectively. Lanes E and F: native GNA. All samples, except native GNA in lane F, were subjected to 10 % 2-mercaptoethanol and heat treatment in boiling water for 10 minutes.

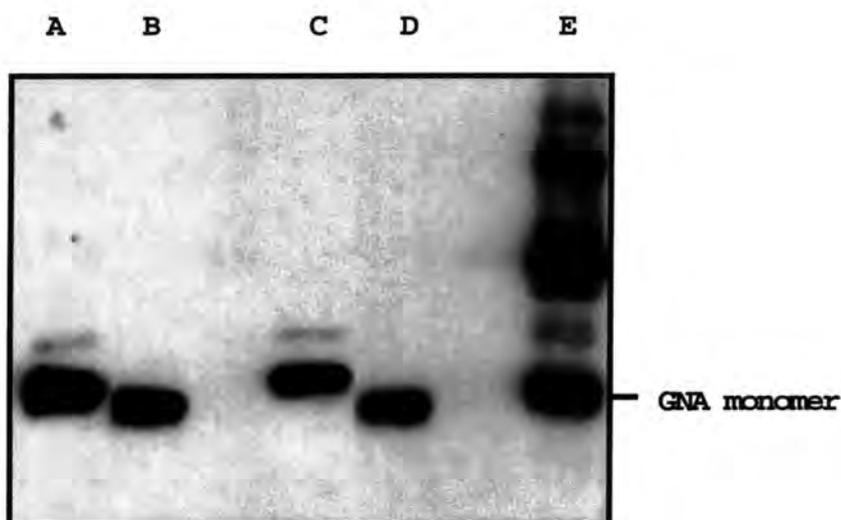


Figure 3.5 (B) Immunoblot analysis using polyclonal anti-nGNA antiserum of denatured/refolded rGNA11 and rGNA105. Solubilisation of insoluble pellets from 10 ml induced culture was carried out in 1 ml 6 M urea / 0.1 % 2-mercaptoethanol; lanes A and B, respectively. Refolding was done by dialysis against 50 mM Tris-HCl pH 8; lanes C and D, respectively. Lane E: native GNA. All samples were subjected to 10 % 2-mercaptoethanol and heat treatment in boiling water for 10 minutes.

strongly to the affinity column. Analysis of the samples of rGNA105 which bound to the mannose column revealed a single band. The yield of purified rGNA105 was 4 - 6 mg per litre culture, representing 10 - 20 % of total recombinant GNA in the inclusion body fraction.

The mutants were expressed as for recombinant GNAs rGNA105 and rGNA109, and were purified from inclusion bodies after denaturation-renaturation as described above. However, since the carbohydrate-binding site of these mutants had been manipulated, binding to the mannose-agarose column was found to be decreased, preventing affinity purification being used. Anion-exchange chromatography was thus employed as an alternative purification method (Fig. 3.8). Although anion-exchange chromatography did not provide an affinity separation, it was possible to purify the recombinant wild-types rGNA105 and rGNA109 and mutants rGNA105(Y97F) and rGNA109(Y97F) from other proteins to a purity of > 85 %, as assessed by SDS-PAGE (Fig. 3.10 and 3.11). Purification of the other mutants rGNA105(D91L), rGNA105(N93L) and rGNA105(D91L/N93L)) using the same method resulted in a much lower yield (i.e. < 50 %) as shown by the Coomassie stained gels (Fig. 3.11). Immunoblotting with anti-nGNA confirmed the presence of the recombinant lectin in all cases. However, the double mutant rGNA105(D91L/N93L) appeared less antigenic since SDS-PAGE gels showed the presence of equal amounts of the protein in question.

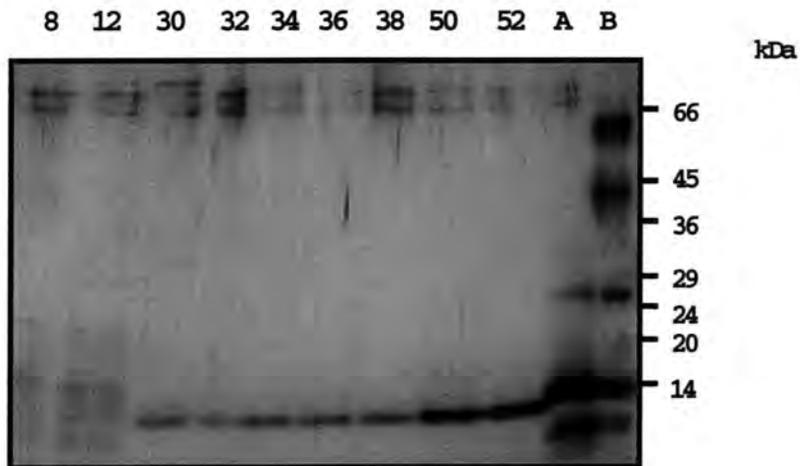
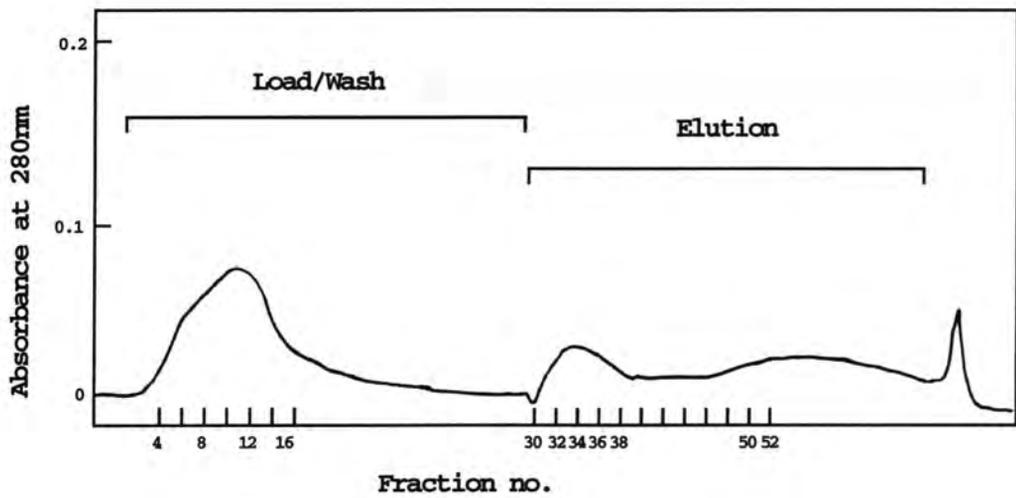


Figure 3.6 Purification of rGNA105 by affinity chromatography on a mannose-agarose column. Elution profile and SDS-PAGE (15 %) analysis by silver staining of unbound and eluted fractions. Lane numbers of gel correspond to fraction numbers of profile. Lanes A and B: native GNA.

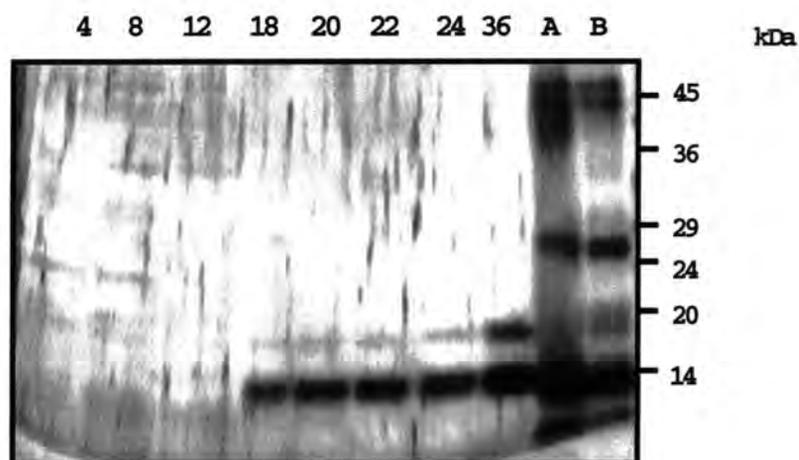
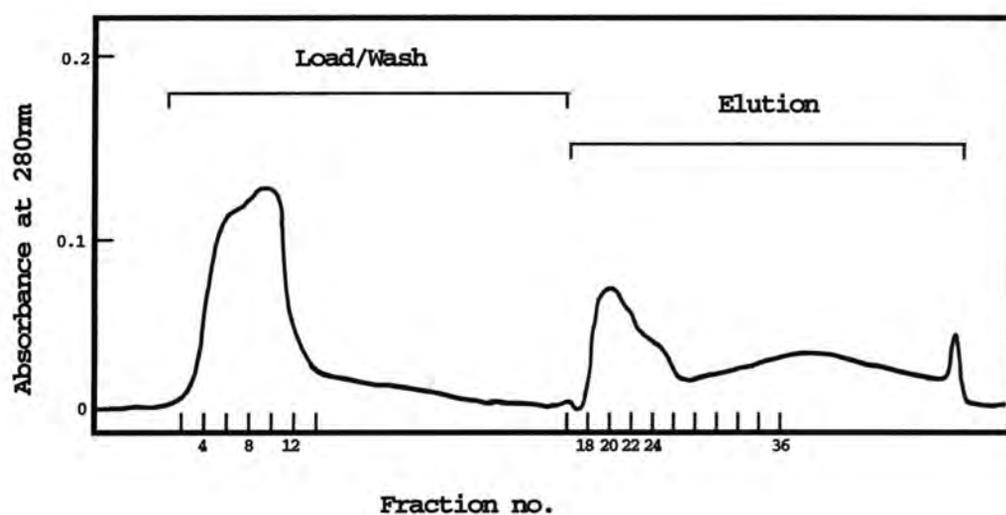


Figure 3.7 Purification of rGNA11 by affinity chromatography on a mannose-agarose column. Elution profile and SDS-PAGE (15 %) analysis by silver staining of unbound and eluted fractions. Lane numbers of gel correspond to fraction numbers of profile. Lanes A and B: native GNA.

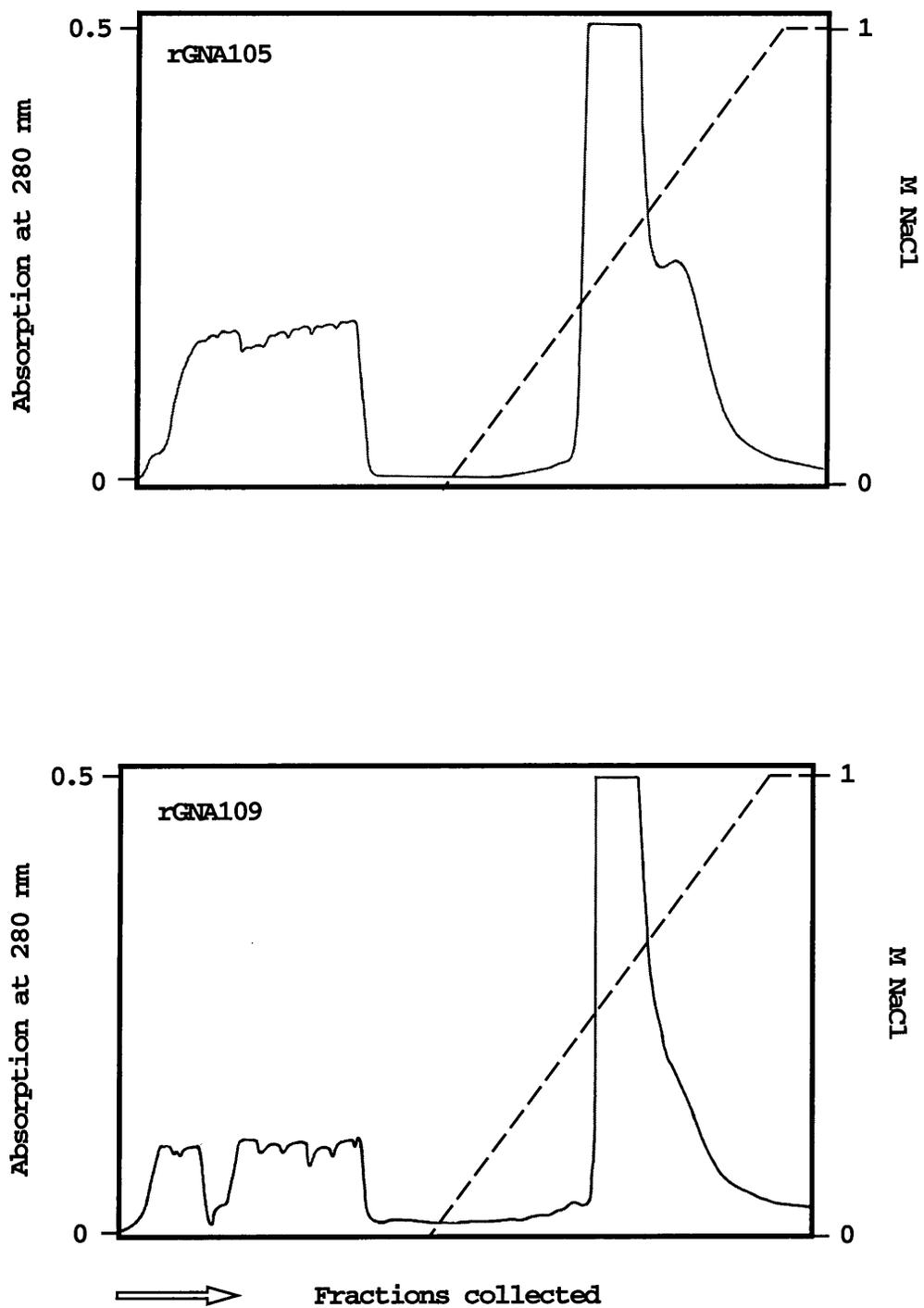


Figure 3.8 Anion-exchange chromatography of rGNA105 and rGNA109.

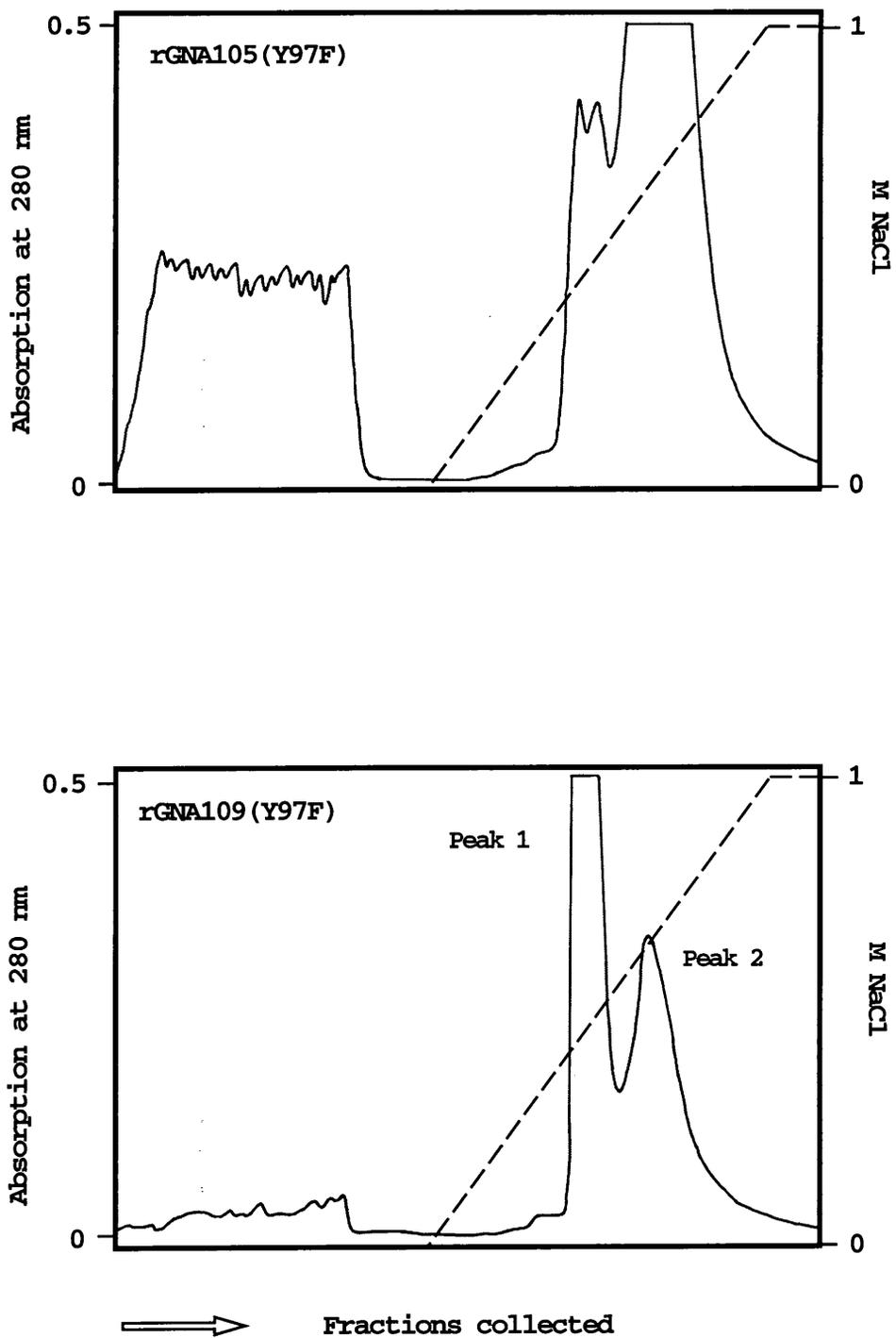


Figure 3.8 continued. Anion-exchange chromatography of rGNA105(Y97F) and rGNA109(Y97F).

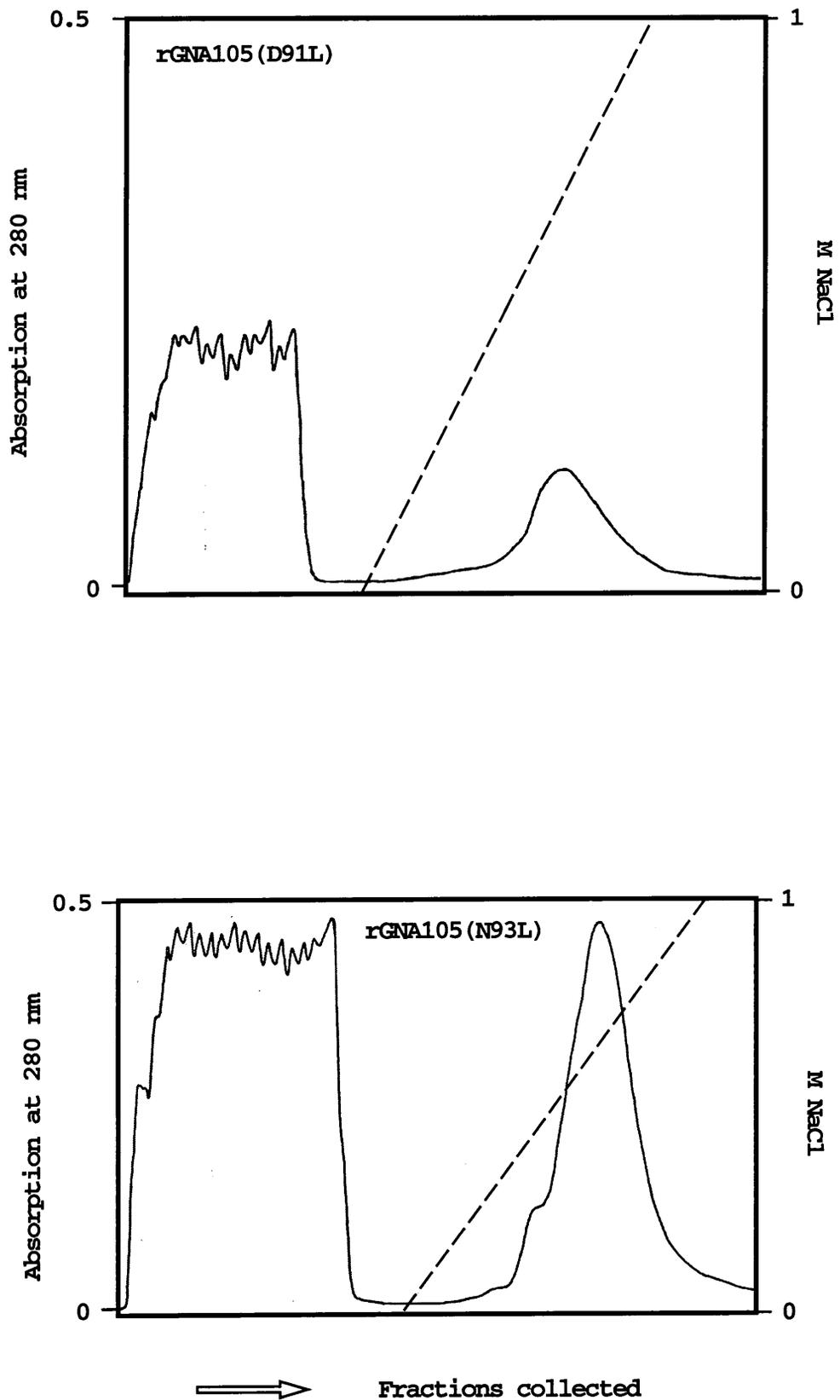


Figure 3.8 *continued*. Anion-exchange chromatography of rGNA105(D91L) and rGNA105(N93L).

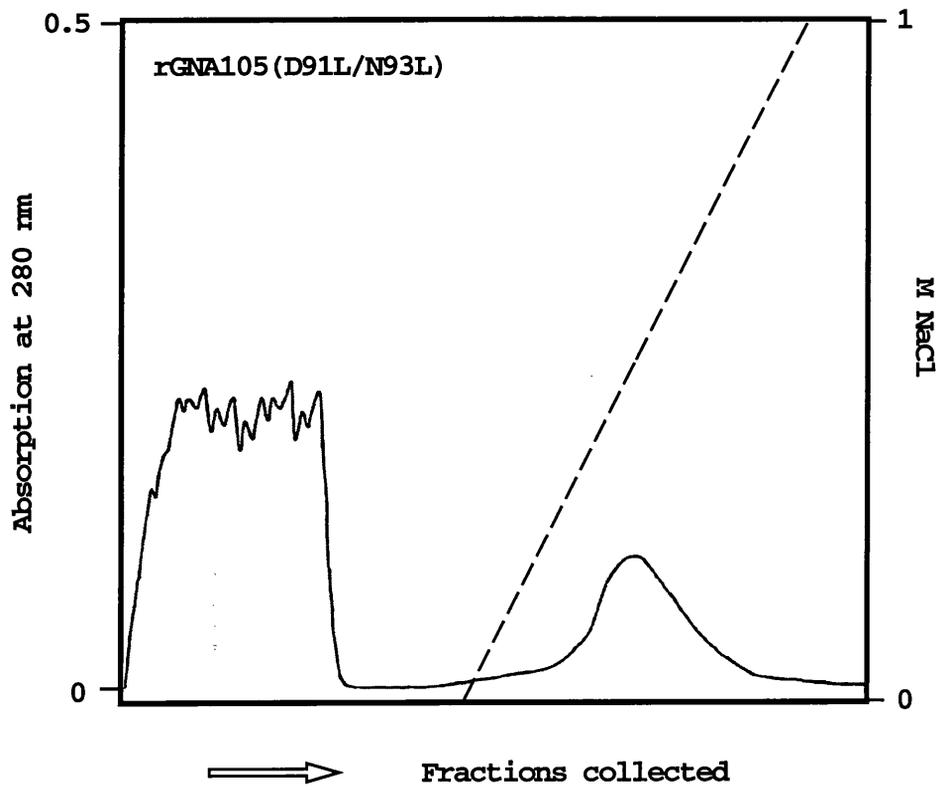


Figure 3.8 continued. Anion-exchange chromatography of rGNA105(D91L/N93L).

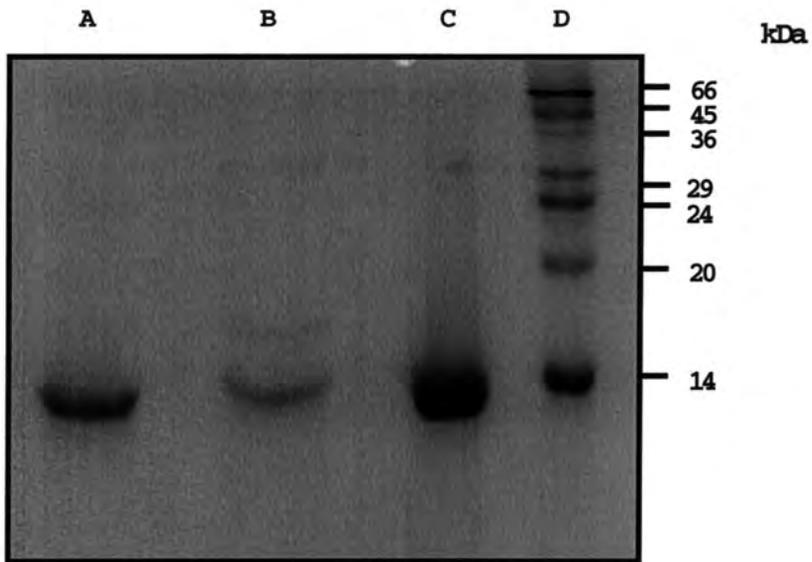


Figure 3.9 SDS-PAGE analysis (12.5 % gel) by Coomassie staining of IEC-purified rGNA109(Y97F). Lane A: Peak 1 fraction; lane B: Peak 2 fraction; lane C: native GNA monomer; lane D: SDS7 molecular weight marker.

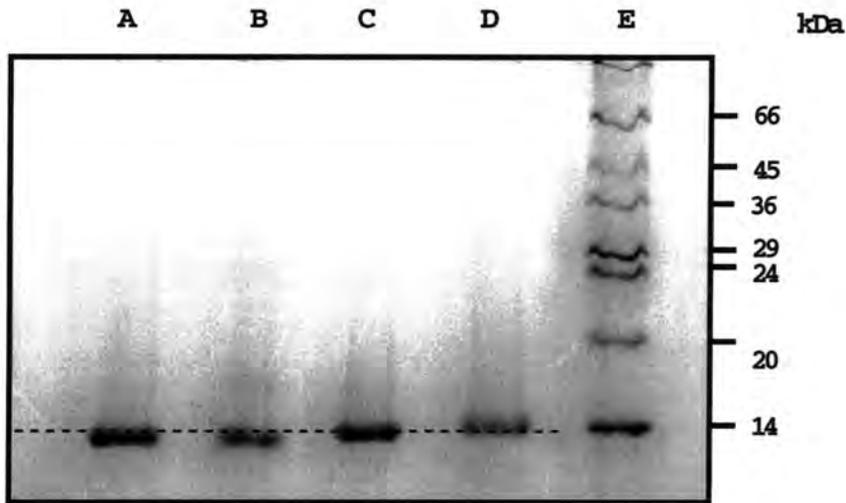


Figure 3.10 (A) SDS-PAGE analysis (15 % gel) by Coomassie staining of purified recombinant wild-type and mutant GNA by anion-exchange chromatography. Lane A: rGNA105; lane B: rGNA105(Y97F); lane C: rGNA109; lane D: native GNA monomer; lane E: SDS7 molecular weight marker. Dashed line indicates differences in migration distances of samples.

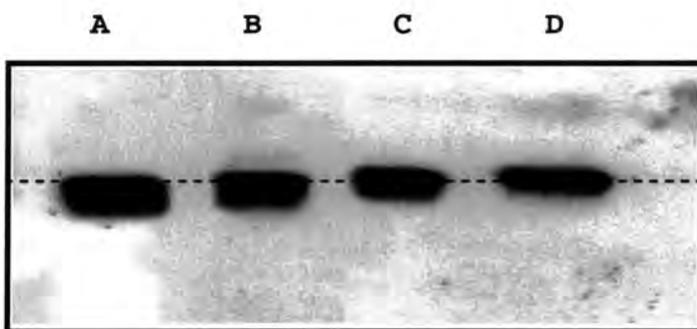


Figure 3.10 (B) Immunoblotting using poly anti-nGNA antiserum of purified recombinant wild-type and mutant GNA by anion-exchange chromatography. Lane A: rGNA105; lane B: rGNA105(Y97F); lane C: rGNA109; lane D: native GNA monomer. Dashed line indicates differences in migration distances of samples.

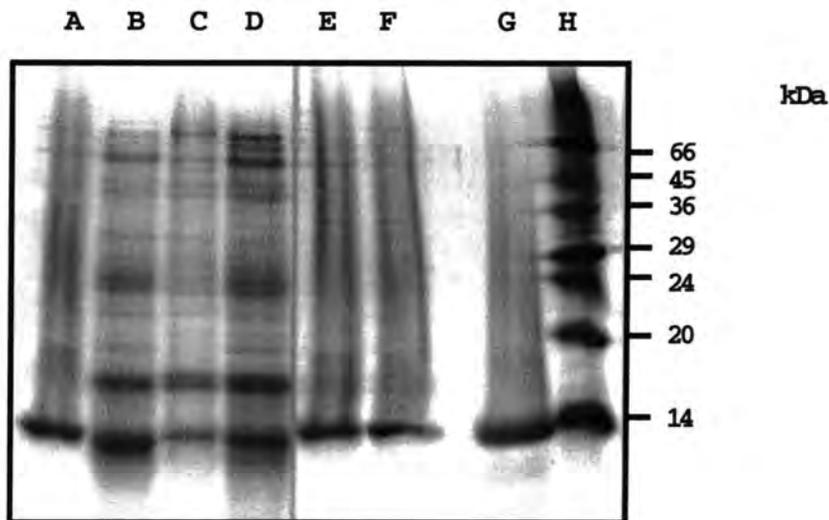


Figure 3.11 (A) SDS-PAGE analysis (12.5 % gel) by Coomassie staining of IEC-purified recombinant GNA variants. Lane A: rGNA105; lane B: rGNA105(D91L); lane C: rGNA105(N93L); lane D: rGNA105(D91L/N93L); lane E: rGNA109; lane F: rGNA109(Y97F); lane G: nGNA; lane H: SDS7 molecular weight markers.



Figure 3.11 (B) Immunoblot analysis using polyclonal anti-nGNA antiserum of IEC-purified recombinant GNA variants. Lane A: rGNA105; lane B: rGNA105(D91L); lane C: rGNA105(N93L); lane D: rGNA105(D91L/N93L); lane E: rGNA109; lane F: rGNA109(Y97F); lane G: nGNA.

3.2.4 Characterisation of rGNA Variants Expressed in *E. coli*

SDS-PAGE of the purified forms of rGNA (Fig. 3.10), after extensive reduction with β -mercaptoethanol, showed that each gave a single band on gel electrophoresis, indicated mol. wt. approx. 12,000 kDa. rGNA105 and rGNA105(Y97F) gave bands migrating slightly faster than that of native GNA monomer, whereas rGNA109 gave a band with mobility very similar to that of native GNA, suggesting that native GNA does indeed contain the extra four amino acids present in rGNA109 (Fig. 3.10). All the forms of rGNA were recognized by polyclonal anti-GNA antiserum raised against protein purified from snowdrop (native GNA) on Western blots, showing that the major epitopes recognised by the antibodies are present in the rGNAs. N-terminal sequencing of rGNA105 showed that an extra methionine residue, added in the expression construct in order to provide a start codon, was present in the rGNA when compared to the native GNA, but that the N-terminal sequence was otherwise identical to the published sequence (data not presented; Fig. 3.1).

The lectin preparations were analysed by gel filtration on a Superose-12 column to determine whether tetrameric molecules similar to native GNA were formed by recombinant GNA forms. Recombinant GNA105, rGNA109 and rGNA105(Y97F) all gave similar gel filtration profiles, with a major protein peak at the same elution volume as that given by a sample of native GNA, corresponding to tetrameric GNA molecules (indicated M_r ~35kDa). Only data for rGNA105 is presented in Figure 3.13. It was concluded that correct assembly of GNA subunits into tetrameric molecules had occurred.

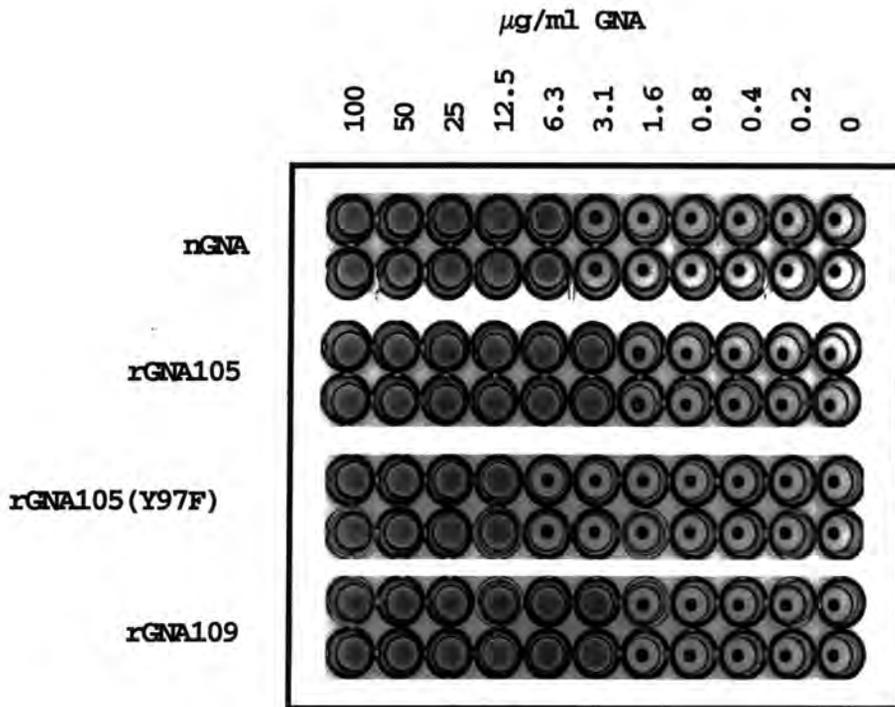


Figure 3.12 (A) Haemagglutination assays of recombinant wild-type and mutant GNA in a microtitre plate. All samples were dialysed against PBS before testing. A two-fold serial dilution of each lectin (initial concentration 100 $\mu\text{g/ml}$) was done in the presence of 1 % rabbit erythrocyte suspension. In the absence of lectin, no agglutination occurs when the erythrocytes precipitate on the bottom of the well.

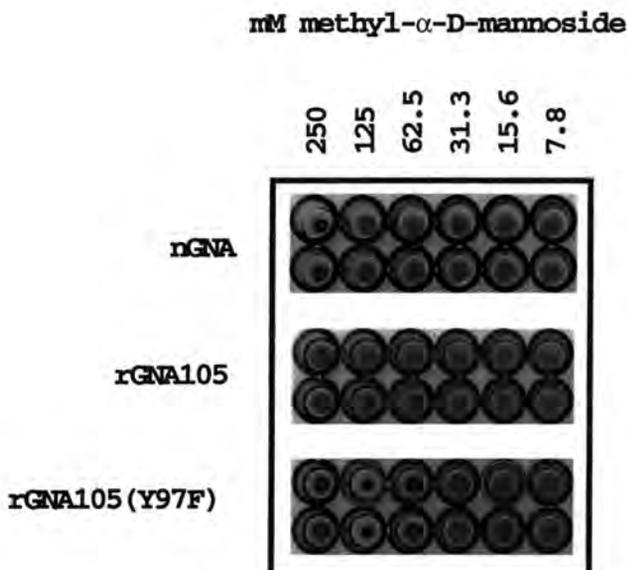


Figure 3.12 (B) Haemagglutination inhibition assays of rGNA105, rGNA105(Y97F) and native GNA. A final lectin concentration of 15 $\mu\text{g/ml}$ was used in each well. Methyl- α -D-mannoside was serially diluted starting from a concentration of 250 mM.

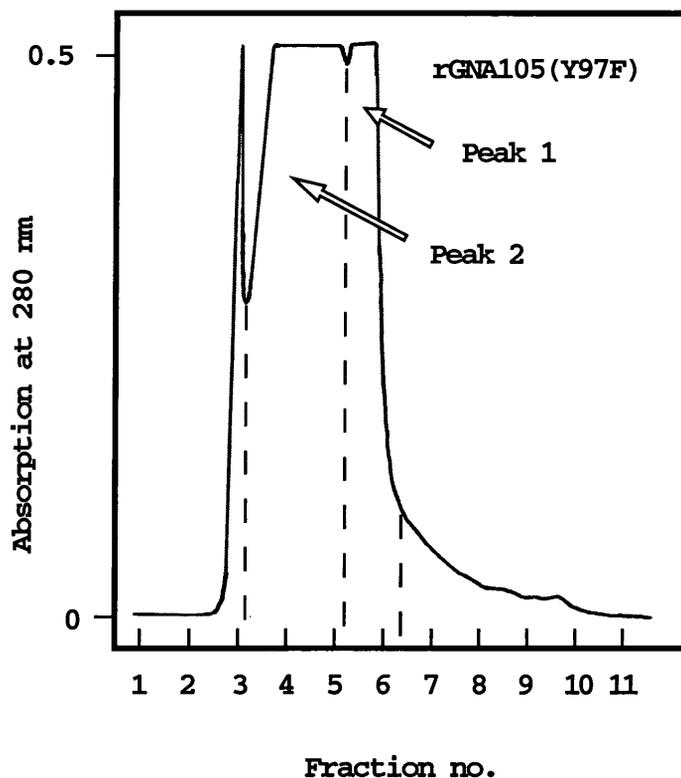
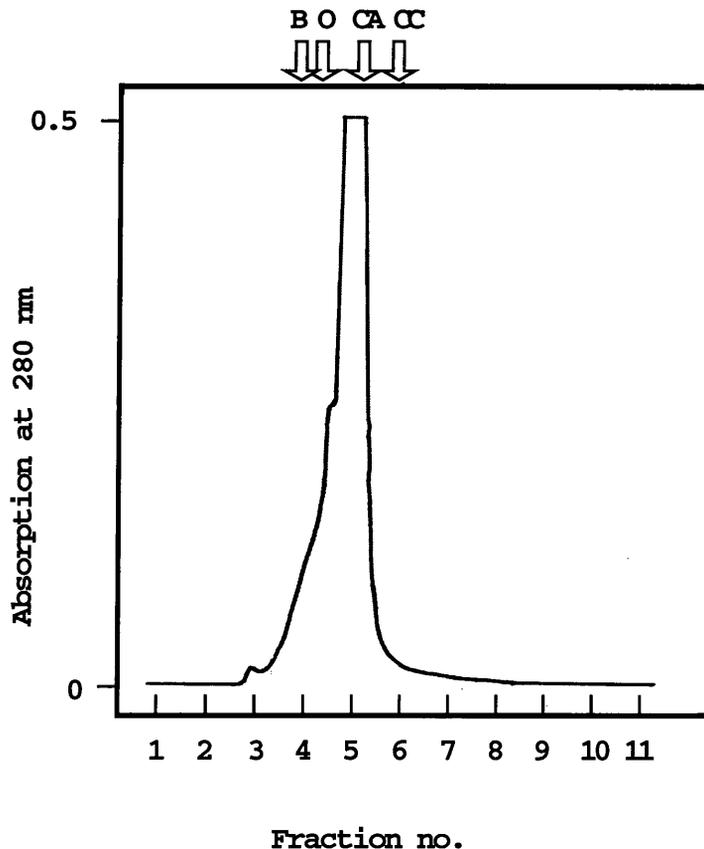


Figure 3.13 Resolution of IEC-purified rGNA105 and rGNA105(Y97F) by gel filtration. Retention time of Peak 1 of rGNA105(Y97F) corresponds to major peak of rGNA105. In top figure, arrows indicate relative elution volumes of standard proteins for calibration of gel filtration column. B: bovine serum albumin (66 kDa); O: ovalbumin (45 kDa); CA: chicken albumin (29 kDa), CC: cytochrome C (12.4 kDa).

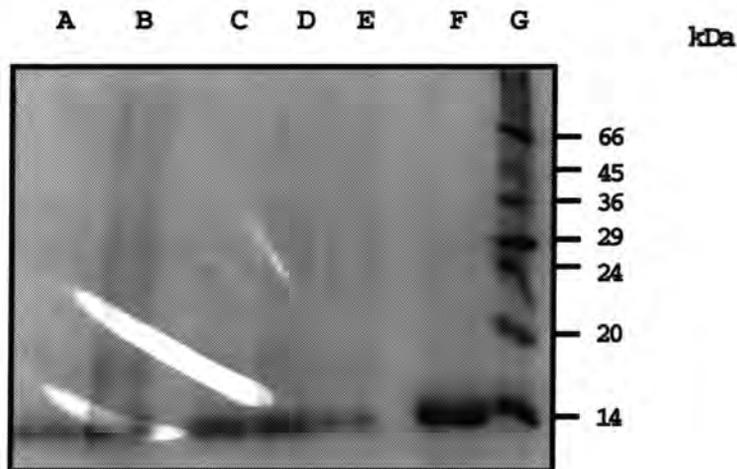


Figure 3.13 continued. SDS-PAGE analysis (15 % gel) by Coomassie staining of IEC- and IEC/GF-purified rGNA105 and rGNA105(Y97F). Lane A: IEC-purified rGNA105; lane B: IEC-purified rGNA105(Y97F); lane C: IEC/GF-purified rGNA105; lane D: IEC/GF-purified rGNA105(Y97F) Peak 1; lane E: IEC/GF-purified rGNA105(Y97F) Peak 2; lane F: native GNA monomer; lane G: SDS7 molecular weight markers. (IEC: anion-exchange chromatography; GF: gel filtration.)

exhibit a similar effect. Pooled fractions corresponding to Peak 1 of IEC-purified rGNA109(Y97F) agglutinated cells at a concentration of 1 $\mu\text{g/ml}$.

3.2.6 GNA-Carboxypeptidase-Y-Binding Assay with rGNA Variants

The snowdrop lectin interacts with the mannose oligosaccharides present on the surface of carboxypeptidase-Y (CpY). This recognition feature forms the basis of a further functional assay which gives a quantitative estimate of GNA binding. GNA, present in excess, is allowed to bind to immobilised CpY, and, after washing, the amount of GNA bound is detected by anti-GNA antibodies in a standard ELISA protocol. In the presence of low concentrations of mannose, the two recombinant wild-type GNA variants differ clearly in their binding activity (Fig. 3.15(A)). Recombinant rGNA109, with the four extra amino acids at the C-terminal end of the mature protein, bound almost twice as well to CpY as recombinant rGNA105, and bound nearly as well as native GNA. rGNA105(Y97F) bound very poorly to CpY, with a final absorbance at almost a third of its rGNA105 counterpart. Direct ELISA was used to show that the differences in binding to CpY between the rGNA forms was not due to differing reactivities with the anti-GNA antibody used. However, the recognition level of rGNA105(Y97F) by the antibodies was lower than that of rGNA105, even under saturating conditions (Fig. 3.16).

In a separate experiment, CpY-binding of rGNA109 was compared to that of rGNA109(Y97F) (Fig. 3.17(A)). Here, a similar result was obtained as for rGNA105(Y97F); the mutant bound considerably less readily (approximately a three-fold reduction) than its native counterpart at low mannose concentrations. No direct

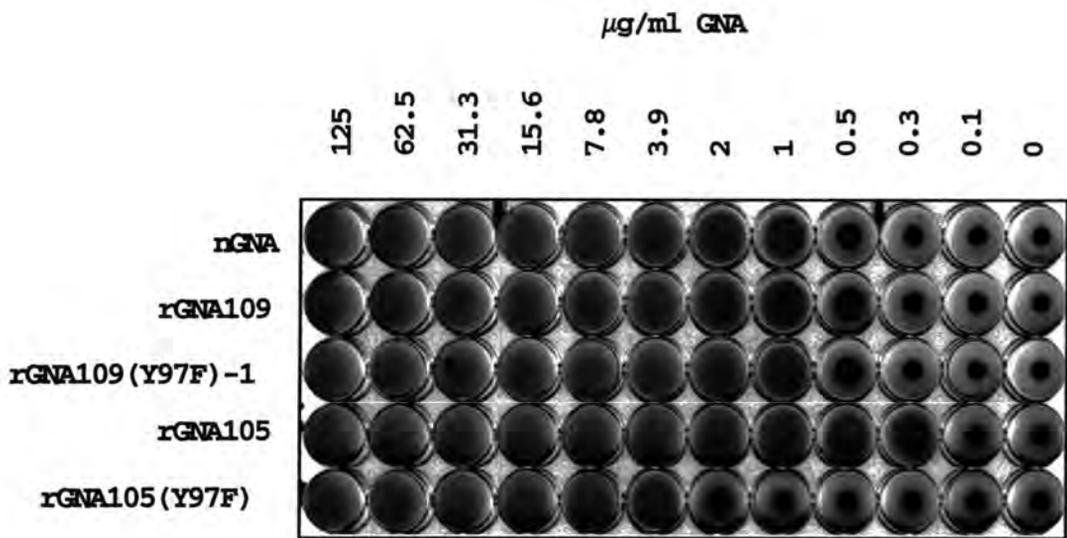


Figure 3.14 Haemagglutination assays of IEC-purified recombinant wild-type and mutant GNA. A two-fold serial dilution of each lectin (initial concentration 125 $\mu\text{g/ml}$) was done in the presence of 1 % rabbit erythrocyte suspension (CSL, York). rGNA109(Y97F)-1: Peak 1 fraction (see Fig. 3.8). IEC: anion-exchange chromatography.

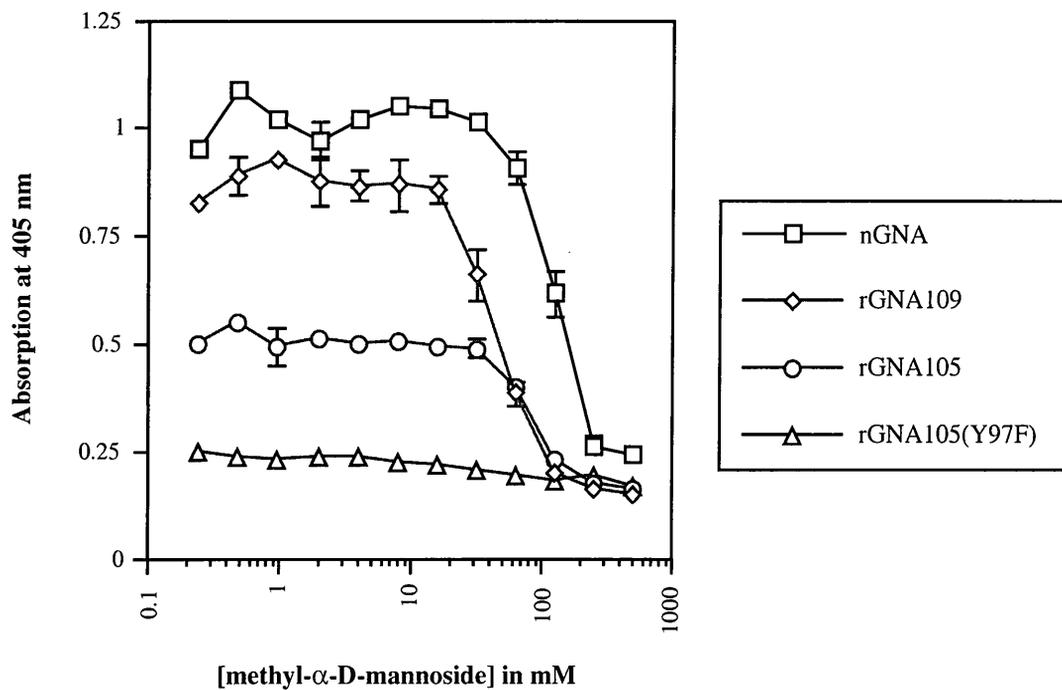


Figure 3.15 (A) Binding of GNA variants to CpY in the presence of varying amounts of mannose, as detected by polyclonal anti-nGNA antibodies.

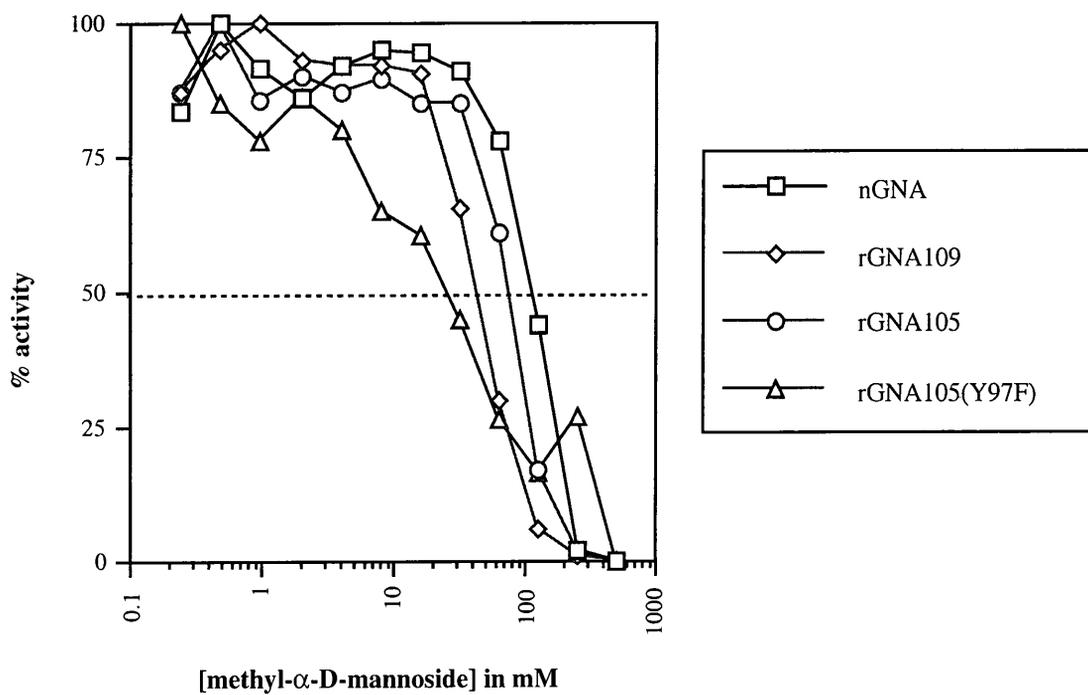


Figure 3.15 (B) Inhibition of binding of GNA variants to CpY by mannose.

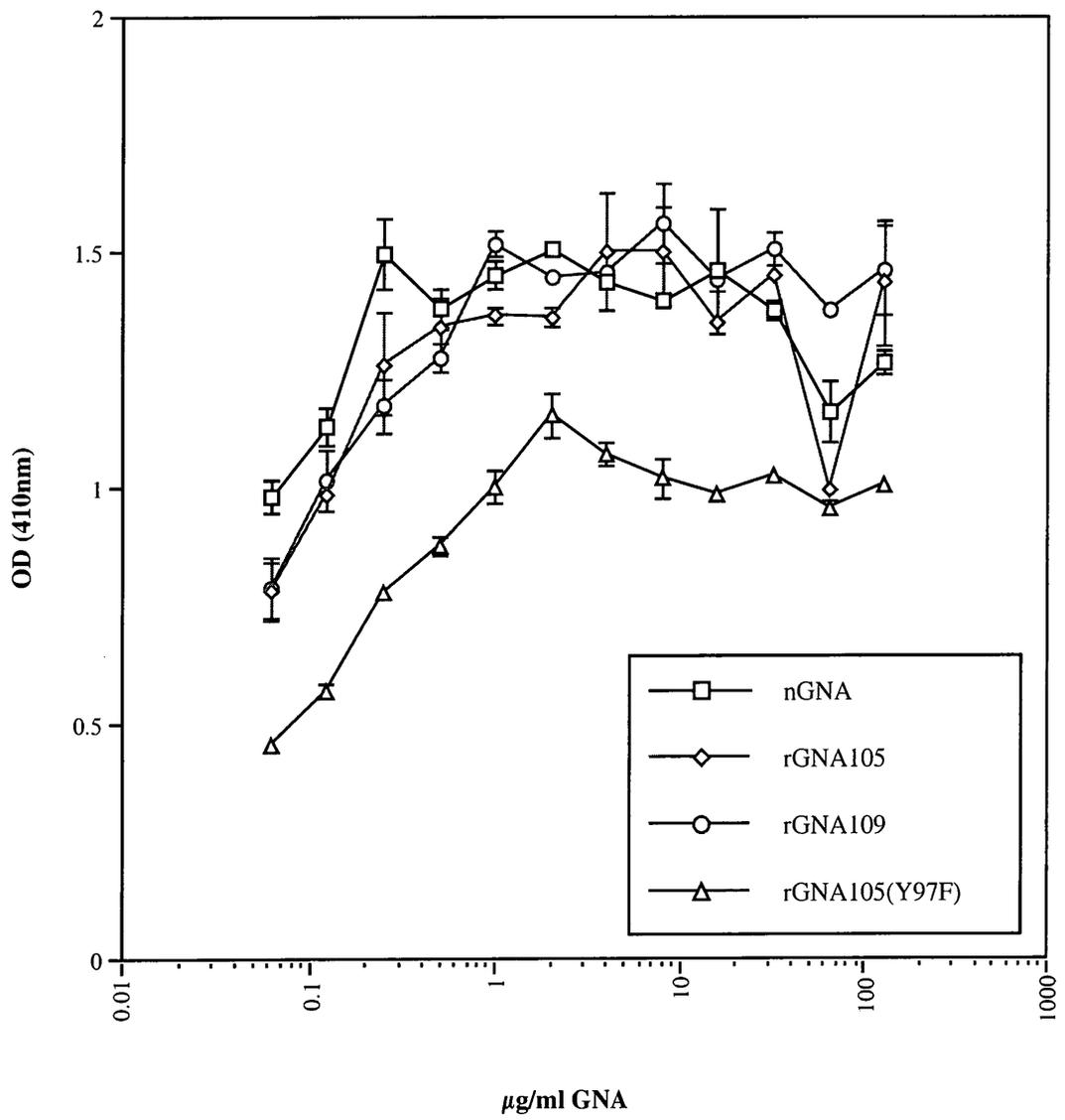


Figure 3.16 Recognition of GNA variants by polyclonal anti-nGNA antibodies by direct ELISA.

ELISA of rGNA109(Y97F) was carried out, but a high purity of the fraction (Peak 1) was demonstrated by SDS-PAGE analysis (Fig. 3.9).

Increase in the mannose concentration present in this assay resulted in a decrease in binding activity for all the GNA forms, as expected, with binding being essentially unaffected below a limiting concentration of mannose (10 – 20 mM mannose) and essentially reduced to a constant minimum above a limiting concentration of mannose (100 – 200 mM mannose). The mannose concentration required to reduce binding activity to 50 % of maximum [I_{50}] was significantly lower for rGNA105(Y97F) than that for native GNA (approximately five-fold; Fig. 3.15(B)). For recombinant rGNA105, the mannose [I_{50}] was similar to that of native GNA, but the mannose [I_{50}] for rGNA109 was lower (by a factor of approx. 2; Fig. 3.15(B)).

In Figure 3.17(B), the mannose concentration required to halve the maximum binding activity for rGNA109(Y97F) was approximately three-fold less than that for native GNA. The mannose concentration [I_{50}] for rGNA109 was two-fold greater when compared to the [I_{50}] of the mutant.

3.2.7 Binding of rGNA Forms to Polypeptides Extracted from BPH Gut

Polypeptides extracted from dissected guts of brown planthopper (BPH) were separated by SDS-PAGE, blotted onto nitrocellulose, and polypeptides binding GNA were detected by reaction with an excess in concentration of rGNA variants, followed by detection of bound GNA using anti-GNA antibodies. The binding activities and specificities of rGNA105 and rGNA105(Y97F) were compared by probing duplicate

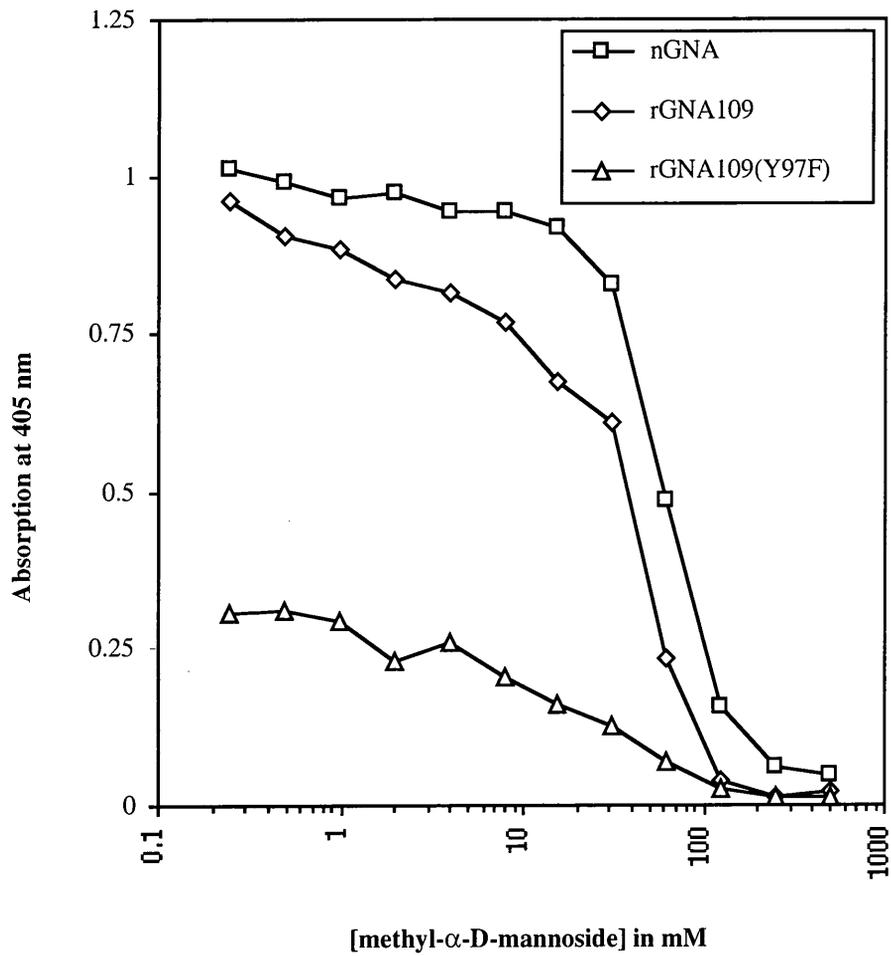


Figure 3.17 (A) Binding of rGNA109 and rGNA109(Y97F) to CpY in the presence of varying amounts of mannose, as detected by polyclonal anti-nGNA antibodies.

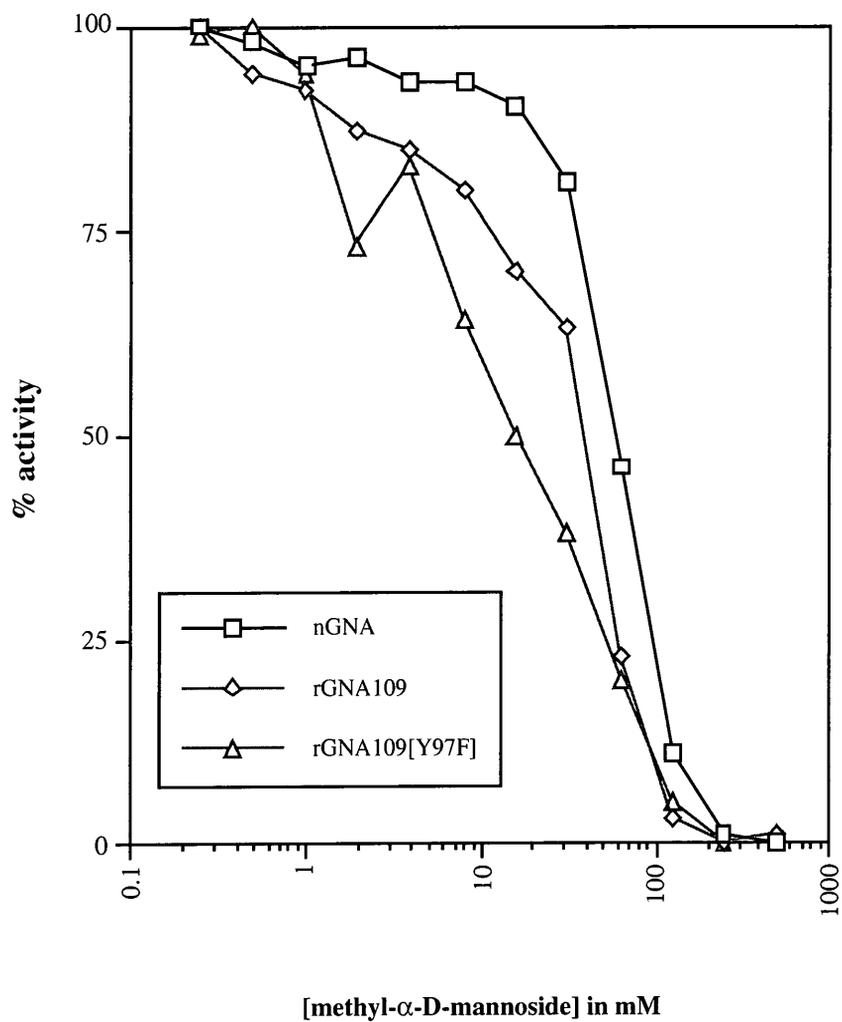


Figure 3.17 (B) Inhibition of binding of rGNA109 and rGNA109(Y97F) to CpY by methyl- α -D-mannoside.

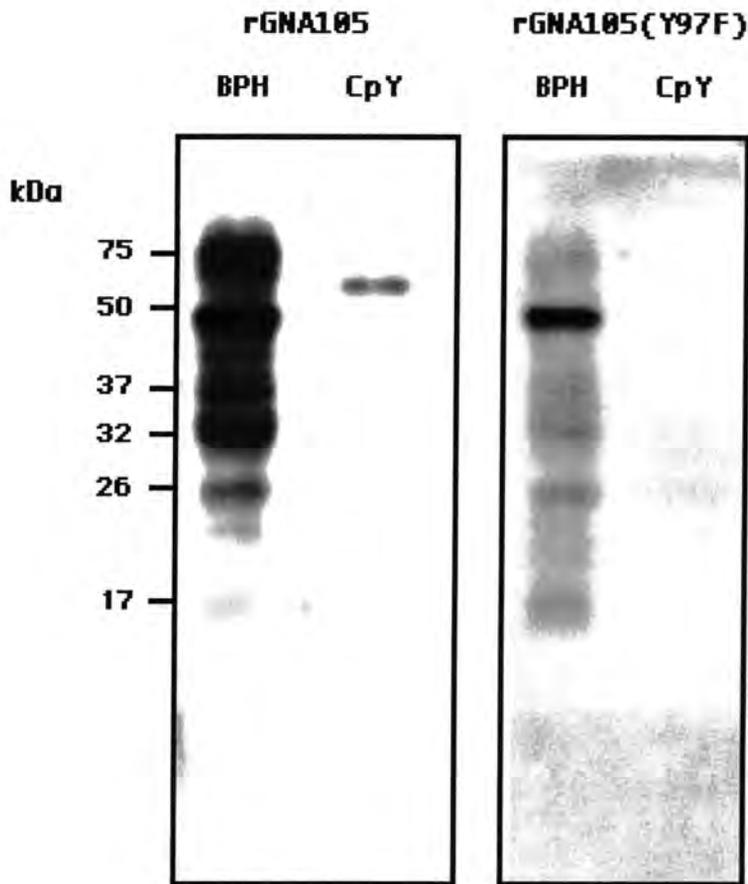


Figure 3.18 Immunoblot analysis of BPH gut glycoproteins probed with rGNA105 or rGNA105(Y97F) by polyclonal anti-nGNA antibodies. CpY was used as control. BPH: Brown Planthopper; CpY: Carboxypeptidase-Y

blots with the two GNA variants. Results are shown in Fig. 3.18. rGNA105 was found to bind to at least 6 BPH gut polypeptides, giving a band pattern similar to that given by probing with native GNA (results not presented); it also bound to carboxypeptidase Y. The molecular weights of glycosylated polypeptides detected by rGNA105 (and native GNA) ranged from 17 to 75 kDa, with three major bands at 32 kDa, 50 kDa and 75 kDa. In contrast, rGNA105(Y97F) only bound detectably to 2 BPH gut polypeptides; it bound to the 50 kDa polypeptide detected by rGNA105, but much more weakly, and also bound to the 17 kDa polypeptide, at an intensity comparable to rGNA105. As expected, binding of rGNA105(Y97F) to carboxypeptidase Y was also observed, although binding was weaker than that shown by rGNA105.

3.2.8 Bioassays of rGNA Variants Against BPH

Native GNA, rGNA105 and rGNA105(Y97F), at added protein levels of 0.05 % (w/v; approx. 10 μ M) were tested for toxicity towards third instar BPH nymphs in an artificial diet bioassay (Fig. 3.19). The concentrations of the recombinant lectins were checked by spectrophotometrical analysis before incorporation into the diet to ensure similar amounts of protein were added. As expected, native GNA showed a significant reduction in nymph numbers when compared to control diet (Powell *et al.*, 1993). Recombinant GNA105 showed a similar survival curve as for native GNA; the two GNA preparations had corrected mortality values (Abbot, 1925) of 84 % and 90 % for native GNA and rGNA105, respectively, indicating a substantial toxic effect. rGNA109 was also assayed for effects on BPH, and gave survival curves not significantly different from those given by native GNA and rGNA105 (Fig. 3.19). On

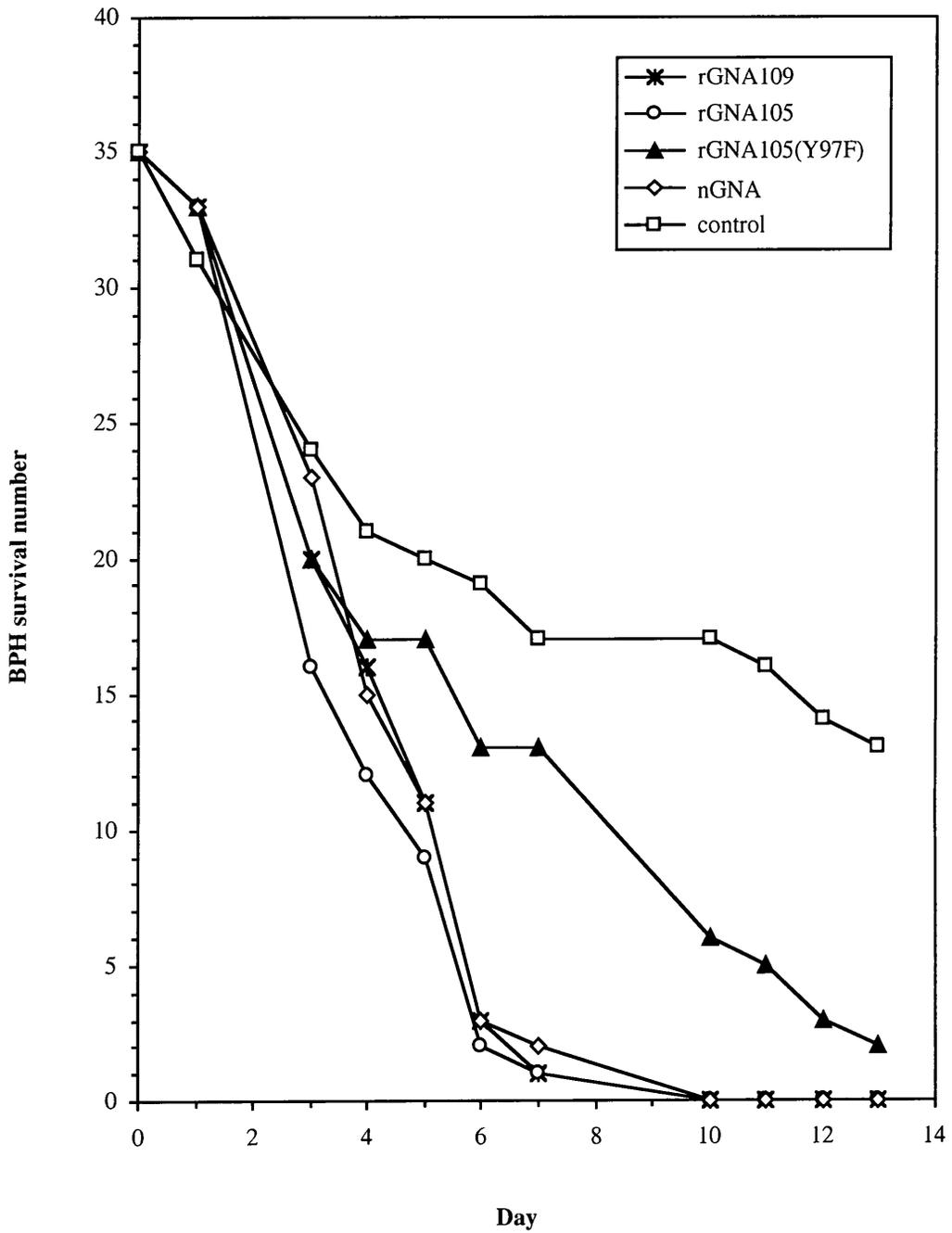


Figure 3.19 Effects of recombinant wild-type and altered GNA to BPH survival. Each treatment consists of 7 replicates with 5 third instar nymphs per replicate. BPH: Brown Planthopper.

the other hand, rGNA105(Y97F) had much less effect on BPH survival, with a corrected mortality value at 30 %. Statistical analysis of the survival curves showed that the curves for all three lectins differed significantly from control (logrank test, $p < 0.05$), but that curves for native GNA and rGNA105 did not differ significantly. The curve for rGNA105(Y97F) differed significantly from both the control and the curves for native GNA and rGNA105.

3.3 Discussion

Expression of both rGNA105 and rGNA109 in *E. coli* has shown that the native GNA purified from snowdrop has a molecular weight on SDS-PAGE that resembled the molecular weight of rGNA109 more closely than that of rGNA105. This lectin is not glycosylated *in planta*, since it does not contain a potential glycosylation site, and thus the molecular weight determined on SDS-PAGE must reflect the polypeptide only. It is noteworthy that the band on SDS-PAGE produced by native GNA is rather broad and diffuse, suggesting the presence of different molecular species; this is consistent with reports in the literature that the lectin exists as multiple isoforms in snowdrop, and is encoded by a multigene family (Van Damme *et al.*, 1991a).

The four C-terminal residues, (Thr₁₀₆His₁₀₇Thr₁₀₈Gly₁₀₉), clearly are not required for GNA to show binding activity to both complex and simple carbohydrates, since both the rGNA105 and rGNA109 forms produced in this work bound to a mannose-agarose column, and agglutinated red blood cells to the same dilution of lectin. The previous expression construct for GNA also lacked these residues, substituting a his-tag sequence (Leu-Glu-His₆), and also gave rise to functional lectin protein. Although serial dilution assays are necessarily inaccurate, and results vary over a tenfold range

when using blood from different rabbits (Longstaff *et al.*, 1998), we reproducibly observed that the rGNA proteins would agglutinate erythrocytes at one serial dilution (2-fold lower concentration) than native GNA. This suggests that the native GNA contains isoforms of the protein which are less active as haemagglutinins than the isoform represented by the cDNA used to produce the expression constructs. Differences in sensitivity to mannose in inhibition of haemagglutination were not observed between native GNA, rGNA105 and rGNA109. When the lectins were assayed for binding to carboxypeptidase Y, however, differences in binding between rGNA105 and rGNA109 were observed, with rGNA109 binding more to the glycoprotein, and resembling native GNA in its binding activity. Direct ELISA experiments demonstrated that quantitative detection of rGNA105 and rGNA109 by anti-GNA antibodies was virtually the same, and thus the observed differences in binding to carboxypeptidase Y were not a result of epitope differences between these two lectins. These results suggest that the four extra C-terminal amino acid residues do play a significant role in high affinity mannose binding, at least in certain cases, since more lectin must have bound per molecule of carboxypeptidase Y for rGNA109 than for rGNA105. Possibly the additional amino acids allow lectin binding to take place at low-affinity sites at which the "truncated" rGNA105 variant is unable to bind. Examination of the published GNA structure shows that residues at positions 106 - 109 are in a suitable location to influence binding. In the subunit pairs A-D and B-C, residues 99-109 of the C-terminal arm from one subunit interact with their complementary subunit forming a tight dimer, and creating at the interface mannose-binding site I, where these residues contribute to subsidiary interactions with the saccharide through polar and nonpolar van der Waals bonding. Interestingly, the rGNA109 variant of GNA was more sensitive to mannose inhibition in this

carboxypeptidase Y binding assay than either rGNA105 or native GNA, suggesting that residues 106-109 also have subtle effects on the specificity of the lectin towards complex carbohydrates vs. simple sugars. Although the four C-terminal residues are involved in interactions between subunits, their removal does not affect the ability of the protein to form tetramers, since similar results were given on gel filtration by rGNA105 and rGNA109 proteins.

When attempting to alter the carbohydrate binding properties of GNA, we chose binding site I as the target for mutation. This site is utilized for binding mannose monosaccharides or terminal mannose with high occupancy (Hester & Wright, 1996). The resulting protein did not have gross alterations in structure, since the gel filtration profile showed that the Y97F mutant was able to comigrate with authentic nGNA. The antibodies raised against native GNA also detected rGNA105(Y97F). The quantitative estimates of antibody binding showed that the curve of antibody binding against lectin concentration had a similar value for 50% of maximum binding, but a lower saturation level, indicating that the amount of antibody bound to the rGNA(Y97F) mutant was approximately 85 % of that bound to the non-mutated rGNA105. This correlates well with the yield estimate (approximately 85 – 90 %) of purified mutant after anion exchange chromatography as observed from Coomassie-stained SDS PAGE gel (Fig. 3.10 (A)).

The functional properties of the Y97F mutant clearly indicate that the alteration to binding site I has altered carbohydrate binding properties. The failure of the protein to be retained on a mannose-agarose column contrasts with the binding of the non-mutated recombinant GNA variants. However, the mutated lectin has not lost carbohydrate binding ability, since it is still active as a haemagglutinin, although four

times the concentration of mutant rGNA105(Y97F) was required to achieve complete agglutination when compared to rGNA105. The mutant was also more sensitive to inhibition by mannose, with approximately four-fold lower mannose concentrations being required to inhibit agglutination of erythrocytes. Binding to the mannose residues on the surface of the red blood cells is most likely to be mediated through participation by all three binding sites per GNA subunit (twelve per tetramer). In the case of the mutant, hemagglutination was not as effective as rGNA105 since the binding activity of site I had been reduced or abolished. Therefore, complete agglutination by rGNA105(Y97F) was only possible in the presence of a lower concentration of free mannose allowing a greater availability of the sites to carbohydrate binding. Similar increased sensitivity to mannose inhibition are observed in the assay comparing binding of the rGNA105(Y97F) mutant and rGNA105 to carboxypeptidase Y. More significantly, however, the mutant showed greatly reduced absolute levels of binding to carboxypeptidase Y in the CpY binding assay, which suggests that binding site I in GNA is much more significant in binding to carboxypeptidase than in interactions with glycoproteins on the surface of red blood cells. This conclusion is supported by the observation of reduced binding to CpY for rGNA105 compared to rGNA109, but similar haemagglutination properties; since the C-terminal residues are near binding site I in the GNA structure, they can influence binding at this site.

The monocot mannose-binding lectins interact most specifically with high mannose glyconjugates that possess α -1,3 linkages (Shibuya *et al.*, 1988; Kaku *et al.*, 1991). However, fine differences in the sugar recognition properties of these lectins exist and it may be that the degree of binding affinity to mannose-containing oligosaccharides

is a factor which contributes to the observed differences in antimetabolic effect of these related lectins towards BPH (Powell *et al.*, 1995a;1995b). It is now clear that in the rice brown planthopper, GNA binds to the surface of cells lining the gut wall (Powell *et al.*, 1998) and can interact with glycoproteins extracted from midgut tissues, as has been suggested for other lectins and insects (Gatehouse *et al.*, 1984; Eisemann *et al.*, 1994). Western blotting analysis of total BPH gut protein probed with rGNA105 revealed a limited number of glycosylated proteins containing mannose residues, that span over a wide molecular weight range with differing band intensities (Fig. 3.18). The pattern is similar to that given when native GNA is used as a probe on a similar gut protein preparation. The differences in band intensities may be attributable to either the abundance or binding affinity of the GNA "receptors". The reduced binding ability of the mutant rGNA(Y97F) was clearly shown in this assay, by a very much reduced level of binding to BPH gut polypeptides, so that only two bands were detectable. Although one of these bands was detected as high intensity by rGNA105, the mutant bound only at low intensity. However, the band at 17 kDa was detected with similar intensity by both rGNA105 and rGNA105(Y97F), showing that the specificity of binding to different glycoproteins had been affected by the mutation made to rGNA(Y97F). This change in specificity may be due to binding site I no longer playing a major role in binding, so that the activities of binding sites II and III become predominant. Mutant rGNA105(Y97F) also bound to carboxypeptidase Y on blots much less readily than rGNA105, further supporting the results obtained from the carboxypeptidase-Y-binding assays (see above). However, the level of binding may only be used qualitatively, not quantitatively, since the protein samples in the blots were denatured for SDS-PAGE.

The results of the insect bioassay show that native GNA, rGNA105 and rGNA109 are indistinguishable in their effects on BPH survival, in agreement with previous results for rGNA expressed as a his-tagged fusion protein (Longstaff *et al.*, 1998) (Fig. 3.19). However, the present data show that mutant rGNA105(Y97F) was significantly less toxic than native GNA, or rGNA105, but the level of toxicity was not completely abolished. These observations are in agreement with the hypothesis that binding of the lectin to “receptors” in the insect gut is necessary for toxicity, since binding of the rGNA(Y97F) mutant to BPH gut glycoproteins was very much reduced, compared to rGNA105, but was still possible, as shown by the blots of BPH gut proteins probed with rGNA variants. The reduced toxicity of rGNA(Y97F) could be due to the lower levels of overall binding shown by the mutant, or its increased specificity towards certain glycopolypeptides. In particular, the glycosylated polypeptides of 17 and 50 kDa may play a significant role in establishing the link between GNA binding and its toxicity towards brown planthoppers, since these are the only bands detected by rGNA(Y97F) on gut protein blots.

Chapter 4

Expression of Functional PHA & GNA in *P. pastoris*

4.1 Introduction

In the previous chapter, it was shown that functional GNA was expressed in the insoluble form in *Escherichia coli*. This lectin must be solubilised by denaturation-renaturation in order to recover its proper native conformation and activity, a step which proves cumbersome and inefficient.

Many plant lectins (not GNA) are also glycosylated and *E. coli* is incapable of glycosylating eukaryotic proteins expressed in this host. The presence or absence of the carbohydrate side chains can have significant effects on the functional and/or physical properties of the recombinant proteins. For example, concanavalin A is activated by deglycosylation *in planta* (Min *et al.*, 1992; Sheldon & Bowles, 1992) and nonglycosylated PHA-L expressed in transgenic plants gave poor quality crystals under conditions where glycosylated PHA-L was crystallized successfully (Dao-Thi *et al.*, 1996).

Alternative heterologous expression systems were looked into. Attempts to express plant lectins in the yeast *Saccharomyces cerevisiae* have met with mixed success. However, associated problems exist, too. PHA-L expressed in *Saccharomyces* accumulates mainly in the vacuole (Tague & Chrispeels, 1987), due to the presence of 'cryptic' vacuolar targeting determinants (Von Schauwen & Chrispeels, 1993), and

only about 1 % is secreted. In addition, a significant proportion of PHA-L expressed in this host was not correctly processed, approximately half of the PHA-L accumulating in the vacuole appeared to contain the uncleaved signal peptide, and all of the secreted PHA-L was in this unprocessed form (Tague & Chrispeels, 1987). A second related lectin, from *Dolichos biflorus*, could be directed into a secretory pathway when expressed at low levels, but accumulated in the cells (not in the vacuole) when expressed at high levels (Chao & Etzler, 1994). In both cases, the functional properties of the recombinant lectin were not reported. The only report of the correct processing and secretion of a plant lectin in *Saccharomyces cerevisiae* is that of wheat germ agglutinin which was secreted and exhibited sugar binding activity. However, yields of protein were relatively low, of the order of 200 $\mu\text{g} / \text{l}$ (Nagahora *et al.*, 1992). Thus it would appear that *Saccharomyces* is not a satisfactory host for lectin expression.

The development of relatively facile expression methodologies for the methylotrophic yeast *Pichia pastoris*, especially for the secretion of recombinant proteins, offers an alternative for expressing plant lectins (Cregg *et al.*, 1993; Sreekrishna *et al.*, 1997). In this chapter, two plant lectins were selected for expression in *Pichia*, *Phaseolus vulgaris* agglutinin (PHA; both the E- and L-forms) and *Galanthus nivalis* agglutinin (GNA). The rationale for this selection was (i) PHA-L is incorrectly processed and accumulates in the vacuole of *Saccharomyces* and (ii) GNA accumulates in inclusion bodies when expressed in *E. coli*. In addition, PHA and GNA are members of two very different lectin families, the legume lectins and the monocot mannose-specific lectins, respectively. These families are unrelated in sequence and we would therefore

predict that, if these lectins could be expressed, then *Pichia* may become a useful system for the expression of a wider range of plant lectins.

In this chapter, expression of functional PHA and GNA and secretion into the culture medium is demonstrated. Also, it is shown that the PHA-E signal peptide directed the secretion of proteins - two plant lectins, PHA-E and GNA, and a protein which is not a lectin and not derived from a plant, GFP - which are correctly processed at the amino-termini, whereas these proteins secreted under the control of the *Saccharomyces* prepro- α -factor sequence have heterogeneous N-terminal extensions. This suggests that the PHA-E signal peptide may have a wider utility in the production of recombinant proteins in *Pichia*.

4.2 Results

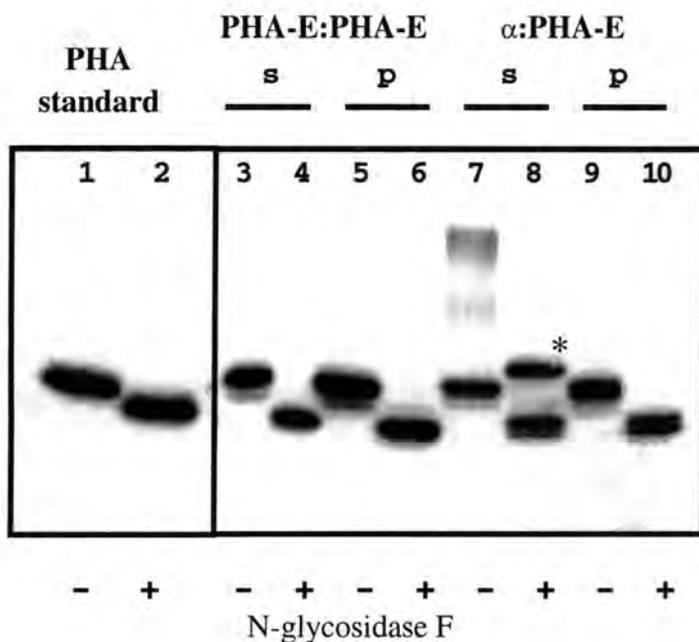
4.2.1 Expression of Active PHA and GNA in *P. pastoris*

Initially, two sets of constructs for the expression of PHA (E and L forms) and GNA were prepared. In the first, the lectin coding sequence contained the 'native' signal peptide sequence and in the second, the signal peptide sequence was replaced by the *Saccharomyces* α -factor prepro--sequence (extending as far as the processing site of the Kex2 protease, i.e. omitting the 'Glu-Ala' repeats). All construct designations and corresponding sequences are shown in Table 1.1 and Figure 4.4. Coding sequences were placed under the control of the methanol-inducible *AOX1* promoter present in the pPICZ series of *Pichia* expression vectors (Invitrogen). Constructs were transformed into *Pichia* strains GS115 or KM71 and the supernatants from methanol-induced cultures were analysed for the presence of the corresponding lectin.

Figures 4.1 (PHA) and 4.2 (GNA) summarise results for the expression and purification studies of lectins in *Pichia pastoris*. Proteins which reacted with the appropriate anti-lectin antibody were present in supernatants of induced cultures for all the PHA expression constructs, both with the native signal peptide (PHA-E:PHA-E and PHA-L:PHA-L) and the α -factor prepro--sequence (α :PHA-E and α :PHA-L) (Figure 4.1 (A), lanes 3 and 7). Results for E- and L-forms of PHA were essentially the same and a representative analysis of PHA-E is shown. In contrast to PHA, eight independent transformants containing the leader and mature GNA coding sequence (GNA:GNA) failed to produce GNA. Even a construct containing the full length GNA coding sequence, including its C-terminal pro-extension, (GNA:GNA(2)), showed that no expression could be detected by dot-blot screening (results not shown). However, GNA expression could be detected readily in *Pichia* clones containing the construct where the native GNA signal peptide was replaced by the α -factor prepro--sequence (α :GNA) (Figure 4.2, lane 1). Transformation into either strain GS115 or KM71 did not produce any significant differences in levels of lectin secretion (results not shown). Time course analyses revealed that maximum expression levels were achieved 5 days (PHA) and 7 days (GNA) after methanol induction (results not shown).

Analysis by SDS-PAGE and Western blotting showed that polypeptides of similar molecular weights to appropriate lectin standards, recognised by anti-PHA or anti-GNA antibodies, were present. However, the polypeptide profiles of the bands recognised by anti-lectin antibodies after Western blotting differed from the 'native'

(A)



(B)

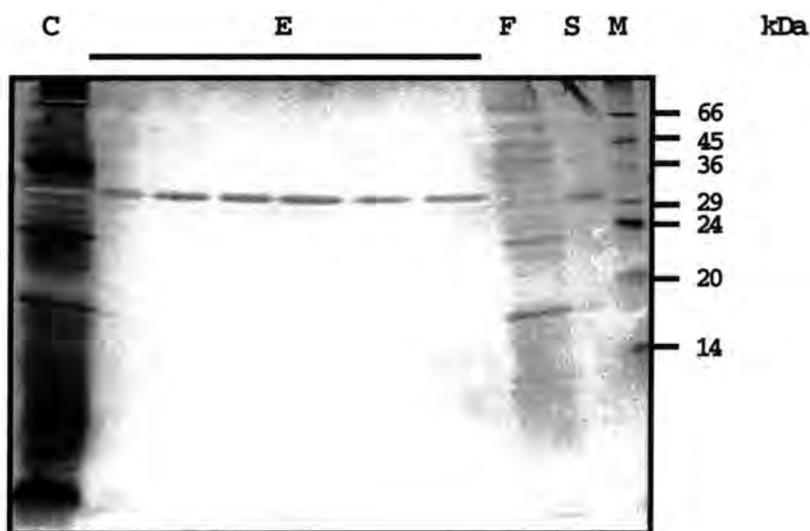


Figure 4.1 (A) Summary of expression and purification of PHA in *Pichia*, and (B) silver stained gel showing the affinity purification of recombinant PHA-L on a thyroglobulin-agarose column, expressed with the native signal peptide. (A) Data for PHA-E but PHA-L gave essentially identical results. Western analysis of PHA-E in culture supernatants (s) or affinity-purified (p) samples expressed using either the native signal peptide or the α -factor prepro sequence. Samples were treated (\pm) with N-glycosidase F to resolve differences in glycosylated forms. The conspicuous 'smear' present in the α -factor supernatant is resolved to a single deglycosylated higher molecular mass form of PHA-E (asterisk) not present in the affinity-purified sample. (B) Lane C: ammonium sulphate precipitate of induced culture supernatant; lane E: eluted fractions from thyroglobulin column; lane F: column flowthrough; lane S: PHA standard; lane M: molecular weight markers.

lectins. In the case of PHA, the bands recognised by anti-PHA antibodies also differed between constructs, depending on whether the α -factor prepro--sequence or native signal peptide sequence was used. The differences in the polypeptide profiles derived from different constructs are due to differences in processing between native signal peptides and the α -factor prepro--sequence, and are described in more detail in the following section. Levels of PHA and GNA in the culture supernatants were estimated from Western blots as being in the range 0.4 - 1.0 and 1 - 2 mg / l, respectively. For PHA, estimates of the relative amounts of protein in the cells and supernatant suggested that at least 50 % of the protein was secreted (results not shown). The proportion of GNA inside cells could not be estimated as GNA also binds to cell wall mannose residues.

Both PHA and GNA were purified from culture supernatants by affinity chromatography, using the functional activity of the lectins, on thyroglobulin- and mannan-agarose columns, respectively. All four forms of PHA (from the PHA-E:PHA-E, α :PHA-E, PHA-L:PHA-L, and α :PHA-L constructs) and GNA (from the α :GNA construct) bound to, and could be eluted from, the corresponding affinity column to yield proteins free of contaminants. A representative purification of PHA-L, expressed using the native signal peptide sequence, is shown in Figure 4.1 (B), indicating the protein is essentially pure as judged by silver staining. Similar results were obtained for the other PHA constructs. Recombinant PHA-E (from both the PHA-E:PHA-E and α :PHA-E constructs) and GNA agglutinated rabbit erythrocytes at concentrations similar to those of native lectins assayed under the same conditions

(Table 4.1), and which are consistent with concentrations reported for the corresponding commercially available plant-derived lectins. Sugar inhibition of haemagglutination was also consistent with the results for native lectins (results not shown). PHA-L does not agglutinate erythrocytes at these concentrations, and thus was not assayed. The data from haemagglutination assays and affinity purification shows that the lectins produced in *Pichia* are functional proteins.

4.2.2 Processing of Proteins Expressed Using the α -Factor Prepro--Sequence

Although the constructs in which the α -factor prepro--sequence was used as a signal peptide led to the secretion of functional lectins from *Pichia*, analysis of polypeptides present in culture supernatants showed that each construct gave rise to multiple bands detected by anti-lectin antibodies, suggesting that processing was not occurring correctly. For both PHA and GNA, although a band of similar mobility to the 'native' protein was present in the culture supernatant, diffuse bands or smears at higher molecular weights were also present (Figure 4.1 (A), lane 7; Figure 4.2, lanes 1 and 3). The diffuse bands were absent from α :PHA constructs after purification by affinity chromatography, showing that these polypeptides do not bind to the affinity column (compare lanes 7 and 9 of Figure 4.1 (A)). For GNA, diffuse higher molecular weight bands were present both before and after purification, although a smear of immunoreactive material of very low mobility on SDS-PAGE was not present in the eluted fraction after affinity chromatography, but was present in the flow-through fraction. The diffuse bands on SDS-PAGE are diagnostic of protein glycosylation. PHA (both E- and L-forms) contains potential endogenous N-linked glycosylation

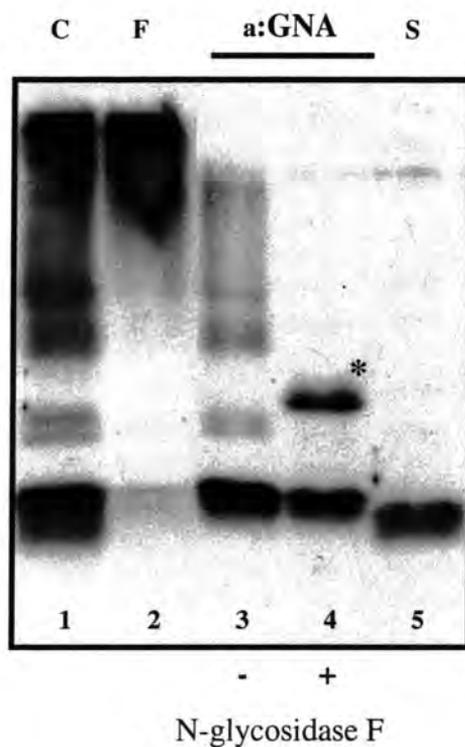


Figure 4.2 Summary of expression and purification of GNA in *Pichia* expressed using the α -factor prepro sequence. Immunoblot analysis of GNA in culture supernatants (C), column flowthrough (F), affinity-purified GNA (α :GNA) and standard native GNA (S). Affinity-purified GNA was treated (\pm) with with N-glycosidase F to resolve differences in glycosylated forms. An asterisk indicates the higher molecular weight form of GNA containing 56 residues of α -factor pro sequence.

sites, and is glycosylated *in planta* (Sturm & Chrispeels, 1986); the diffuse bands could thus reflect hyperglycosylation occurring in *Pichia*. However, mature GNA is non-glycosylated, and does not contain any potential N-linked glycosylation sites. The diffuse bands in this case therefore cannot be due to glycosylation of the mature GNA sequence.

To resolve differences in apparent molecular weight on SDS-PAGE, polypeptides were deglycosylated by treatment with N-glycosidase F. Prior to deglycosylation, affinity-purified GNA, produced by the α :GNA construct, contains a major polypeptide band at a molecular weight slightly larger than 'native' GNA, and a series of diffuse bands at higher molecular weight (Figure 4.2, lane 3). In the deglycosylated protein, the mobility of the major polypeptide band is unchanged (Figure 4.2, lane 4). However, after deglycosylation the higher molecular weight diffuse bands resolve to a single polypeptide of molecular weight approximately 6 kDa greater than native GNA (Figure 4.2, lane 4, asterisk). The higher molecular weight diffuse bands must thus represent glycosylated GNA, and since the glycosylation cannot be taking place on residues present within mature GNA, it must be occurring on incompletely processed α -factor pro-sequence. In agreement with this conclusion, GNA contained two distinct N-terminal sequences (Figure 4.4) with extensions of 9 and 56 amino acids corresponding to part of the α -factor pro-sequence. The 56-amino acid N-terminal extension on the larger polypeptide contains two of three potential N-glycosylation sites present in the α -factor pro-sequence. Taken together, this demonstrates that GNA expressed using the α -factor prepro--sequence is incompletely processed and highly glycosylated on the partially cleaved α -factor pro-sequence.

Analysis of the results obtained for PHA expressed from constructs containing the α -factor prepro--sequence produced similar results to GNA. Deglycosylation decreases the indicated molecular weight of native, glycosylated PHA (Figure 4.1, lanes 1 and 2) due to the removal of the N-linked carbohydrate side chain from the protein, although both before and after N-glycosidase treatment the native protein (E- and L-form) gives a single band after SDS-PAGE and Western blotting. For PHA expressed from the construct containing the α -factor prepro--sequence, culture supernatants contain a major polypeptide detected by anti-PHA antibodies at a similar molecular weight to glycosylated native PHA (Figure 4.1, lanes 7 and 1), and a series of diffuse bands at higher molecular weight. A faint band at slightly lower molecular weight is also present. N-glycosidase treatment of culture supernatants results in three bands being observed on Western blots (Figure 4.1, lane 8). Two bands form a close doublet, of similar molecular weight to deglycosylated native PHA. N-glycosidase treatment removes the high molecular weight diffuse bands observed in the culture supernatant on SDS-PAGE, which resolve to a single band, of higher indicated molecular weight (approximately 6 kDa) than deglycosylated native PHA (Figure 4.1, lane 8, asterisk). This band is not observed in α :PHA samples after purification and deglycosylation (compare lanes 8 and 10 of Figure 4.1), showing that it must have come from the glycopolypeptides which produce the diffuse bands on SDS-PAGE, since these do not bind to the affinity column (see above). N-terminal sequencing of PHA expressed from the α :PHA constructs after affinity purification (i.e. corresponding to Figure 4.1, lane 10) gave two N-terminal sequences, corresponding to the two closely spaced polypeptides observed in the deglycosylated recombinant protein. The lower molecular weight band of the doublet had the same N-terminal sequence as the mature, native PHA produced *in planta*, thus showing that some

correct processing had taken place. However, the higher molecular weight band of the doublet had a sequence consistent with an N-terminal extension of 9 amino acid residues derived from the C-terminal region of the α -factor pro-sequence (Figure 4.4). Thus, this polypeptide (present in comparable amount to the correctly processed form) was incompletely processed, as observed for the equivalent GNA construct described above. By analogy with the results obtained for GNA (see above) the highly glycosylated forms of PHA most probably contain the 56 amino acid N-terminal extension derived from the α -factor pro-sequence. In contrast to GNA, this incompletely processed form does not bind to carbohydrates.

4.2.3 Processing of Proteins Expressed from Constructs Containing a Modified α -Factor Prepro--Sequence Containing Glu-Ala Repeats

As the goal of this work was to produce functional, correctly processed, recombinant lectins, a second GNA construct with the addition of Glu-Ala repeats between the α -factor prepro-sequence and the GNA mature N-terminus (α EA:GNA; Figure 4.4) was prepared. The Glu-Ala repeats have been used to enhance the processing of secreted proteins when the α -factor pro-sequence is used, improving the efficiency of cleavage by the Kex2 protease; the Glu-Ala repeats are then cleaved by the Ste13 protease (Brake, 1989; Sreekrishna *et al.*, 1997). This construct produced functional GNA in the culture supernatant, which on examination by SDS-PAGE and Western blotting, proved to contain a homogeneous polypeptide of molecular weight slightly larger than 'native' GNA (Figure 4.3 (A), lanes 1 and 5). No evidence for the presence of GNA polypeptides with large N-terminal extensions was observed, nor was the protein glycosylated as judged by the lack of shift in mobility after N-glycosidase F treatment

(Figure 4.3 (A), lanes 1 and 2). However, although the inclusion of the Glu-Ala repeats clearly improved processing of the α -factor prepro--sequence, the protein still contained a four amino acid residue extension, EAEA, to the correct GNA N-terminus, showing that processing of the Glu-Ala repeats themselves was inefficient.

Similar results were obtained with a construct in which the α -factor prepro--sequence plus Glu-Ala repeats was fused to the green fluorescent protein (GFP) coding sequence (α EA:GFP). Expression in *Pichia* resulted in the production of functional protein, which was secreted into the culture supernatant. Analysis of the secreted protein by SDS-PAGE and Western blotting showed that it contained several polypeptides showing a slight size heterogeneity (Figure 4.3 (B), lane 1). In agreement with the results of gel electrophoresis, the protein had a heterogenous N-terminal sequence. The sequence data were consistent with most of the polypeptides containing either a 4 or 6 amino acid N-terminal extension (EAAA or EAEAAA, respectively), although a minor fraction of the protein did show evidence of correct processing (Figure 4.4). Thus, although the use of a modified α -factor prepro--sequence including the Glu-Ala repeats can improve processing efficiency in *Pichia*, it can be concluded that in the case of the proteins studied here the *Saccharomyces* α -factor prepro--sequence is not optimal for obtaining correctly processed protein.

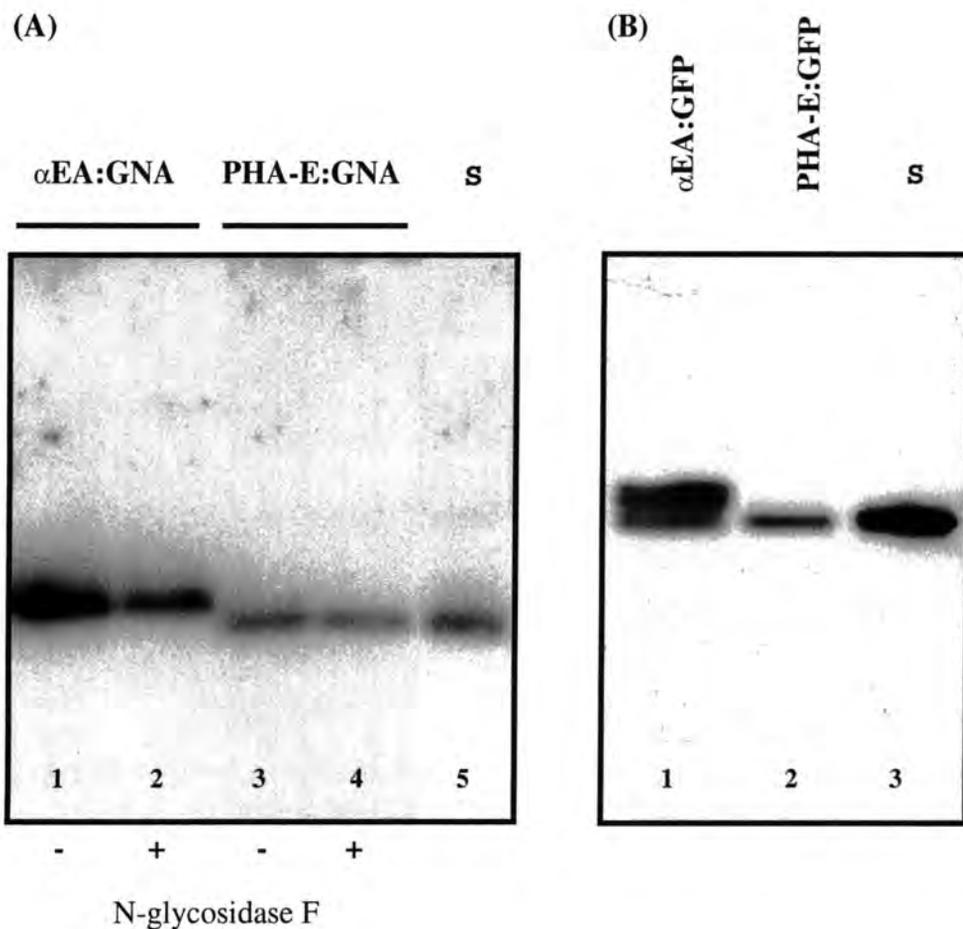


Figure 4.3 GNA and GFP expression. (A) GNA expressed in *Pichia* with Glu-Ala repeats between the α -factor prepro sequence (α EA:GNA) or using the PHA-E signal peptide (PHA-E:GNA). Immunoblot analysis of culture supernatants treated (\pm) with N-glycosidase F. (S) standard native GNA. (B) GFP expressed in *Pichia* using the α -factor prepro sequence with the Glu-Ala repeats (α EA:GFP) or using the PHA-E signal peptide (PHA-E:GFP). (S) standard recombinant GFP (rGFPuv).



Figure 4.4 Summary of processing and N-terminal sequences of lectin and GFP constructs expressed in *Pichia pastoris*. Construct designations are as for Table 1.1. A colon (:) denotes the start of mature protein coding sequences. Full arrows show the sites of processing determined by N-terminal sequencing and the residues determined are underlined. For α EA:GFP, the grey arrow denotes a minor component for which the first residue (in grey) could not be determined. Consensus N-linked glycosylation sites are shown on a grey background. A dashed grey line shows the site of cleavage of the signal peptide of the *Saccharomyces* α -factor prepro sequence. For α :PHA-E, the arrow with an asterisk (*) denotes processing by analogy with α :GNA as an N-terminal sequence was not obtained.

4.2.4 Correctly Processed Heterologous Proteins Using the PHA Signal Peptide

Since PHA could be successfully expressed in *Pichia* from constructs containing the protein's native signal peptide sequence, the processing of the protein was examined in more detail (Figure 4.1 (A)). PHA expressed from constructs containing the native signal peptide (both E- and L-forms) gave one major band and one minor band on SDS-PAGE both before, and after affinity purification; the major band was of a similar molecular weight to native PHA (compare lanes 1, 3 and 5 of Figure 4.1 (A)). When the recombinant protein was treated with N-glycosidase F, the two bands were resolved to a single deglycosylated polypeptide which migrated at a rate equivalent to deglycosylated native PHA (compare lanes 2, 4 and 6 of Figure 4.1 (A)). Finally, the N-terminal sequence of affinity purified recombinant PHA (both E- and L-forms) was determined. A single, homogeneous N-terminal sequence was present, which corresponded exactly to the N-terminal sequence of mature native PHA as purified from plant sources (Figure 4.4). The PHA isoforms expressed in *Pichia* from constructs containing the native signal peptides are thus correctly processed.

PHA expressed from the constructs containing the native signal peptides did not contain diffuse bands at a high molecular weight on SDS-PAGE, although the protein was glycosylated. In this case, the glycosylation that took place in *Pichia* resulted in carbohydrate side chains comparable in size to those observed when the protein is synthesized *in planta*. However, the composition of the carbohydrate side chains on PHA synthesized in *Pichia* is likely to differ from that of native PHA, due to differences in the respective glycosylation systems (Cregg *et al.*, 1993). PHA

synthesized in *Pichia* contains α -1,3 or 1,6-linked mannose residues, since GNA (specific for mannose residues) binds to recombinant PHA (result not shown).

The correct processing of PHA expressed from constructs containing the complete prepro-tein suggested that this signal sequence might be useful for expression of secreted proteins in *Pichia*. In the absence of a readily available alternative, the PHA signal peptide was tested to (i) obtain correctly processed GNA and (ii) direct the secretion and correct processing of an unrelated protein, in this case, green fluorescent protein (GFP). The corresponding expression constructs were (i) PHA (E-form) signal peptide fused to mature coding sequences of GNA (PHA-E:GNA) and (ii) PHA-E signal peptide fused to GFP (PHA-E:GFP) (Table 1.1 / Figure 4.4).

Expression of the PHA-E:GNA construct in *Pichia* resulted in the presence of soluble non-glycosylated protein, reactive with anti-nGNA antibodies, in the culture supernatant (Figure 4.3 (A), lanes 3 and 4). This contrasts with attempts to express GNA from constructs containing the native GNA signal peptide, where no expression could be detected. GNA produced from the PHA-E:GNA construct was functional, as shown by affinity purification through binding to a mannan-agarose column, and by haemagglutination assays of the purified protein (Table 4.1). The purified recombinant GNA contained a single N-terminal sequence, identical to the N-terminal sequence of native GNA (Figure 4.4), showing correct processing of the signal peptide.

Similar results were obtained with the PHA-E:GFP construct, where soluble functional GFP accumulated in the culture supernatant. The culture supernatant, after

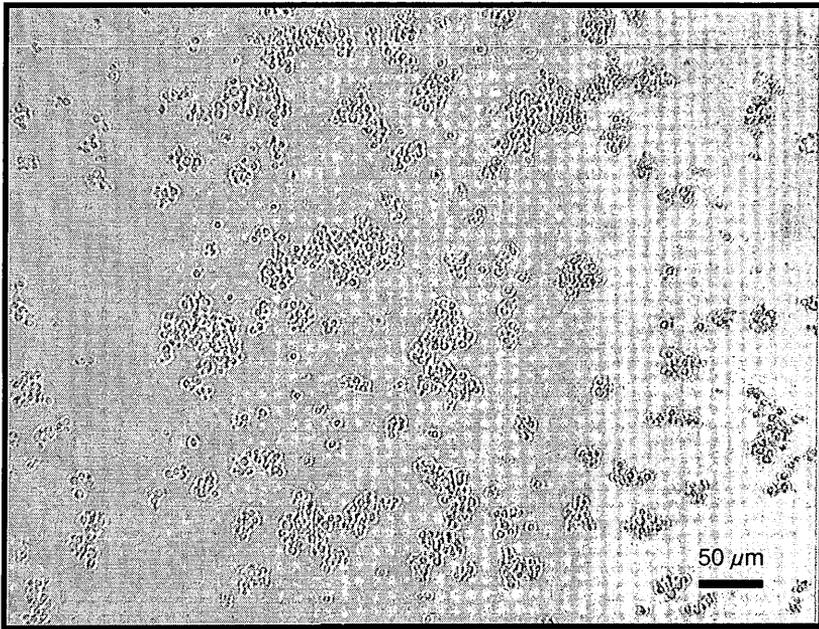
concentration by ultrafiltration, was fluorescent (with no fluorescence observed in a control supernatant), and analysis by SDS-PAGE and Western blotting with anti-GFP antibodies showed a single polypeptide migrating at a rate equivalent to a GFP standard (Figure 4.3 (B), lane 2). The GFP protein band was blotted and the N-terminal sequence corresponded to the N-terminal sequence of mature GFP (Figure 4.4), again demonstrating correct processing of the heterologous signal peptide. These results demonstrate that the PHA-E signal peptide is correctly processed from two heterologous proteins which are directed to a secretory pathway.

4.2.5 Choice of Culture Medium in GNA Binding

It was reported earlier that highly branched mannans obtained from *Pichia pastoris* (and from *Saccharomyces cerevisiae*) reacted strongly with GNA in quantitative precipitation studies (Shibuya *et al.*, 1988). Figure 4.5 shows clearly that *Pichia* cells agglutinated in the presence of native GNA. This suggests that expression of recombinant GNA using *Pichia* would inevitably lead to binding of some of the lectin to the cell surface.

The extracellular medium into which recombinant GNA is secreted plays a significant role in sensitive functional assays such as CpY-binding assay (see section 2.8.3). One of the constituents of the buffered growth/induction medium of the *Pichia* cell culture (i.e. BMGY and BMMY, see section 2.6.3) is yeast extract. Yeast extract is the water soluble portion of autolysed yeast and inevitably contains mannans. These mannans contributed to partial binding inhibition of recombinant GNA in CpY-binding assays. Even purification of the recombinant lectin does not seem to remove the mannans.

(A)



(B)

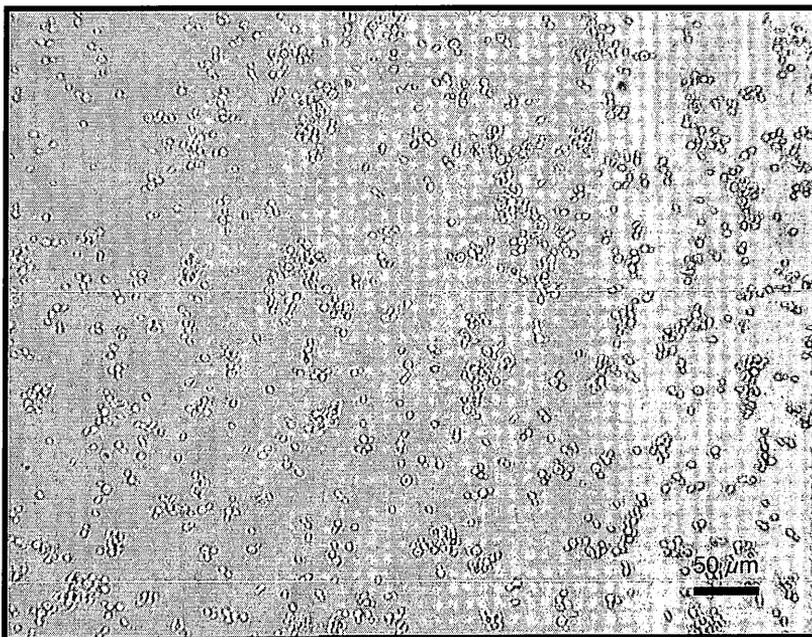


Figure 4.5 Agglutination of 10 μl *Pichia* cell suspension (7×10^8 cells / ml) in the presence (A) and absence (B) of 20 μg of native GNA. Cells were studied under a Nikon Type 104 light microscope at x 40 magnification and the image captured by JVC TK-1281 video camera.

Figure 4.6 shows that, under low concentrations of the hapten methyl- α -D-mannoside, PHA-E:GNA, affinity-purified from culture supernatant in BMMY medium, bound less readily (almost by a half) to carboxypeptidase-Y when compared to native GNA dissolved in PBS. Substituting anti-nGNA antibodies with anti-rGNA antibodies gave similar results (result not shown). BMMY culture medium containing *Pichia* spiked with native GNA prevented the lectin from binding to CpY (Fig. 4.7). Native GNA in a 1 % solution of yeast extract virtually did not bind to carboxypeptidase. Replacing yeast with beef extract resulted in a slightly better binding, but the difference was not significant; beef extract also contains interfering mannose structures. In the presence of YNB, there was no difference in binding of native GNA when compared to binding of the lectin in PBS (Fig. 4.8).

Thus, for sensitive assays, such as those based on ELISAs, which include GNA lectin, care must be taken to avoid contaminating mannans. However, for more robust assays, such as haemagglutination assays (see earlier), no differences in recombinant GNA activity were observed.

The presence of yeast extract in the culture medium also created problems when trying to ammonium sulphate precipitate in an effort to concentrate recombinant GNA. After the addition of salt, the lectin-sugar binding complex created a 'sticky' precipitate which was difficult to recover by centrifugation. Redissolving the pellet also proved difficult which resulted in a low yield of recovery. Adding 20 - 50 mM diaminopropane to the pellet was needed here to release the recombinant lectin from the sugar complex (results not shown).

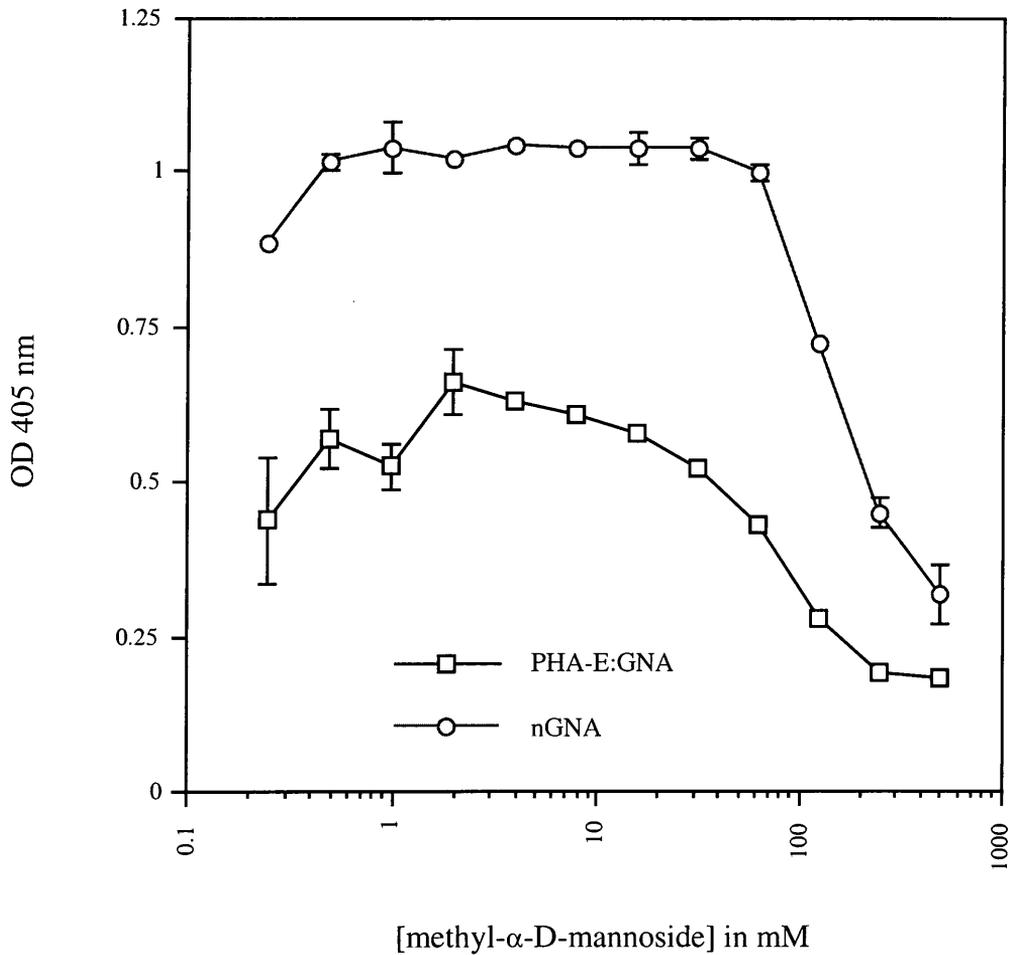


Figure 4.6 Binding of affinity-purified *Pichia*-expressed PHA-E:GNA (from BMMY medium) to carboxypeptidase-Y in the presence of varying amounts of mannose, as detected by polyclonal anti-nGNA antibodies. Native GNA (nGNA) was dissolved in PBS.

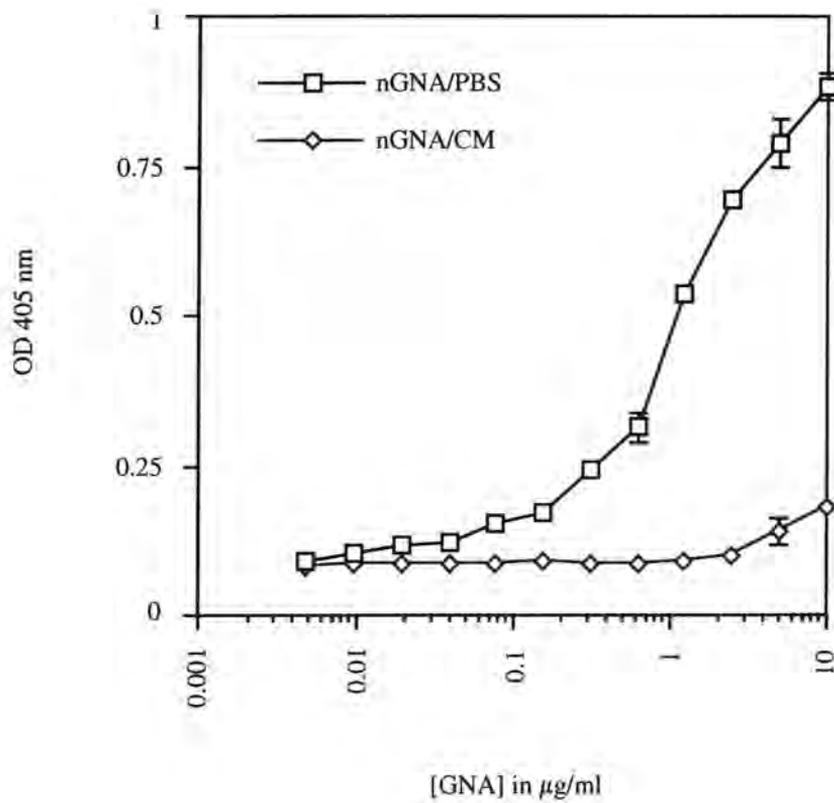


Figure 4.7 Binding to carboxypeptidase-Y of varying concentrations of native GNA dissolved in *Pichia* BMMY culture medium (nGNA/CM) or in PBS alone (nGNA/PBS).

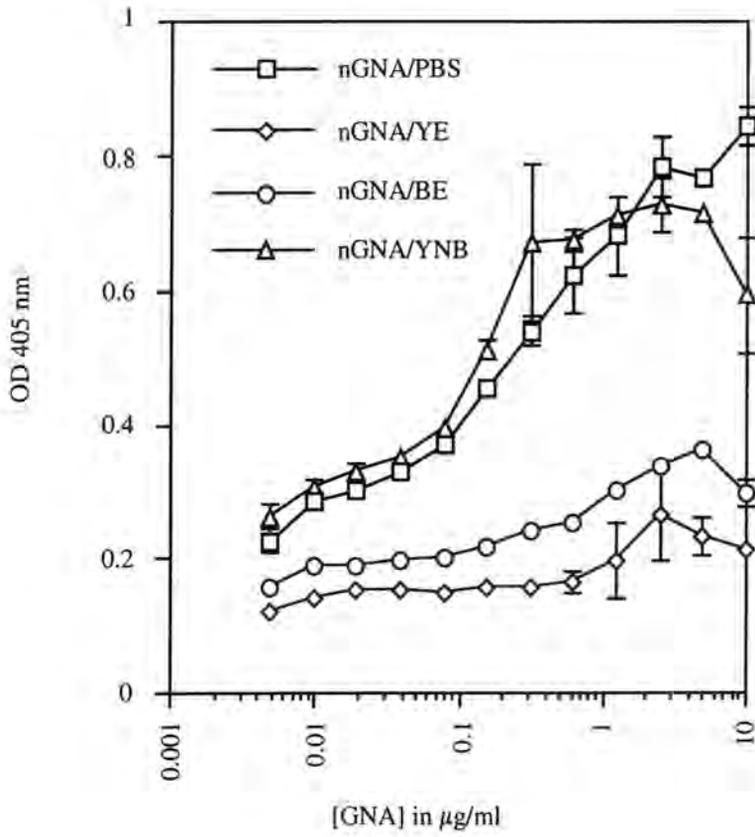


Figure 4.8 Binding to carboxypeptidase-Y of native GNA dissolved in PBS (nGNA/PBS), 1 % yeast extract (nGNA/YE), 1 % beef extract (nGNA/BE) and 1.34 % yeast nitrogen base supplement (nGNA/YNB).

An attempt to remove and separate the soluble mannans of the yeast extract from the GNA molecules by a two-step dialysis procedure (against excess (i) diaminopropane (final concentration 50 mM; pH 10 - 11) and (ii) PBS (pH 7.0)) proved futile. To circumvent this GNA-intrinsic problem, BMGY/BMMY medium can be replaced by another medium (BMGH/BMMH) lacking yeast extract supplement thus eliminating contaminating mannans. However, histidine is required as supplement since the *Pichia* strains used for expression are auxotrophic for this residue.

4.3 Discussion

To date, this is the first description of the use of *Pichia pastoris* to produce secreted, functional plant lectins which can be successfully purified from culture supernatants by one-step affinity chromatography. However, there are particular problems in expressing a mannose-specific lectin like GNA in *Pichia pastoris*, due to the presence of high-mannose carbohydrates on the surface of the cells and secreted in the medium and also due to the soluble yeast extract component of the buffered medium. As is suggested by the gel blot shown in Figure 4.2, significant amounts of GNA may be present as complexes with soluble carbohydrates in the culture supernatants. These complexes give rise to the intense smear at, and near, the origin on SDS-PAGE gels, due to incomplete denaturation, and are not retained on mannan-agarose affinity columns. GNA-carbohydrate complexes can be precipitated from the culture medium by increasing the pH, but the protein cannot be recovered from the precipitate without denaturation (result not presented). Thus, although GNA can be purified from culture supernatants, further research will be necessary to optimise the recovery of active GNA from *Pichia*. Interestingly, α :GNA species containing most of the pro-

sequence were present in the protein after affinity purification on the mannan-agarose column, showing that GNA was still functional even with a highly glycosylated N-terminal extension. In contrast, PHA molecules containing the glycosylated N-terminal extension derived from the α -factor pro-sequence were not retained on an affinity column. This may be due to the carbohydrate side chains interfering with the sugar-binding site in PHA, or altering the structure of the molecule to abolish its activity, or blocking access to the binding site.

N-linked oligosaccharide side chains produced on glycoproteins expressed in *Pichia* have been shown to be mainly of the 'high-mannose' type, where a core structure of NAcGlu-(Man)₂ is modified by the addition of (mannose)_n branches to the terminal mannose residues. However, the vast majority of these oligosaccharides in *Pichia* contain 8 - 14 mannose residues, and thus would be expected to increase the molecular weight of a glycoprotein by approximately 2 kDa (Cregg *et al.*, 1993). The decrease in molecular weight observed when PHA expressed in *Pichia* from the PHA-E:PHA-E construct is deglycosylated is consistent with the molecule containing 1 or 2 carbohydrate side chains of this type, and it was also observed that GNA binds strongly to PHA expressed in *Pichia* (result not shown). The mature PHA-E sequence contains three potential N-linked glycosylation sites, although only two of these are utilised *in planta* (Sturm & Chrispeels, 1986). The glycosylation that is present on both PHA and GNA expressed in *Pichia* from constructs containing the α -factor prepro--sequence (α :PHA-E and α :GNA) is different in nature, and appears to involve long carbohydrate side chains resulting from glycosylation on the uncleaved α -factor pro-sequence. This is likely to be a result of inefficient processing which has

previously been shown to result in the secretion of hyperglycosylated unprocessed proteins in *Saccharomyces* ((Kjeldsen *et al.*, 1998) and references therein).

While the yields of lectins obtained from *Pichia* cultures in this work were fairly low, no attempt was made to optimise yields by selection of clones with multiple inserts (Clare *et al.*, 1991), or to optimise culture conditions. Consequently, the potential yields of recombinant lectins produced in *Pichia* are likely to be at least an order of magnitude greater than those reported here as, after optimisation, yields of > 1 g / l for HIV-1 gp120 (Scorer *et al.*, 1993) and > 10 g / l for a tetanus toxin fragment (Clare *et al.*, 1991) have been reported.

Many heterologous signal peptides do not function in yeast, either to direct protein secretion or to generate the 'natural' N-terminus of the protein (reviewed in (Tuite, 1991)). Early reports (eg. (Zsebo *et al.*, 1986)) demonstrated the secretion and correct processing of proteins when fused to the α -factor prepro--sequence. As a consequence, the α -factor prepro--sequence has been widely used in the synthesis of recombinant proteins in *Saccharomyces* and in *Pichia*, where, as with *Saccharomyces*, it has been shown to direct secretion and correct N-terminal processing in a number of examples (Hoffman & Donaldson, 1985). Processing of the α -factor prepro--sequence in *Saccharomyces* involves several distinct proteolytic cleavage steps, carried out by different enzymes. The 19 amino acid signal peptide (pre-sequence) is removed by a signal peptidase system; further processing of the pro-sequence then involves the action of an endopeptidase encoded by *KEX2* gene, which cleaves C-terminally to a specific Lys-Arg sequence, and a dipeptidyl aminopeptidase encoded by the *STE13* gene, which removes N-terminal Glu-Ala repeats (Brake,

1989). The presence of the Glu-Ala repeats enhances the activity of the Kex2 protease, but subsequent processing of these repeats by the Ste13 aminopeptidase has been found to be inefficient in many cases (Tuite, 1991).

The correct processing of the α -factor prepro--sequence is very much case-dependent and numerous examples have been reported in both *Saccharomyces* and *Pichia* where proteins have failed to give correct processing to yield mature N-termini (Sreekrishna *et al.*, 1997). In *Pichia*, the pro-peptide is not removed from protein disulphide isomerase and remains, hyperglycosylated, on a proportion of recombinant protein (Vuorela *et al.*, 1997); the propeptide is only partially processed from human procarboxypeptidase A2, resulting in heterogeneity (Reverter *et al.*, 1998); expression of influenza neuraminidase (in *Saccharomyces*), or the β subunit of bovine follicle-stimulating hormone, gave products where the Glu-Ala repeats were not removed (Martinet *et al.*, 1997; Samaddar *et al.*, 1997). The results obtained in this chapter agree with and extend these previous observations. Both PHA and GNA show inefficient processing by the *Pichia* equivalent of the Kex2 protease. Addition of Glu-Ala repeats to the α :GNA construct resulted in efficient cleavage by this enzyme but the products of both the α EA:GNA and α EA:GFP constructs are inefficiently processed by the *Pichia* equivalent of the Ste13 aminopeptidase, resulting in proteins with N-terminal extensions.

Given the problems sometimes associated with the α -factor prepro--sequence, efforts have been made to develop alternative signal peptides for use in *Pichia* (Sreekrishna *et al.*, 1997); for example, the *Pichia* acid phosphatase *PHO1* signal peptide. This signal peptide fused to mature protein sequences can lead to secretion of either

correctly processed protein (Weiss *et al.*, 1995), or protein containing N-terminal heterogeneity (O'Donohue *et al.*, 1996), again showing case-dependency. A hybrid PHO1 signal peptide, containing a Kex2 cleavage site to improve processing efficiency, has also been used (Laroche *et al.*, 1994). Some prepro-teins expressed in *Pichia* produce secreted and correctly processed protein (Juge *et al.*, 1996; Ferrari *et al.*, 1997; Sreekrishna *et al.*, 1997), but it is not known if the corresponding signal peptides can confer correct processing on heterologous proteins. This work describes such an analysis and demonstrates the secretion and correct N-terminal processing of two unrelated proteins (GNA and GFP) using the PHA-E signal peptide. This is unexpected in view of the failure of this signal peptide to function correctly in *Saccharomyces* (Tague & Chrispeels, 1987), whether this reflects a difference in signal peptide recognition between *Saccharomyces* and *Pichia* or some other factor(s) remains to be determined. The amount of protein accumulating in the culture supernatant appears to be slightly less with constructs using the PHA-E signal peptide compared to equivalent constructs using the α -factor prepro--sequence (unpublished observations). Notwithstanding, these results suggest the PHA-E signal peptide could be used for proteins where secretion and processing using other signal peptides has proven problematic.

Chapter 5

Use of *P. pastoris* to Express Functional GNA-GFP Fusion Protein

5.1 Introduction

An important application of some plant lectins in biotechnology is the expression of lectin genes to confer insect resistance in transgenic crop plants. One prime example of such an insecticidal lectin is the snowdrop lectin GNA. This lectin consists only of mannose-binding domains, so its toxicity towards insects may be dependent upon the recognition of and binding to mannose-containing structures (Powell *et al.*, 1995a, 1995b; Du *et al.*, 2000).

Chimerolectins, i.e. chimeras which are built up of a carbohydrate-binding domain arrayed in tandem with an independent unrelated domain, for example a catalytic domain, already exist in nature. For example, Class I plant chitinases are composed of an N-terminal chitin binding domain linked to a catalytic domain (Collinge *et al.*, 1993). However, since this protein only has one binding domain it does not have the ability to agglutinate cells. Type 2 ribosome-inactivating proteins (RIPs), such as ricin and abrin, are fusion proteins of an N-terminal toxic A chain, which has the N-glycosidase activity characteristic of all RIP, and a C-terminal carbohydrate-binding B-chain (Barbieri *et al.*, 1993). Type2 RIPs have two binding sites on each B-chain and therefore agglutinate cells.

In this work, the fluorescent GFP was fused to GNA so as to allow direct visualisation of lectin binding. The previous chapter (Chapter 4) described the development of an heterologous expression system based on *Pichia pastoris* for the expression of functional and correctly processed plant lectins and the green fluorescent protein GFP. Utilising this system, expression of functional chimera GNA-GFP and secretion into the culture medium is demonstrated. Also, the dual activities of the lectin and reporter moieties are demonstrated in independent assays. To complement on-going studies on the mode of action of GNA, the binding of the GNA-GFP chimera to cells in the haemolymph of the tomato moth (*Lacanobia oleracea*) was investigated *in vitro* as well as *in vivo* by confocal microscopy.

This is the first report of a functional recombinant GNA chimera consisting of another unrelated protein. The chimera is a prerequisite to creating more practical fusion proteins in which the GFP protein may be substituted by other insecticidal domains, for example, Bt Cry domains. In this way, the chimera may be more effective towards controlling certain insect pests when expressed from transgenic plants.

5.2 Results

5.2.1 Expression of GNA-GFP in *P. pastoris*

Initially, two constructs were prepared, α :GNA-GFP and PHA-E:GNA-GFP, in a triple ligation reaction as described in Materials and Methods (Chapter 2.5.6). Constructs containing GFP alone (i.e. α :GFP and PHA-E:GFP; see Chapter 2.5.5) were used for comparative purposes. For each construct, 50 *Pichia* strain GS115

transformants were randomly selected, picked and transferred to a gridded plate containing minimal media MMH. After three to four days of methanol induction, *Pichia* colonies expressing GFP held under a UV transilluminator were shown to fluoresce at different intensities, with some colonies not fluorescing at all (results not shown). Generally, those that expressed GFP alone were brighter than those that expressed the GNA-GFP chimera. In both cases, the colonies that contained the PHA-E leader constructs had a higher proportion of fluorescent colonies than those that contained the construct with the α -factor prepro--sequence (12 % α :GFP and 28 % PHA-E:GFP; 8 % α :GNA-GFP and 16 % PHA-E:GNA-GFP).

At the cellular level, location of expressed recombinant GNA-GFP and GFP in the *Pichia* cell was observed to be similar: accumulation occurred in isolated area(s) within the cytoplasm, presumably the vacuole (Fig. 5.1). Generally, the localised fluorescence of rGFP was more intense than that of the fusion. However, upon closer inspection of the yeast cell, it appeared that in those cells that contained the GNA-GFP constructs (i.e. α :GNA-GFP (Fig. 5.1(D)) and PHA-E:GNA-GFP (Fig. 5.1(H))) a significantly higher degree of fluorescence on the periphery of the cells was shown than in those cells that only harboured the GFP constructs (i.e. α :GFP (Fig. 5.1(B)) and PHA-E:GFP (Fig. 5.1(F))). This finding is most likely the first indication of dual functionality of the GNA-GFP fusion: binding of the GNA moiety to mannan structures in the cell wall and simultaneous fluorescence of the GFP moiety.

Pichia containing constructs α :GNA-GFP and PHA-E:GNA-GFP were grown and induced in BMGH/BMMH culture medium. This medium, which was free of the yeast extract component found as an ingredient in BMGY/BMMY medium,

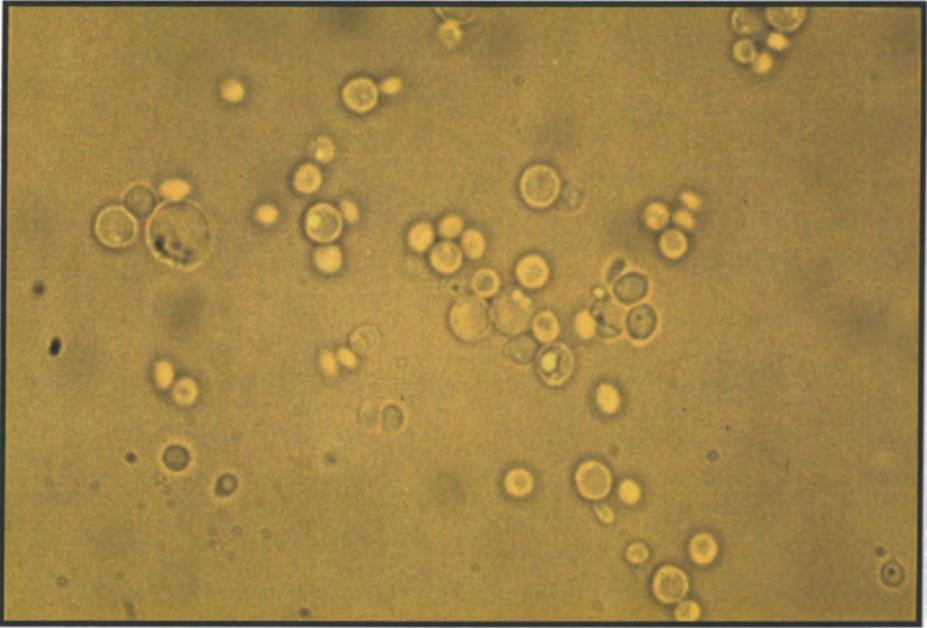
circumvented the problem of undesired GNA binding to the soluble mannans (see Chapter 4.2.5). The supernatants and cell extracts from methanol-induced cultures were analysed for the presence of the chimeric product.

Figures 5.2 & 5.3 summarise the results for the expression of GNA-GFP in *Pichia pastoris*. Proteins which reacted with anti-GNA and/or anti-GFP antibodies were present in supernatants and cell extracts of induced cultures for the GNA-GFP constructs, both with the α -factor prepro--sequence (α :GNA-GFP) and the PHA-E signal peptide (PHA-E:GNA-GFP).

For both constructs, analysis by SDS-PAGE and Western blotting of the supernatants using both anti-GNA and -GFP antibodies showed several polypeptides, including the expected polypeptide representing the GNA-GFP fusion product of molecular weight approximately 40 kDa. Anti-GNA antibodies revealed the presence of two extra low molecular weight bands slightly larger (between 14 and 20 kDa) than the native GNA monomer (lanes C & E, Fig. 5.2 (A)). Apart from the fusion product band, anti-GFP antibodies also picked out a second band of similar mobility to that of standard rGFPuv (lanes C & E, Fig. 5.2 (B)) and recombinant GFP (lanes B & D, Fig. 5.2 (B)).

These results suggest that the GNA-GFP chimera had been partially cleaved presumably in the linker region or at the immediate flanking regions during the synthesis and secretion process since the truncated bands detected on the anti-GNA blot were not repeated on the anti-GFP blot. An attempt to reduce proteolysis by mutation of a suspected lysine basic residue (possible cleavage point for trypsin-like enzymes) to threonine in the linker region (Thr-Leu-Glu-Asp-Pro-Arg-Val-Pro-Val-

(A)



(B)

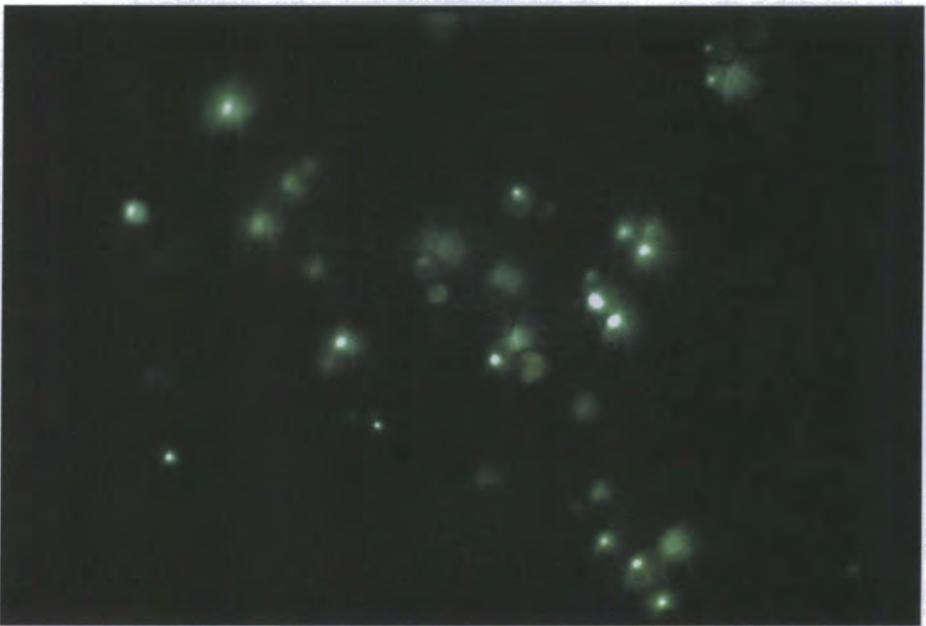
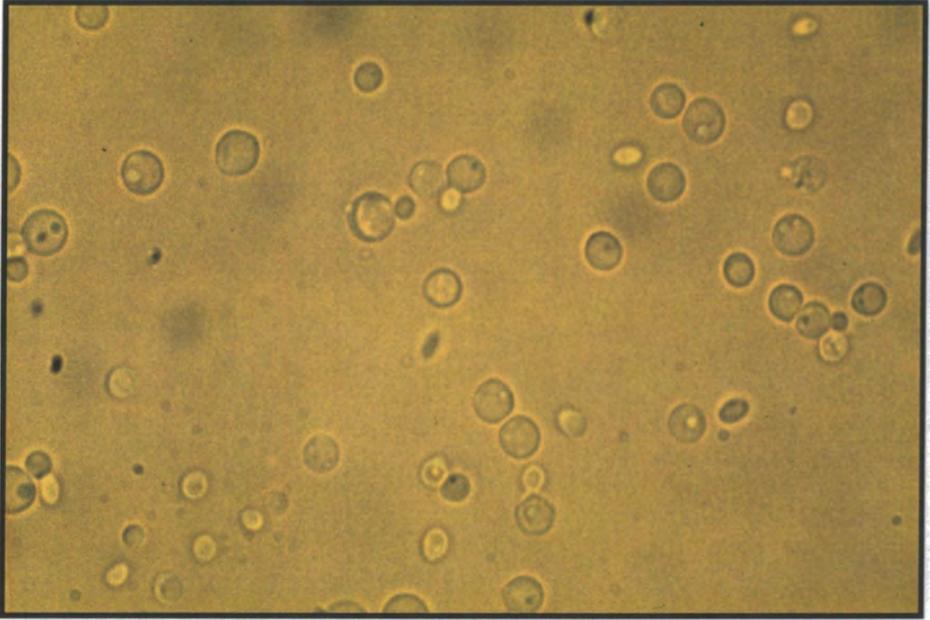


Figure 5.1 Distribution of rGFP and GNA-GFP in *Pichia* GS115 cells. The cells were first grown in BMGY medium overnight and then induced in BMMY medium for 24 hours. The cells were studied by light microscopy and photographed through Normasky optics (A, C, E, G) and fluorescent microscopy (B, D, F, H). (A,B) α :GFP, (C,D) α :GNA-GFP, (E,F) PHA-E:GFP (G,H) PHA-E:GNA-GFP.

(C)



(D)

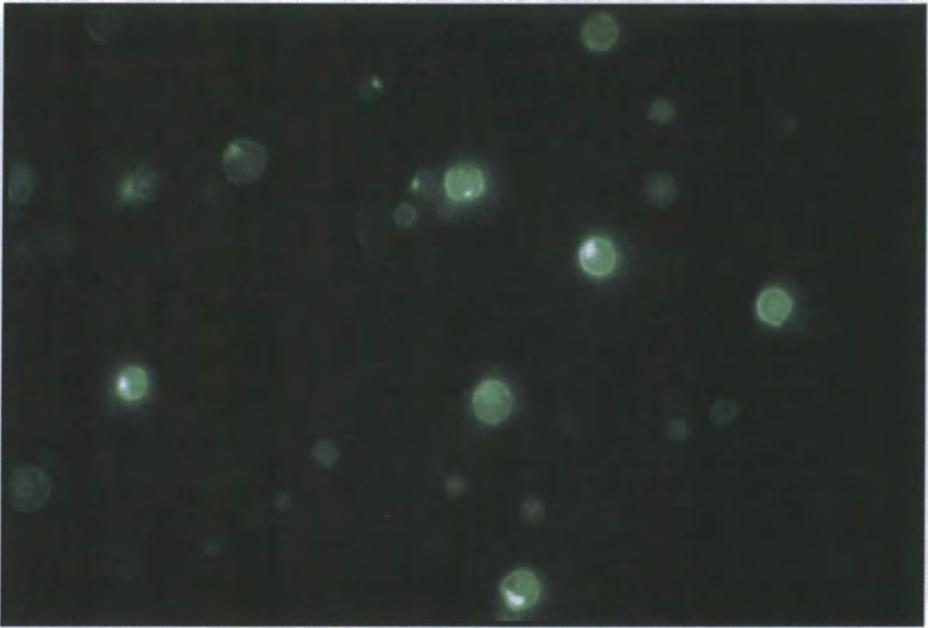
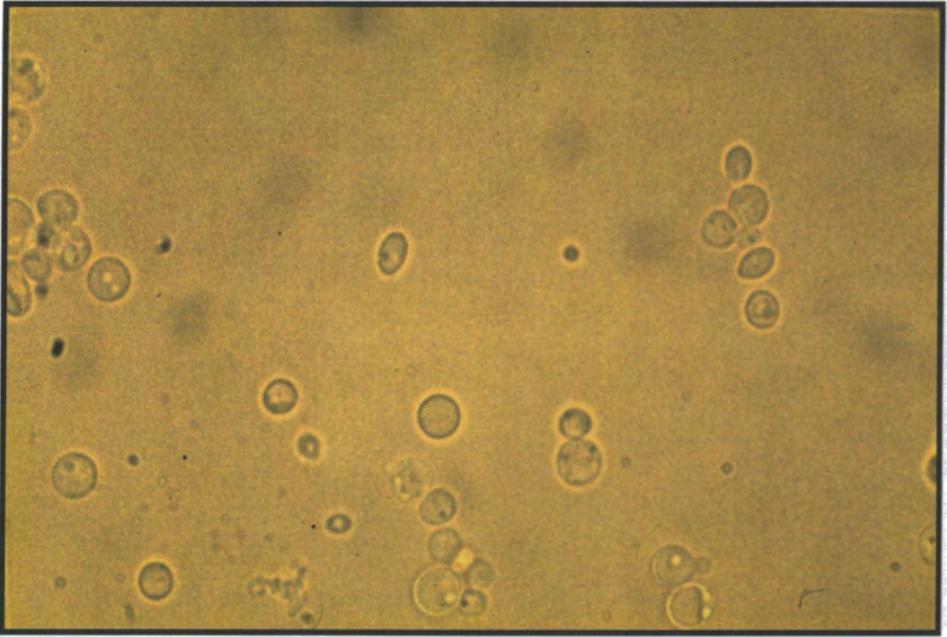


Figure 5.1 continued.

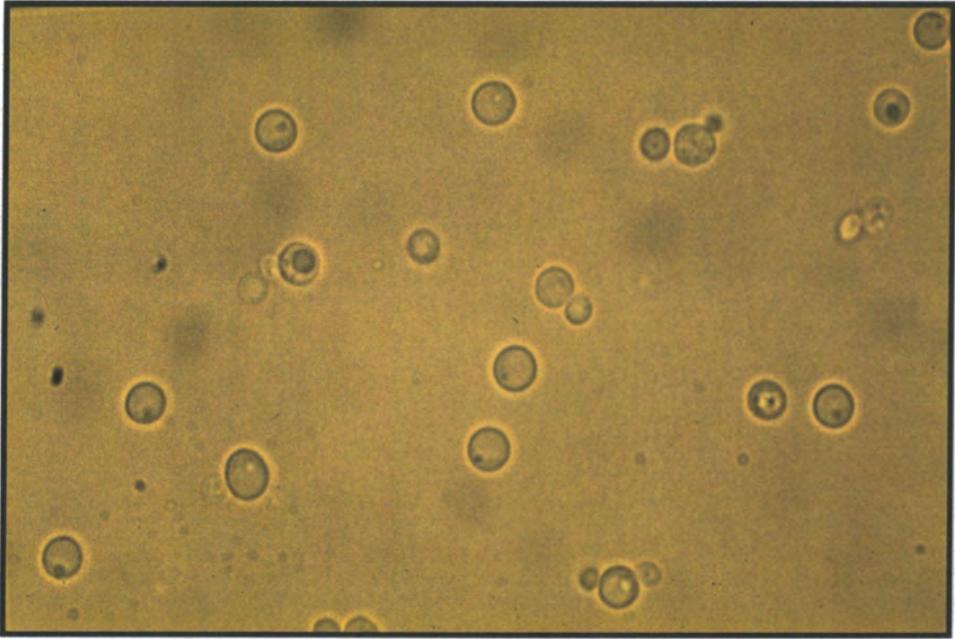
(E)



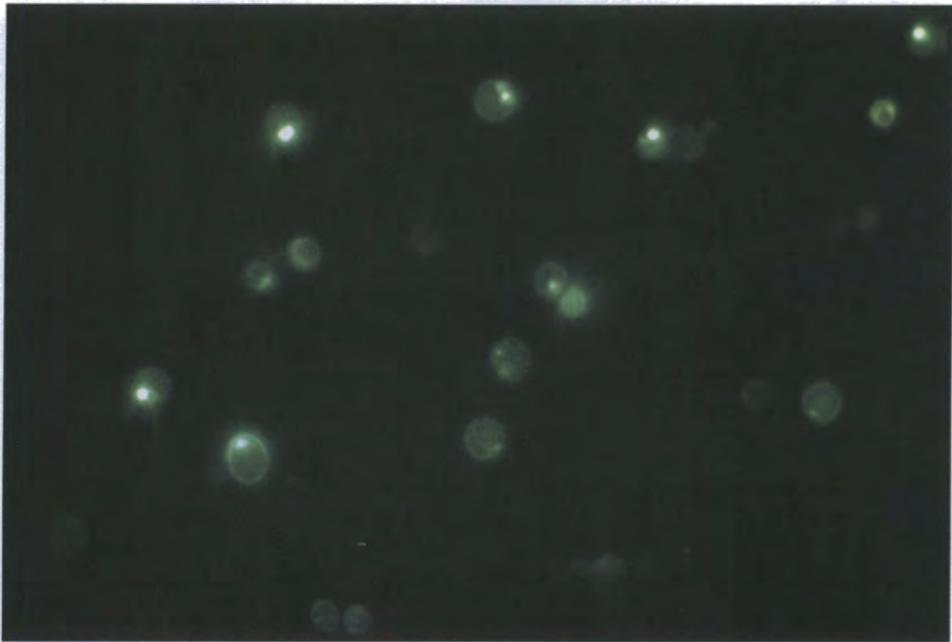
(F)

Figure 5.1 continued.

(G)



(H)

Figure 5.1 continued.

Glu-Lys-Met) did not produce any convincing positive changes in the polypeptide patterns (results not shown). Also, the addition of yeast extract did not prevent partial proteolytic cutting (result not shown).

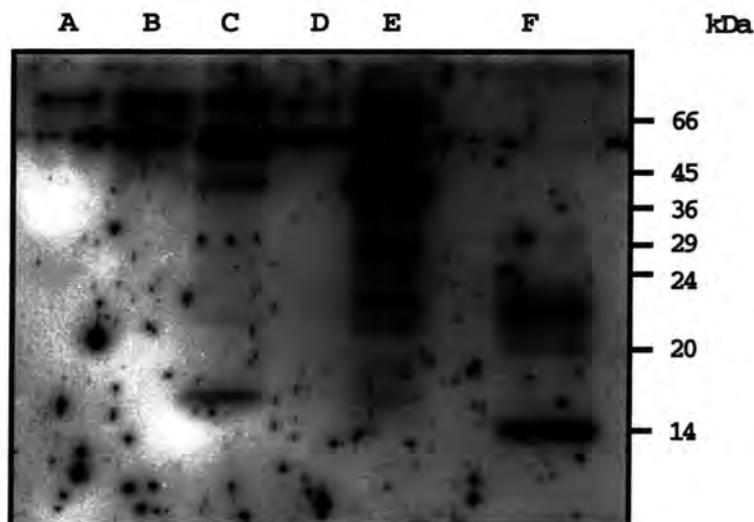
Surprisingly, the polypeptide band profiles on the Western blots of supernatants for both α :GNA-GFP and PHA-E:GNA-GFP were quite similar. However, there were differences in the relative amounts of fusion protein and cleaved polypeptides between the two constructs. For construct α :GNA-GFP, the fusion protein was slightly more abundant than the individual cleaved products, although the difference was not great. In the case of PHA-E:GNA-GFP, there was less fusion product than its truncated products, as shown clearly in lane E of Figure 5.2 (B).

Analysis by SDS-PAGE and immunoblotting of *Pichia* cell extracts for both constructs showed that the number of bands recognised by anti-GNA and -GFP antibodies was greater than that obtained from blots of culture supernatant (Fig. 5.3). A repertoire of bands ranging between 14 and 45 kDa, as indicated by the molecular weight markers, could be detected by anti-GNA antibodies (lanes C & E, Fig. 5.3 (A)). For both constructs, the presence of the GNA-GFP chimera was verified by the two antibodies. Some of the bands recognised by anti-GNA antibodies overlapped with those that were detected by anti-GFP antibodies. There were no bands with a molecular weight lower than or similar to that of 'native' GNA monomer that could be detected by the anti-lectin antibodies. This suggests that proteolytic cleavage occurred not only within the linker but also within the amino terminal of the GFP moiety. Polypeptide bands with a much higher molecular weight than the fusion protein band (lane E, Fig. 5.3 (B)) are indicative of the presence of lectin and fusion products

complexed with mannose-containing components from *Pichia pastoris*. The prominent band found between the 45 and 66 kDa molecular weight markers in lane C of both blots (Fig. 5.3 (A) & (B)) could be due to incorrect amino terminal processing of the α -factor prepro- leader sequence. Recombinant GFP produced from construct α :GFP (lane B, Fig. 5.3 (B)) resulted in a number of bands some of which correspond to the standard rGFPuv (lane F, Fig. 5.3 (B)). This result is due to incomplete processing of the α -factor prepro- leader sequence with glycosylation as a result (see Chapter 4). Recombinant GFP from construct PHA-E:GFP produced a band profile similar to that of standard rGFPuv. Judging from the blots, the concentration ratios of GNA-GFP fusion product to its cleaved products for both constructs were the opposite from, and more pronounced than, those observed for the culture supernatant. Anti-GNA and -GFP antibodies revealed that for construct PHA-E:GNA-GFP the fusion protein was much more abundant than the cleavage products (75 % and 25 %, respectively) (lane E, Fig. 5.3 (A) & (B)). For construct α :GNA-GFP, more than 50 % of heterologous expressed product that was estimated to be formed resulted in the individual GNA and GFP cleavage products (lane C, Fig. 5.3 (A) & (B)).

Estimating the concentration of heterologous protein synthesized and secreted into the culture supernatant or directed to cell compartments was carried out by comparing the band intensity corresponding to the fusion protein with those of the GNA and GFP standards. The densities of the cell cultures were normalised. For construct α :GNA-GFP, the cells harboured 0.02 mg of fusion protein per litre culture as opposed to 0.1 mg / l culture of the protein found in the supernatant. For construct PHA-E:GNA-GFP, 0.2 mg of GNA-GFP per litre culture was found in the cells, whereas only 0.01

(A)



(B)

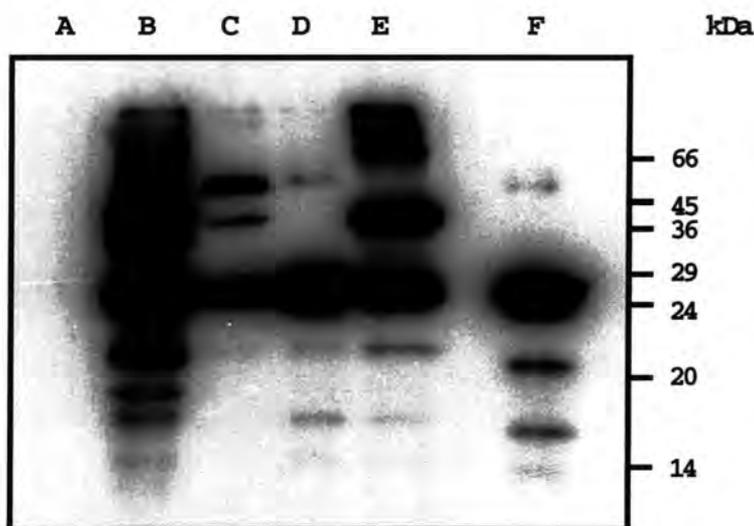


Figure 5.3 Immunoblot analysis using polyclonal anti-GNA antiserum (A) or monoclonal anti-GFP antibodies (B) of *Pichia* cell extracts containing GFP and GNA-GFP. (A) Lane A: control (pPICZB); lane B: α :GFP; lane C: α :GNA-GFP; lane D: PHA-E:GFP; lane E: PHA-E:GNA-GFP; lane F: 10 ng standard native GNA. (B) same as (A) except for lane F: standard rGFPuv. All samples were adjusted by volume to maintain equal cell densities as determined by OD measurements at 600nm. One ml of cell culture was centrifuged and the cells resuspended in 130 μ l of cell lysis buffer before cell disruption. Samples α :GFP and PHA-E:GFP were diluted by a factor of 5. Ten μ l of sample plus an equal volume of 2x SDS sample buffer was subjected to 10 % β -mercaptoethanol and boiling for 10 minutes prior to loading onto a 15 % SDS-PAGE.

mg / l culture of the product was all that could be detected in the supernatant. Inevitably, the fusion protein concentration in the cell extract takes into account the GNA-GFP which bound to the cell wall.

Recombinant GFP expressed from construct α :GFP was secreted into the culture supernatant at a concentration of 5 mg / l culture, whereas the remainder of the protein was kept in the cell at a concentration of 15 mg / l culture. As for rGFP expressed from PHA-E:GFP, only 0.5 mg / l was secreted and 5 mg / l was detected in the cell.

5.2.2 Purification of GNA-GFP

PHA-E:GNA-GFP was purified from culture (grown and induced in BMGH/BMMH medium) supernatant by affinity chromatography using a mannan-agarose column. The proteins in the culture supernatant were precipitated by $(\text{NH}_4)_2\text{SO}_4$ at 80 % saturation and redissolved in PBS prior to loading the column. Figure 5.4 shows the SDS-PAGE and Western blot analysis of the load, wash and eluted fractions using both anti-GNA and -GFP antibodies. GNA-GFP bound to and could be eluted from the affinity column. However, as shown in lane D of Figure 5.4 (A), anti-GNA antibodies also detected in the eluted fraction the two main smaller bands that were the products of proteolytic cleavage. These bands, which were not detected by anti-GFP antibodies (lane D, Fig. 5.4 (B)), represent GNA with an extended C-terminal region encompassing part or whole of the 'linker' region. The blot showing the wash fraction (lane C, Fig. 5.4 (B)) contained a single band with a similar mobility as rGFPuv standard and which was only recognised by anti-GFP. Both the wash and the

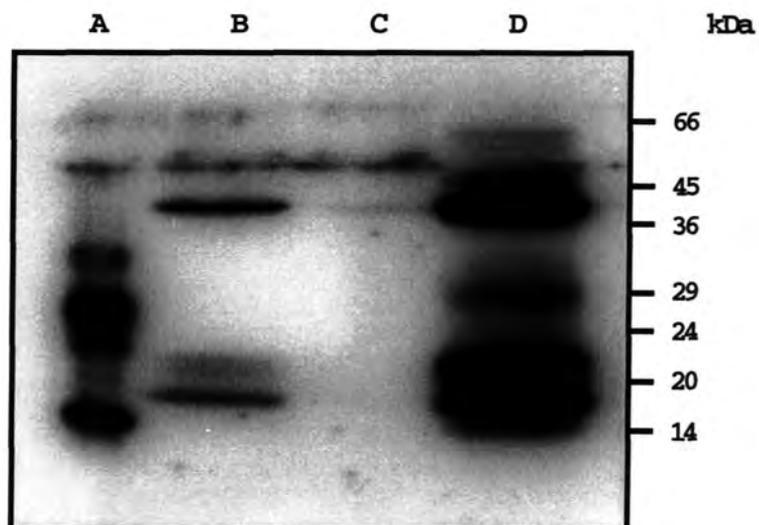
eluted fractions fluoresced green under UV light (result not shown). Thus, the fraction containing the GNA-GFP in elution buffer 50 mM DAP with a pH value of approx. 11 did not seem to have a detrimental effect on GFP fluorescence activity. The yield of purified GNA-GFP chimeric product was < 0.5 mg / l culture.

As these results show, it is clear that the co-purification of the GNA moieties with GNA-GFP did not hamper the binding of the fusion protein to the affinity column. In fact, the presence of GNA monomers may instead prove to be beneficial in relation to the functional activity of GNA-GFP since formation of tetramers seems more sterically plausible with a mixture of small and large molecules than with large molecules alone. Immunoblot analysis of samples from *Pichia* culture supernatants containing secreted GNA-GFP that were not subjected to heat treatment and addition of β -mercaptoethanol prior to loading onto an SDS-PAGE gel (Fig. 5.5) showed a reduction in the amounts of GNA and GNA-GFP monomers and the appearance of high molecular weight (> 45 kDa) fusion complexes possibly due to oligomer assembly (lane B). This pattern clearly differs from that of the treated samples as shown in lane A.

5.2.3 Functional Activities of Recombinant GNA-GFP

The ability of GNA-GFP to bind to and elute off the mannan column, albeit in the presence of cleaved GNA products, is evidence of GNA's functional carbohydrate-binding role of the fusion protein. In turn, the GFP moiety of GNA-GFP was shown to fluoresce, but only once eluted; because GNA-GFP was not concentrated enough fluorescence could not be detected under UV light while still bound to the column matrix. A hint of synchronous activities of the GNA and GFP moieties was shown by

(A)



(B)

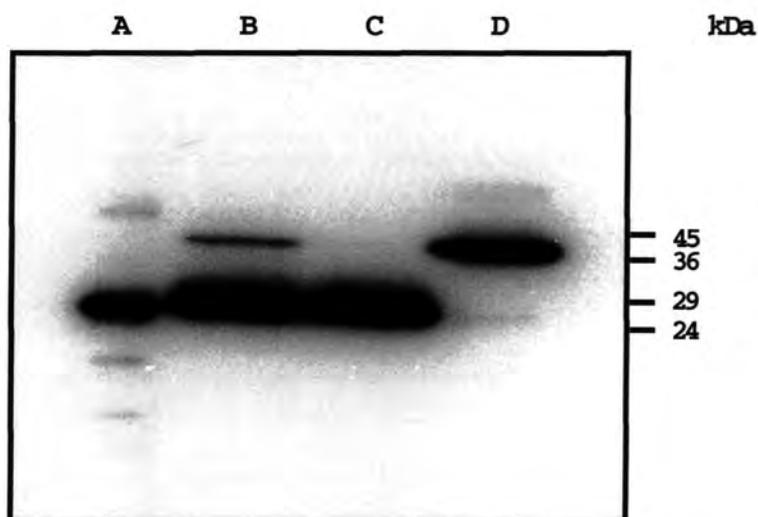


Figure 5.4 Immunoblot analysis using polyclonal anti-GNA antiserum (A) or monoclonal anti-GFP antibodies (B) of affinity-purified rGNA-GFP. (A) Lane A: 25 ng standard native GNA ; lane B: load fraction; lane C: wash fraction; lane D: eluted fraction. (B) Lane A: 25 ng of standard rGFPuv; lanes B-D: same as in (A).

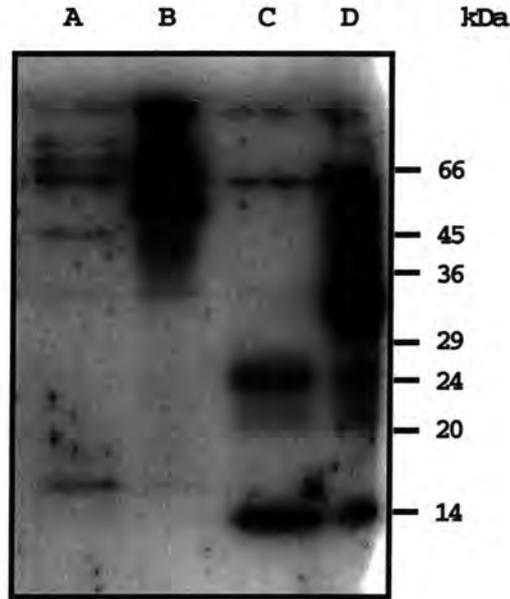
the fluorescence of bound GNA-GFP to the surface of *Pichia* cells expressing the chimera from constructs containing either the α -factor prepro--sequence or the PHA-E leader sequence (Fig. 5.1). However, these results alone were not sufficiently convincing, and, therefore, more independent functional assays were carried out to further characterise the binding activity of GNA-GFP fusion protein more fully.

Sufficient GNA-GFP fusion protein for the assays was purified from a 2 l *Pichia* culture supernatant. The elution peak profile together with the concentration ratios of GNA-GFP to GNA in the eluted fractions (2-6) are shown in Figure 5.6. The concentrations were determined by Bradford assay (GNA-GFP plus GNA) and by fluorometric measurement (GNA-GFP only). For the latter, it was assumed that the fluorescence of the GFP moiety of the fusion protein resembled that of the rGFPuv standard.

5.2.3.1 Haemagglutination Activity of Recombinant GNA-GFP

Both fractions 4 and 5 of affinity-purified recombinant GNA-GFP completely agglutinated untreated rabbit erythrocytes obtained from CSL (York) at a minimum concentration of 20 $\mu\text{g} / \text{ml}$ (Fig. 5.7). At this concentration, fraction 3 did not agglutinate whereas fractions 2 and 6 showed signs of incomplete haemagglutination. Native GNA and rGNA109 required a minimum of 2.5 and 1.25 $\mu\text{g} / \text{ml}$, respectively, for complete agglutination to occur. GNA will agglutinate cells most effectively when present as tetrameric molecules. The low agglutination activity of the GNA-GFP fusion protein may be a result of an inability to form homogenous tetramers, because steric hinderance by the GFP moieties would prevent this from happening. If so, either 1 or 2 GNA-GFP molecules could associate with 3 or 2 GNA molecules,

(A)



(B)

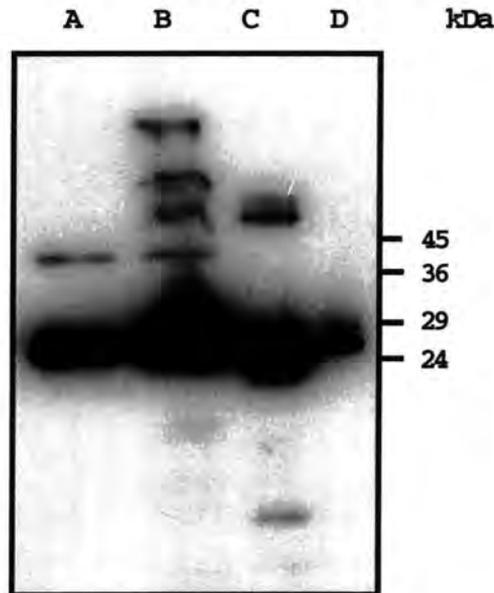


Figure 5.5 Immunoblot analysis of culture supernatant containing secreted GNA-GFP using polyclonal anti-GNA antiserum (A) or monoclonal anti-GFP antibodies (B) of treated (10 % β -mercaptoethanol and boiling for 10 minutes) and untreated (no reduction and no heat) GNA-GFP. (A) Lane A: treated GNA-GFP; lane B: untreated GNA-GFP; lane C and D: 10 ng treated and untreated standard native GNA, respectively. (B) Lanes A and B: same as in (A); lanes C and D: 10 ng standard rGFPuv treated and untreated, respectively.

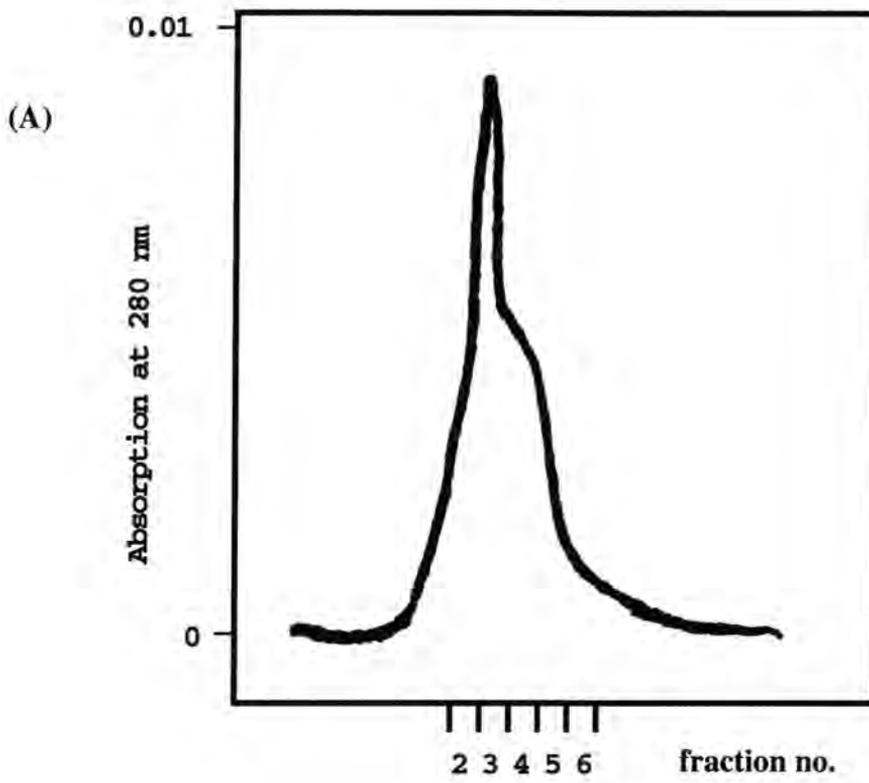
respectively. If the former holds, then $13 \mu\text{g} / \text{ml}$ ($[(12.5 \text{ kDa} \times 4 \text{ GNAs}) / [(12.5 \text{ kDa} \times 4 \text{ GNAs}) + (27 \text{ kDa} \times 1 \text{ GFP})] \times 20 \mu\text{g} / \text{ml})$, and if the latter holds, then $10 \mu\text{g} / \text{ml}$ ($[(12.5 \text{ kDa} \times 4 \text{ GNAs}) / [(12.5 \text{ kDa} \times 4 \text{ GNAs}) + (27 \text{ kDa} \times 2 \text{ GFPs})] \times 20 \mu\text{g} / \text{ml})$, would be the molar agglutination concentrations predicted for haemagglutination activity. These are not similar to the minimum agglutination concentrations of native and recombinant lectin.

5.2.3.2 Binding of Recombinant GNA-GFP to Mannan-Agarose Beads

A simple experiment to verify the dual activities of GNA-GFP fusion protein was to mix an equal volume of mannan-agarose bead suspension (also used for affinity purification) to $50 \mu\text{g} / \text{ml}$ of purified GNA-GFP (fraction 4). With the aid of a fluorescent microscope, it was observed that the beads were fluorescing brightly (Fig. 5.8). This was indicative of GFP's activity while, at the same time, GNA was holding the chimera to the mannan structures. In the presence of rGFPuv standard alone and at the same concentration, the beads were not bright green. Instead, the buffer surrounding the cells was found to be fluorescing slightly. A similar result was observed when mannose was added at a final concentration of 250 mM to the mannan beads in the presence of GNA-GFP.

5.2.3.3 Binding of Recombinant GNA-GFP to Haemocytes *in vitro*

In vitro analysis by confocal microscopy of GNA-GFP binding to *Lacania oleracea* cells was investigated using two approaches. The first approach involved direct incubation of GNA-GFP with fresh haemolymph, whereas the second, more controlled approach involved probing of prepared immobile cell monolayers with



(B)

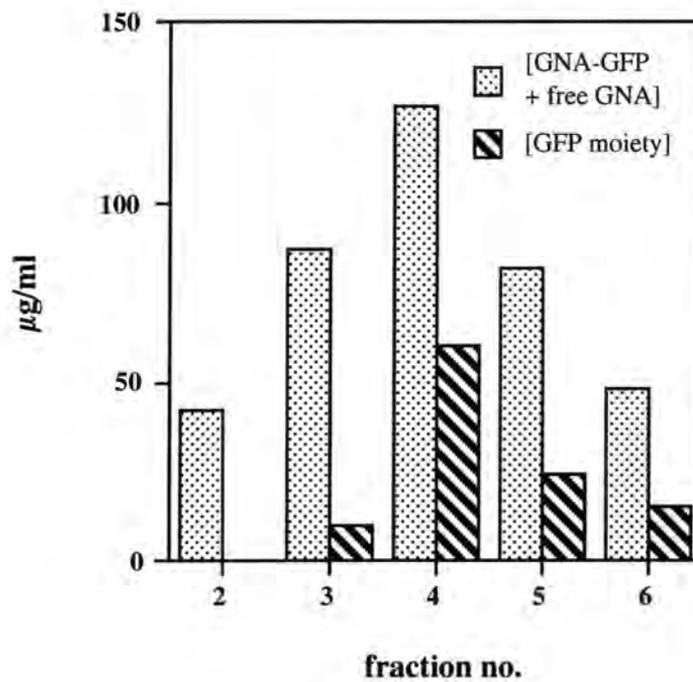


Figure 5.6 Elution peak profile of GNA-GFP during affinity purification (A) and concentration ratios of GNA-GFP to GNA in eluted fractions (2-6) collected (B). [GNA-GFP plus free GNA] was determined by Bradford assay using nGNA as standard; [GFP moiety] was determined by fluorimetry assay.

GNA-GFP, including a wash step to remove non-specific binding or any undesired contaminants.

When an equal volume of purified GNA-GFP fusion protein (fraction 4; initial concentration 120 $\mu\text{g} / \text{ml}$, as determined by Bradford assay or 60 $\mu\text{g} / \text{ml}$ of GFP moiety, as determined by fluorimetric measurements) was added to haemolymph extracted from 6th instar *L. oleracea*, fluorescence under UV light of some, but not all, cells was observed at low power magnification (x 20 objective lens) (Fig. 5.9(A)). When the haemolymph was directly incubated with rGFPuv standard protein (initial concentration 60 $\mu\text{g} / \text{ml}$) no fluorescence at all was detected on the cells (Fig. 5.9(C)). Also, in the presence of PBS (no GNA-GFP, no GFP), no fluorescence was observed (result not shown). Under a higher power of magnification (x 40 objective lens), fluorescent staining of certain cell-types provided further evidence that this observation was a consequence of GNA-GFP chimera binding (Fig. 5.9(B)).

To confirm that the interaction of the GNA-GFP fusion protein was through the lectin moiety and not through a non-specific binding of GFP, GNA-GFP (final concentration 5 $\mu\text{g} / \text{ml}$) binding to haemocytes was investigated by overlaying cell monolayers with the fusion protein, as described in Chapter 2 (section 2.8.8). Again, as shown by Figure 5.10(A), fluorescence was restricted to certain cell-types. In the presence of GFP (rGFPuv protein; final concentration 5 $\mu\text{g} / \text{ml}$) alone, no fluorescence was detected (Fig. 5.10(B)). When overlaying the monolayer with TBS, no fluorescence was observed (result not shown). Thus, fluorescence was not a consequence of non-specific interactions of GFP with haemocytes or of endogenous haemocyte fluorescence.

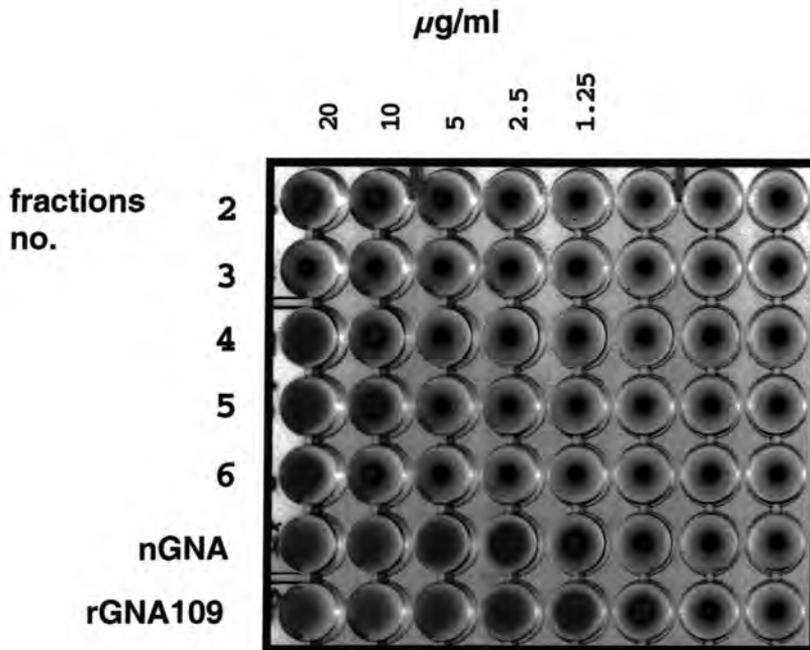


Figure 5.7 Haemagglutination assays of affinity-purified recombinant GNA-GFP. A two-fold serial dilution of each eluted fraction (see Fig. 7) with initial concentration $20 \mu\text{g/ml}$ was done in the presence of 1 % rabbit erythrocyte suspension (CSL, York).

(A)

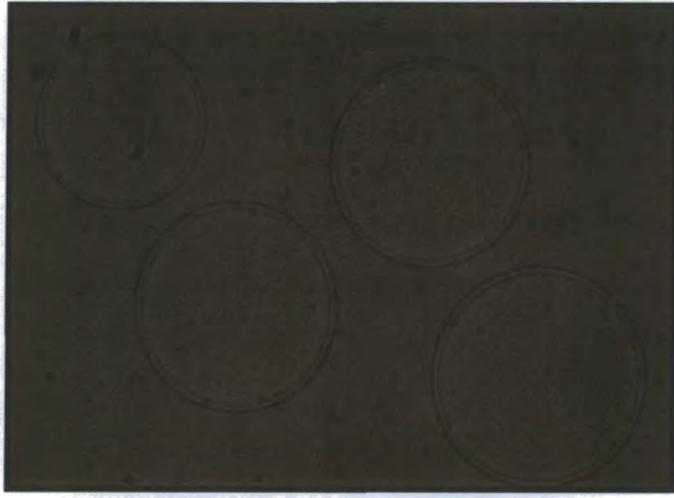
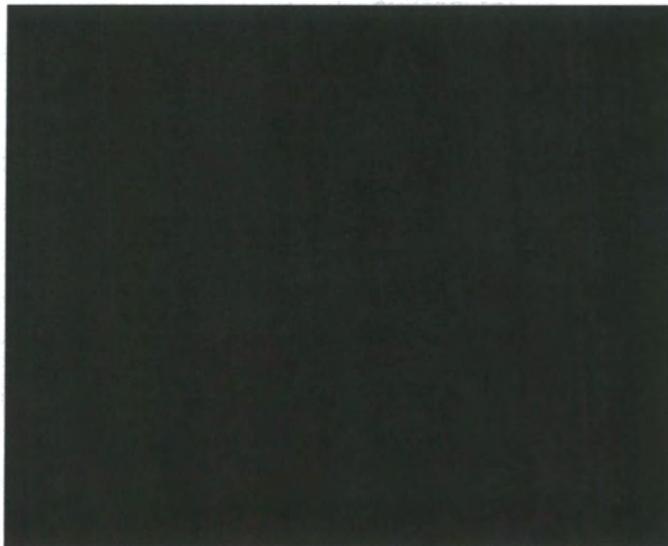
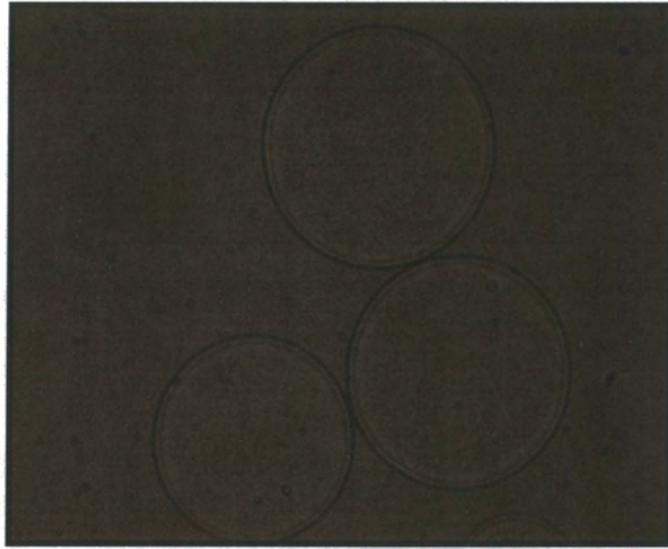


Figure 5.8 Binding of affinity-purified recombinant GNA-GFP to mannan-agarose beads. An equal volume of 50 $\mu\text{g/ml}$ of GNA-GFP (A) or standard rGFPuv (B) was added to a mannan-agarose bead suspension and incubated at room temperature for 1 hour. The beads were studied by light microscopy (x 20 objective lens) and photographed through Normasky optics (above) and fluorescent microscopy (below).

(B)

Figure 5.8 continued.

5.2.3.4 Binding of Recombinant GNA-GFP to Haemocytes *in vivo*

Five 4th instar *Lacanobia oleracea* caterpillars were fed for 24 hours on artificial diet alone or containing either PHA-E:GNA-GFP, rGFPuv or rGNA109. An average of 52 % of 0.5 g of diet, with or without the test protein, was consumed. Thus, it was estimated that 12 μ g of PHA-E:GNA-GFP (fraction 4), 6 μ g of rGFPuv and 12 μ g rGNA109 were ingested. Confocal microscopic studies of haemolymph extracted with meticulous care from caterpillars showed results that were consistent with those obtained in the *in vitro* work (section 5.2.3.3). Of the five caterpillars that were tested on GNA-GFP, only one showed no evidence of fluorescence. For the other four, although the fluorescence was not as abundant or bright as was clearly shown *in vitro*, a very limited number of cells was detected which showed either a faint overall or sharp localised GNA-GFP binding (Fig. 5.11). These observations could be accounted for by the low concentration of GNA-GFP that was taken through the gut into the haemolymph. The haemolymph of all caterpillars fed on artificial diet alone or on diet containing GFP or GNA revealed no evidence of fluorescence due to non-specific interactions of cells to GFP or autofluorescence.

Figures 5.12 (A) & (B) reveal the results of immunoblotting analysis of pooled haemolymph for each protein tested. The presence of the GNA-GFP fusion protein in the haemolymph was confirmed by both anti-GNA and anti-GFP antibodies (lane C, Fig. 5.12 (A) & (B)). However, further proteolytic cleavage had occurred since anti-GFP antibodies picked out a second band with molecular weight similar to that of the GFP standard, and anti-GNA antibodies had detected a number of bands with molecular weights between that of native GNA monomer and of the chimeric product.

(A)

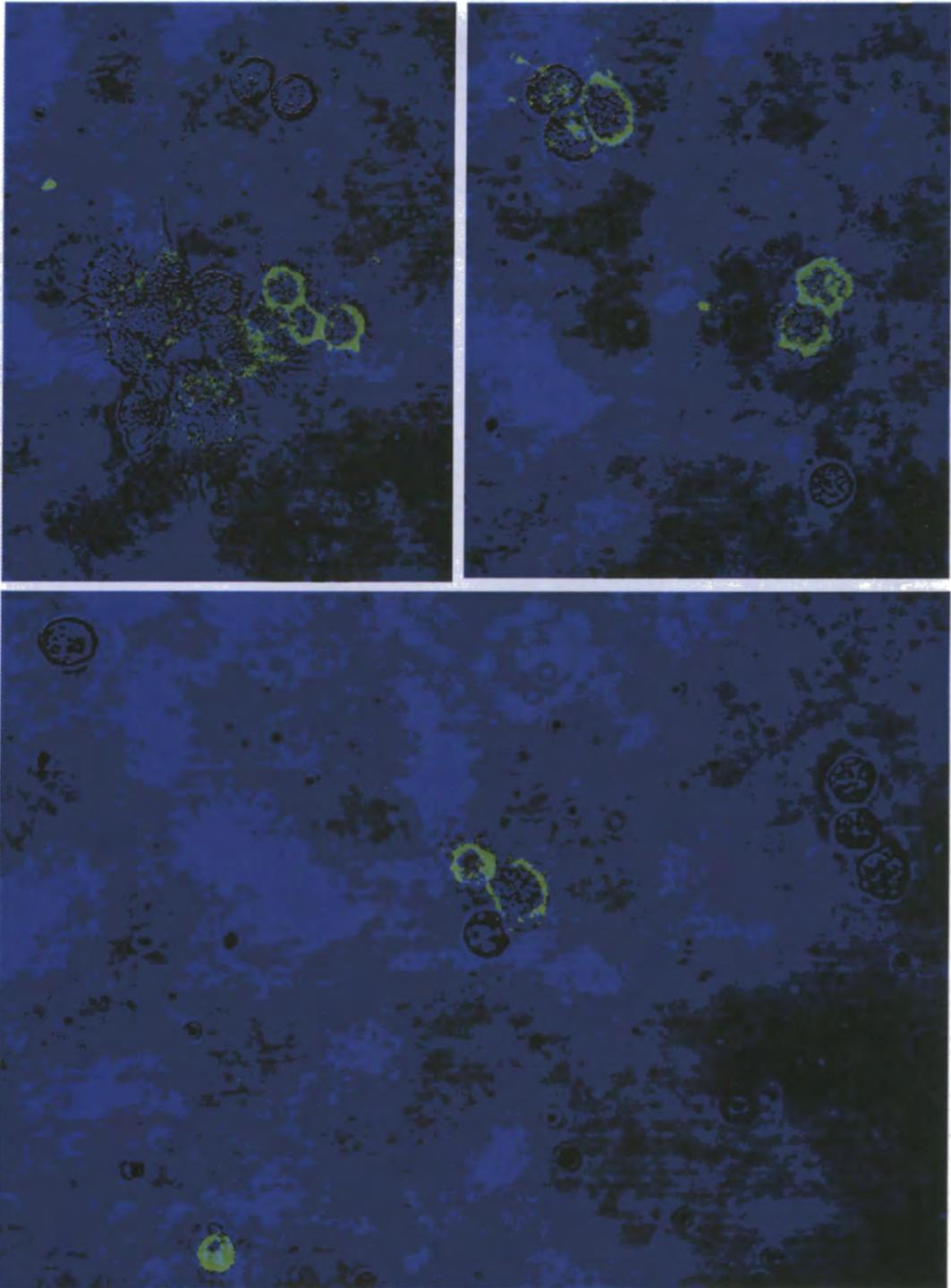
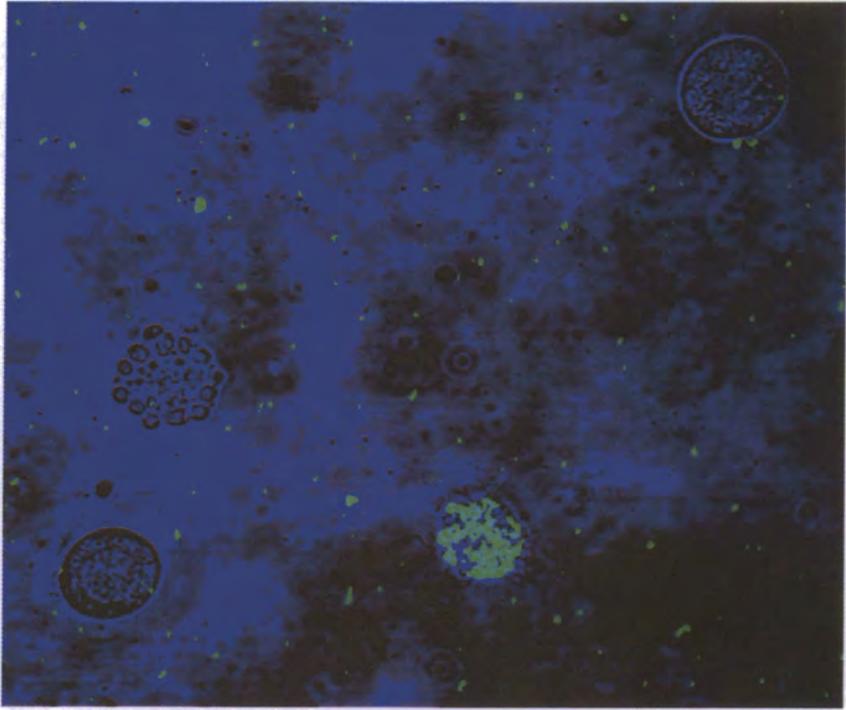
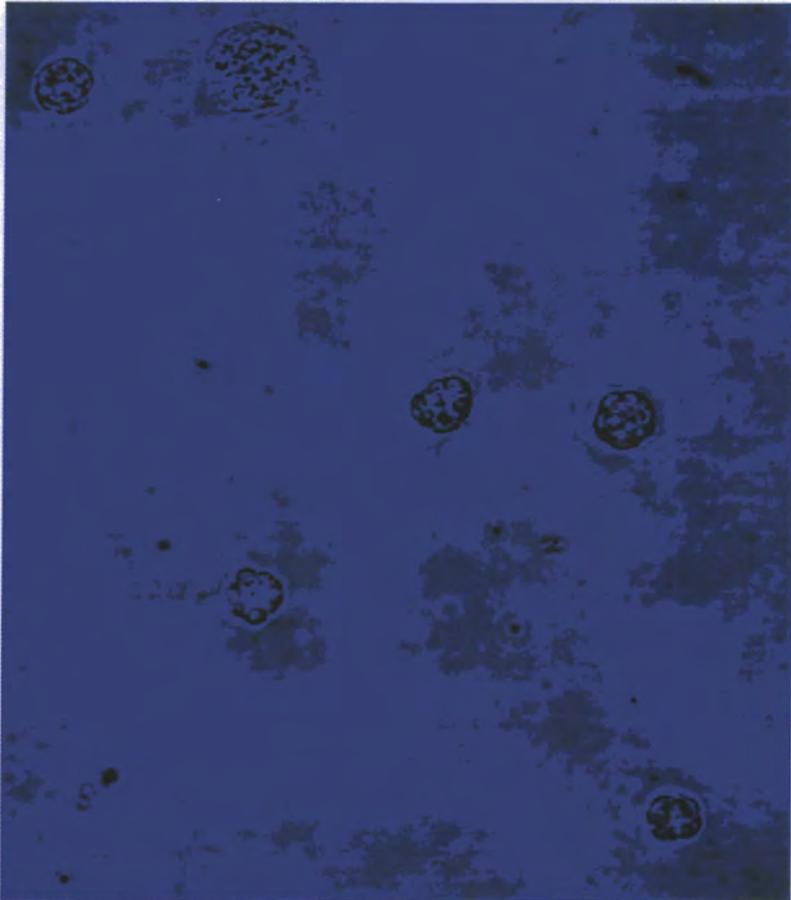


Figure 5.9 Binding of affinity-purified recombinant GNA-GFP to 6 th instar *Lacanobia oleracea* haemocytes in vitro - crude technique. Three μl of haemolymph was added to 3 μl of GNA-GFP (fraction 4) to a final concentration of 60 $\mu\text{g/ml}$ (A & B) or to standard rGFPuv to a final concentration of 30 $\mu\text{g/ml}$ (C) and the suspension incubated at room temperature for 5 minutes. The haemocytes were studied by confocal microscopy (x 20 objective lens for A & C; x 40 objective lens for B).

(B)



(C)

Figure 5.9 continued.

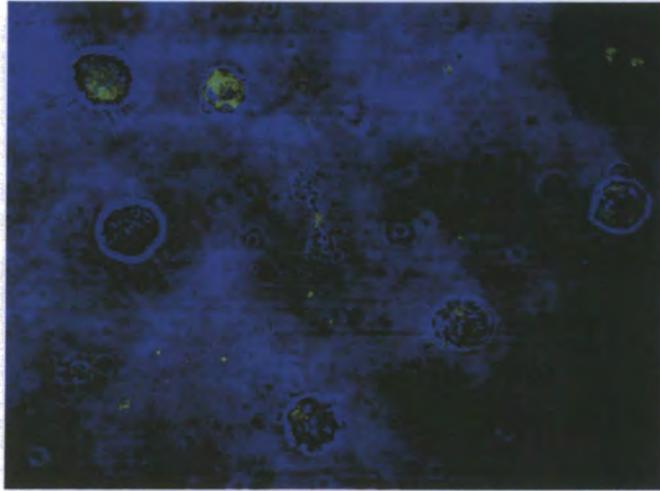
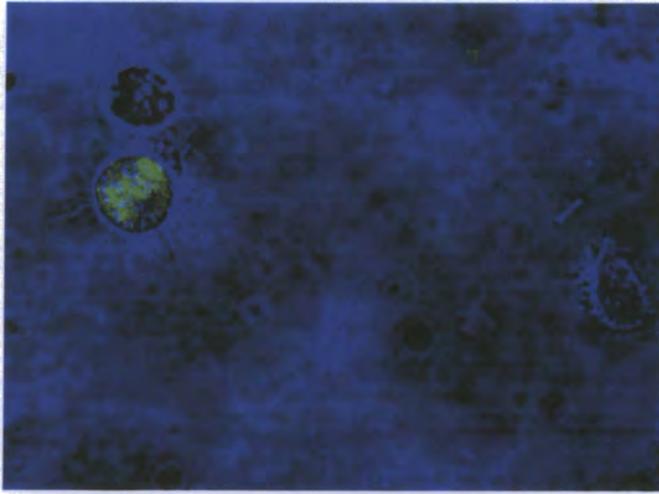
A degree of non-specific binding had also occurred since several bands appeared in the control lanes of the anti-GNA antibody blot (lane E, Fig. 5.12 (A)). Anti-GNA antibodies also detected recombinant GNA109 in the haemolymph; lane B of Figure 5.12 (A) shows a band that migrated as far as the native GNA monomer. Surprisingly, the more specific anti-GFP antibodies detected rGFPuv in the haemolymph (lane D, Fig. 5.12 (B)), too. No proteolytic cleavage of GFP seemed to have taken place since its molecular weight remained unchanged at 27 kDa.

5.3 Discussion

The goal of this study was to determine whether GNA and GFP could be linked into a chimeric molecule that combined the functional properties of the individual proteins. The results above clearly demonstrate that the lectin-based chimeric protein, GNA-GFP, consisting of GNA isolectin 2 fused to the GFPuv protein through a vector-derived linker, was expressed successfully in *Pichia pastoris* exhibiting dual activity. Although GFPuv was optimised for bacterial expression (Crameri *et al.*, 1996), it fluoresced very efficiently in *Pichia*. In contrast, human optimised EGFP (Clontech) did not express at all in *Pichia pastoris* (results not shown) and is generally not efficiently expressed in yeast (Cormack *et al.*, 1997).

The expressed fusion protein (and rGFP), using either leader (α -factor prepro- or PHA-E leader) construct, was not only secreted into the culture supernatant, but was also directed to intracellular organelles in the yeast cell. Presumably these organelles were vacuoles. The GFP fluorescence was displayed in large structures which stood out brightly from the dim fluorescent cytoplasm. Form and size of vacuoles may

(A)



(B)

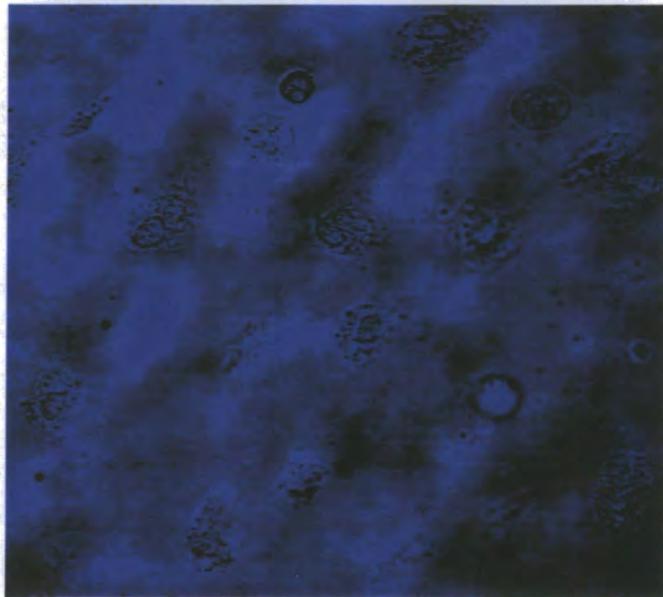


Figure 5.10 Binding of affinity-purified recombinant GNA-GFP to 6 th instar *Lacanobia oleracea* haemocytes in vitro - monolayer technique. The cell monolayer was incubated in TBS with adjusted osmolarity containing 5 $\mu\text{g/ml}$ GNA-GFP (A) or 5 $\mu\text{g/ml}$ standard rGFPuv (B). The haemocytes were studied by confocal microscopy.

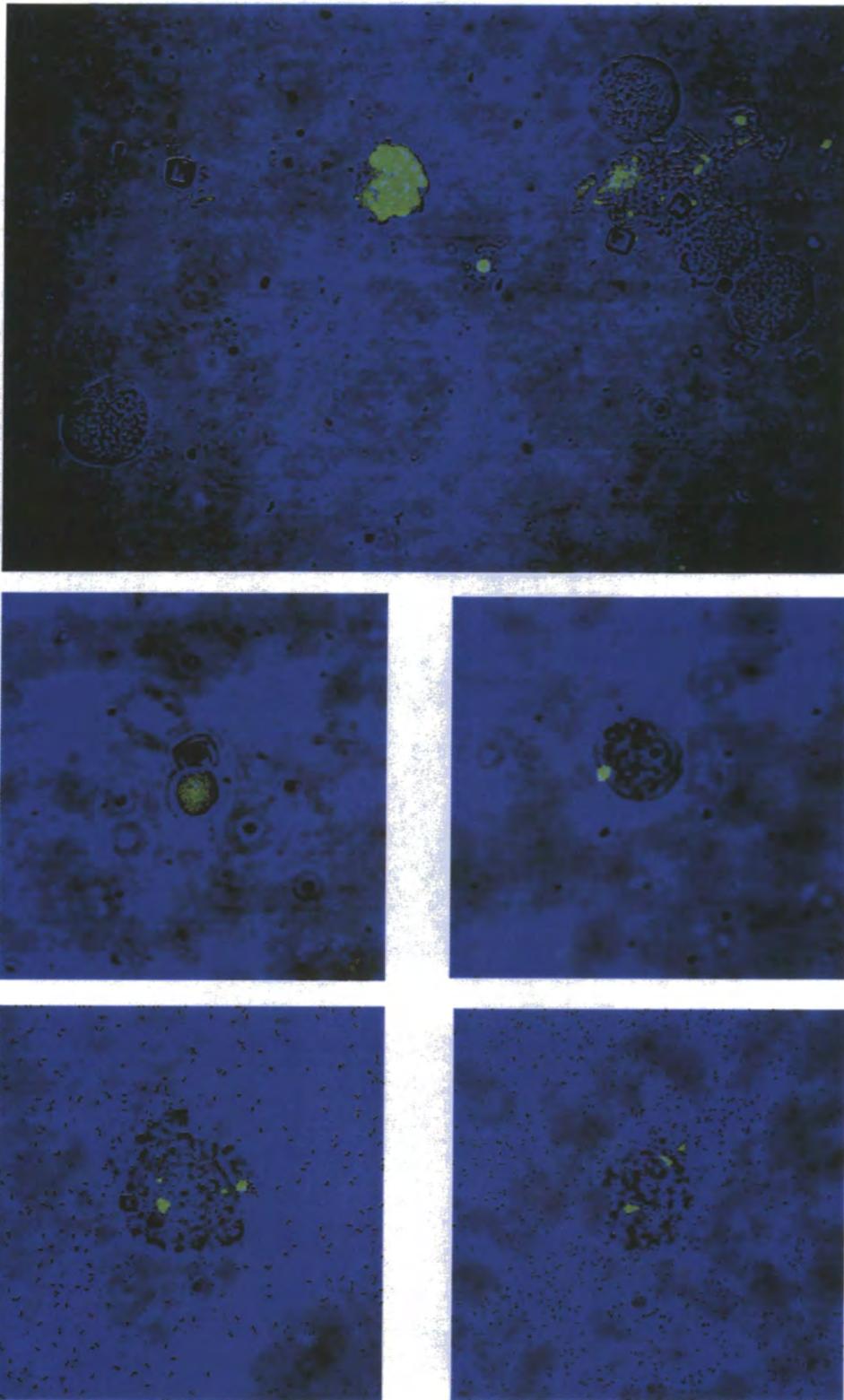


Figure 5.11 Binding of affinity-purified recombinant GNA-GFP to 4th instar *Lacanobia oleracea* haemocytes in vivo. Haemolymph extracted from different caterpillars fed on artificial diet containing GNA-GFP was studied by confocal microscopy.

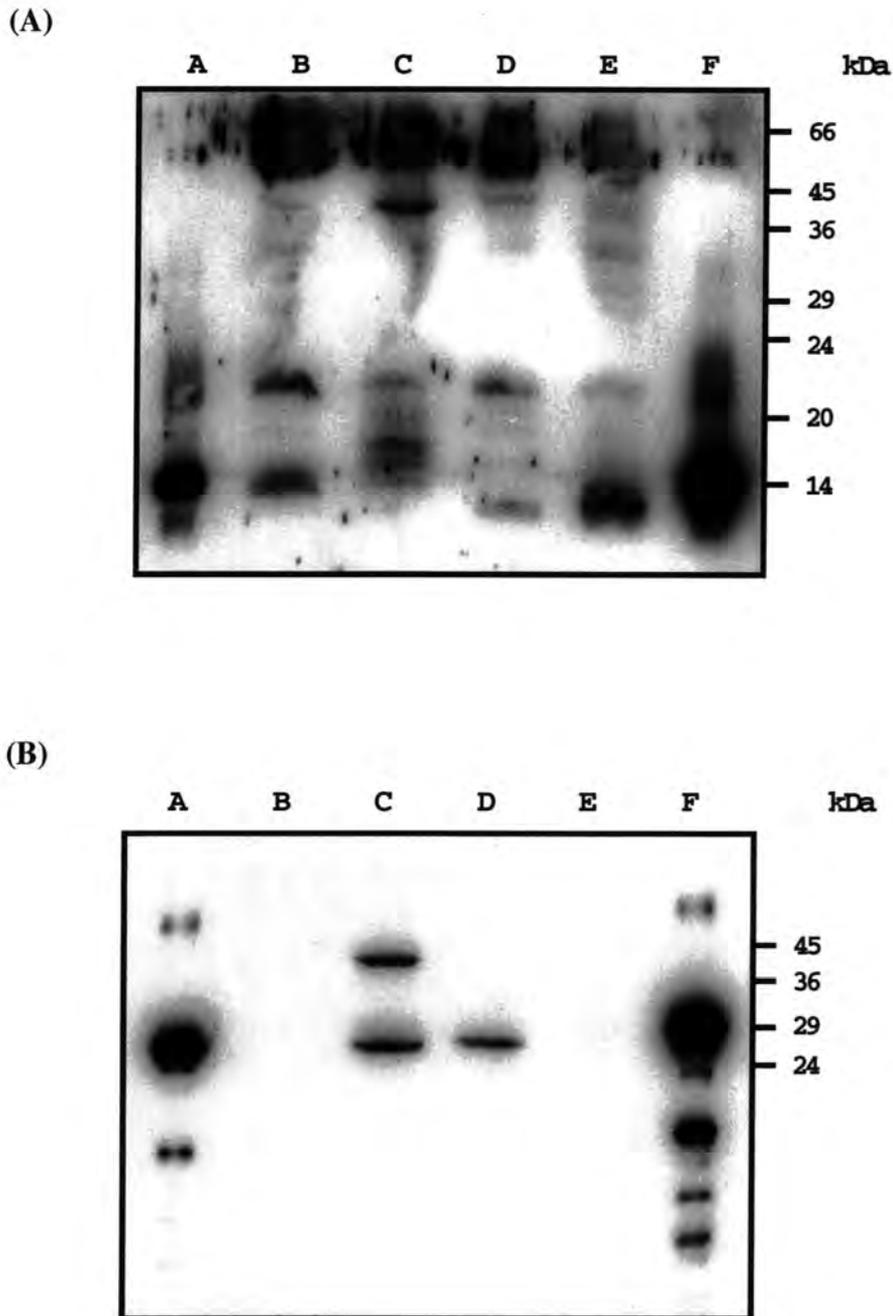


Figure 5.12 Immunoblot analysis using polyclonal anti-GNA antiserum (A) or monoclonal anti-GFP antibodies (B) of pooled haemolymph extracted from 4th instar *Lacanobia oleracea*. The caterpillars were fed on artificial diet containing recombinant GNA109 (lane B), affinity-purified recombinant GNA-GFP (lane C), standard rGFPuv (lane D), and casein (lane E). Lanes A and F: 25 and 50 ng, respectively, of native GNA (A) or standard rGFPuv (B). For all samples, 10 μ l of haemolymph was loaded.

change considerably during the yeast cell cycle (Tuite, 1991), which could account for observations of fluorescent subcellular structures of different sizes in different cells. Western blots confirmed the presence of the fusion product both in the supernatant and in the cell. However, the band profiles were not the same. From the blots of cell extracts, more bands were noticeable suggesting that this was most probably due to proteolytic cleavage from the abundant proteases contained within the vacuoles (Martinoia *et al.*, 1979). Also, in order to estimate the relative concentrations of proteins from the blots differences in sensitivity of the two antibodies had to be taken into consideration. The polyclonal anti-nGNA antibodies were not as specific as the mix of monoclonal anti-GFP antibodies which were raised against both GFP alone and a GFP fusion product (Boehringer-Mannheim).

The first evidence of dual activities of the lectin and GFP moieties of the fusion product was observed at the cellular level when, for both fusion constructs, the *Pichia* cell wall fluoresced strongly at the GFP wavelength. The outer surface of the yeast cell wall contains highly immunogenic mannoproteins which consist of large numbers of mannose groups covalently bound to N-acetylglucosamine groups that in turn are attached to specific amino acid side groups of a number of cell wall-specific polypeptide chains. These mannose structures are recognised and bound to by GNA as previously observed, and the fluorescence is most likely due to bound GNA-GFP fusion. Furthermore, this hypothesis is supported by the finding that GNA agglutinated *Pichia pastoris* cells (see Chapter 4.2.5), through mediation of the lectin with mannose structures on the cell surface.

Interestingly, 95 % of the fusion product expressed from construct PHA-E:GNA-GFP was confined within the cells, whereas 83 % of total GNA-GFP expressed from construct α :GNA-GFP was secreted into the culture supernatant. These results do not reflect directly the observations made on the fluorescent cells (Fig. 5.1), since the fluorescence was not only contributed by the fusion product, but also by the cleaved GFP. In contrast, recombinant GFP was found mostly in *Pichia* cells when expressed from α :GFP or PHA-E:GFP (75 % and 91 % of total GFP protein, respectively). These results are surprising since it was expected that most of the heterologous protein (with the exception of α :GNA-GFP) would be secreted when directed by the leader sequences. These foreign proteins may contain positive sorting information within their polypeptide chains which may explain their diversion from the 'default' secretory route (Stevens *et al.*, 1986).

Despite the slight differences in band profiles as conveyed by the immunoblots shown in Figure 5.2, expression from either construct α :GNA-GFP or PHA-E:GNA-GFP could potentially be chosen for functional analysis. However, since the results of the previous chapter (Chapter 4) indicated that constructs containing the PHA-E leader sequence expressed in *Pichia pastoris* produced heterologous proteins that were correctly processed and that no N-terminal sequencing was carried out here, further work focussed mainly on construct PHA-E:GNA-GFP.

Purifying GNA-GFP from culture supernatant using mannan-agarose proved much less cumbersome than purifying the chimera from cell extract (result not shown); breaking the *Pichia* cells open with glass beads was a tedious and inefficient process resulting in low yields. Binding of GNA-GFP to the mannan-agarose column matrix

was most probably mediated by the oligomeric association of the fusion protein with cleaved GNA; tetrameric formation of homogenous GNA-GFPs was sterically unlikely. In fact, evidence of oligomerisation was shown by immunoblotting of samples that were not subjected to heat treatment or β -mercaptoethanol addition (Fig. 5.5). Thus, partial proteolysis of the fusion product and co-purification of GNA were an advantage, not an interference, to binding activity. A theoretical relationship between the number of molecules of free GNA with the number of molecules of GNA-GFP in the collected eluted fractions of affinity-purified recombinant GNA-GFP sample is presented in the Appendix.

The agglutination activity of rabbit erythrocytes was caused by the concerted activities of both GNA-GFP and GNA. Agglutination solely by homogenous GNA tetramers was potentially possible because sufficient levels of free GNA were present in the purified fractions (4 and 5). Fraction 2 did not contain any GFP fluorescence and hence no GNA-GFP, only free GNA. However, complete haemagglutination using this fraction was not achieved at a concentration of 20 $\mu\text{g} / \text{ml}$. This is evidence that the lectin activity of co-purified cleaved GNA was reduced because the C-terminal linker residues may have obstructed proper binding of the lectins to the sugars on the outer surface of the red blood cells. The activity was expected to be the same as that of rGNA109 since Longstaff *et al.* (1998) showed that recombinant GNA with a C-terminal hexa-histidine tail was as active as native GNA.

However, when GNA-GFP binding to mannan-agarose beads was observed under UV light, it was apparent that, without doubt, the binding involved the fusion product

since the beads were fluorescing. This finding clearly suggests that the fusion protein is able to take part in lectin binding activity.

The ability of recombinant GNA-GFP to bind *in vitro* and *in vivo* to haemocytes from haemolymph samples extracted from larvae of the lepidopteran *Lacanobia oleracea* was demonstrated. That only certain cell types were fluorescing indicates that binding of the fusion protein was specific; of the haemocytes that were identifiable (Ribeiro *et al.*, 1996), GNA-GFP fluorescence was observed on either plasmatocytes or granulocytes, whereas no fluorescence was observed on spherular cells. This is in strong contrast to the observation made when FITC-labelled GNA was found to bind to haemocytes non-selectively (Fitches, 1998). Since FITC in itself showed signs of binding to the cells, the distinction between selectively and non-selectively bound GNA-FITC proved difficult. In this work, it was observed that GFP alone does not stick to cells, even at relatively high concentrations (30 $\mu\text{g} / \text{ml}$). Fitches (1998) demonstrated that GNA could be detected in the haemolymph of *Lacanobia oleracea* when fed on artificial diet containing the plant lectin. Demonstration of *in vivo* binding of GNA-GFP to haemocytes extracted from lectin-fed larvae showed that the GNA-GFP fusion was also taken up into the haemolymph, presumably through the functional GNA part of the fusion mediating uptake through binding to the insect gut surface, as observed previously for GNA. It was shown, by immunoblotting, that the fusion protein was present in the haemolymph, and although its concentration was low, fluorescence by the GFP moiety was detectable as a result of specific binding to specific cells in the haemolymph. Interestingly, cleaved GFP as a result of partial proteolysis (most probably by gut proteases at the linker region of the fusion protein) was also detected in the haemolymph. In addition, immunoblotting of blood samples

extracted from the caterpillars fed on artificial diet containing GFP alone showed that GFP had managed to traverse the midgut barriers into the haemolymph at low efficiency, but did not show binding to cells in the haemolymph. As expected, cleaved GNA and GNA used as control were found in the haemolymph, too.

Chapter 6

General Discussion & Future Prospects

This work has as its aim the development of microbial expression systems based on *Escherichia coli* and *Pichia pastoris* to produce functional recombinant plant lectins. Towards this aim, two unrelated plant lectins were chosen: (i) the lectin from the snowdrop plant (*Galanthus nivalis* agglutinin, or GNA), a member of the Amaryllidaceae family, and (ii) the lectin from the French bean *Phaseolus vulgaris* (phytohaemagglutinin, or PHA), a member of the Leguminosae family. These lectins are members of two distinct families of proteins. Much interest has focussed on the snowdrop lectin as candidate for genetic engineering of crops since it has established insecticidal effects towards a wide range of insects including homopterans (e.g. *Nilaparvata lugens*, or rice brown planthopper, BPH) and lepidopterans (e.g. *Lacanobia oleracea*, or tomato moth), and no evidence of toxicity towards mammals. PHA, on the other hand, is toxic to higher animals but is of limited toxicity to insects. Following the successful production of correctly processed plant lectins, the heterologous expression systems were used as a basis for carrying out site-directed mutagenesis on one of the carbohydrate-binding sites of GNA and for the creation of a functional GNA-GFP chimaera in order to complement on-going studies on the mode of action of the snowdrop lectin with respect to insect toxicity.

Like most plant lectins, snowdrop lectin and phytohaemagglutinin exist as a mixture of multiple isoforms in nature. GNA isolectins only vary slightly in amino acid residue sequence, most noticeably in the C-terminal region (Van Damme *et al.*,

1991a) and this variation is not thought to affect the specificity of carbohydrate binding. It was important to ensure that the isoform previously selected for expression in a transgenic plant showed the anticipated antimetabolic effects. Whether expressed in *E. coli* or *P. pastoris*, recombinant GNA isoform 2 consistently imparts a two-fold higher haemagglutination activity than native snowdrop lectin. This consistent difference is significant even though such a difference in an individual assay would not be. Also, the native GNA was shown to have a slightly more pronounced binding to carboxypeptidase Y in the ELISA assays than bacterially expressed GNA. However, no difference in toxicity between recombinant and native snowdrop lectin was established in the brown planthopper bioassay. Therefore, it may be assumed that the difference in activity is the consequence of the slight changes in amino acid residue content of the different GNA isoforms, and that GNA isoform 2 represents a “high activity” lectin sequence compared to other isoforms present in the mixed native GNA preparation.

Almost all recombinant plant lectins to date expressed in bacteria are from the legume family. Most of them accumulate as insoluble inclusion bodies in the bacteria. Although GNA comes from a different protein family than the legume lectins, and no homology is present between the monocot lectins represented by GNA and legume lectins, expression of the snowdrop lectin in *E. coli* was no exception. In this system, expression of GNA was driven by an inducible T7 promoter in a pET plasmid vector. The construct was engineered so that the protein was expressed without the N-terminal leader sequence. A procedure to resolubilise denatured recombinant GNA was optimised so that most of the lectin was refolded into a functionally active state. Denaturation and refolding procedures are case-to-case dependent, but typically do

not give high yields. For example, Arango *et al.* (1992) showed that only a very small percentage (1 - 4 %) of insoluble EcorL, a lectin from the legume *Erythrina corallodendron*, was recovered into an active form.

The four extra C-terminal residues of the recombinant GNA109 polypeptide manifested themselves in the ELISA assays where rGNA109 bound more strongly to carboxypeptidase Y than rGNA105. No difference in activities between the two recombinants was observed in other functional assays. It is probable that these residues do contribute significantly in GNA's activity since they participate in the C-terminal strand exchange which is important for (i) stabilisation of the dimers, and (ii) binding of terminal mannose monosaccharides at site I. The strand cross-over to form hybrid beta sheets do exist in other proteins, however, this is rare (Bennett *et al.*, 1995). For example, the monomer-monomer association of the panicum mosaic virus is stabilized by strand insertion (Ban & McPherson, 1995).

Despite the refolding problem, the *E. coli* expression system does provide sufficient recombinant GNA for allowing alterations of individual amino acids in the binding sites to study their effect on the sugar specificity of the protein and consequently on their toxicity towards the rice brown planthopper. By altering Tyr97 to Phe97, the hydrogen bond between the amino acid side chain phenyl OH and the mannose 4OH group was deleted. However, this did not abolish the biological activity of rGNA105. Instead, the mutant, rGNA105(Y97F) was less active (and less toxic) than its wild-type counterpart. In contrast, rGNA109(Y97F) showed no difference in haemagglutination activity when compared to wild-type rGNA109. This finding further supports the important role the 4 extra C-terminal residues play in sugar

binding. The importance of the contribution of aromatic amino acid residues can be reflected in other cases. Nishiguchi *et al.* (1997) revealed that the aromatic ring of Phe130 of the *Robinia pseudoacacia* bark lectin is essential for carbohydrate binding. The lectin still had haemagglutinating activity when Phe was replaced by tyrosine, however, when Phe was substituted for Ala or Leu, the bark lectin's activity was completely abolished. Substitution of Tyr134 of *Griffonia simplicifolia* lectin (GSII) to aspartic acid or glycine eliminated carbohydrate-binding and biological activity (Zhu-Salzman *et al.*, 1998). However, Tyr134 replaced with Phe134, in which the aromatic ring was preserved, did not change the binding or insecticidal activity. Similarly, loss of activity was shown when Phe131 of EcorL was subjected to mutations (Adar & Sharon, 1996).

The two different isoforms which constitute PHA (E and L forms), are also very similar in residue content (> 80 % homologous) (Hoffman and Donaldson, 1985). In contrast to GNA, the activities of both isoforms are different: only PHA-E is erythroagglutinating, whereas only PHA-L confers both leucoagglutinating and mitogenic activity. Isolation of pure native E₄ or L₄ forms by conventional techniques is difficult due to the similar physicochemical properties of both E and L types. Therefore, contamination with tetramers containing either subunit type would be inevitable. In an attempt to obtain a pure PHA isoform, Hoffman and Donaldson (1987) expressed soluble PHA-L in *E. coli*, but only a low yield (75 µg/l culture) of affinity-purified lectin was attainable. Although the plant signal peptide was recognized by the bacterial proteases, the recombinant protein was incorrectly processed. Interestingly, the absence of N-linked carbohydrates in bacterially expressed PHA-L did not abolish biological activity confirming the report by Bollini

et al. (1985) that unglycosylated PHA from bean cotyledons treated with tunicamycin had mitogenic and erythroagglutinating activities similar to those of glycosylated PHA.

When expressed in *Saccharomyces cerevisiae*, part of the recombinant PHA molecules had their N-terminal signal peptide still attached to the mature polypeptides, most probably due to failure of recognition of the plant signal peptide by the yeast proteases (Tague & Chrispeels, 1987). Most of the protein was directed to the vacuole with a small percentage of the protein containing the leader peptide secreted. Surprisingly, in this work, it was shown that expression of PHA in *Pichia pastoris* resulted in complete removal of the signal peptide. Furthermore, the lectin was correctly processed and a higher percentage was secreted into the yeast culture medium. It would be of interest to determine the factors causing the unexpected difference in signal peptide recognition between *Saccharomyces* and *Pichia*.

Expression of GFP and GNA-GFP in *Pichia pastoris* resulted in intracellular fluorescence indicating that not all of the recombinant proteins were secreted. It would be desirable to more accurately localise the recombinant proteins produced in these cells by, for instance, immunogold labelling, in order to verify localisation to vacuoles. Stevens *et al.* (1986) hypothesized that secretion is the default pathway taken by proteins with no specific vacuolar targetting signals. In the event of saturation of the sorting or modifying machinery after overexpression may result in secretion. However, Chao and Etzler (1994) suggested that yeast and plants utilize different vacuolar targetting signals after observing that *Dolichos biflorus* lectins, naturally confined to the plant vacuoles, were secreted into the yeast medium even at

low expression levels. Furthermore, these lectins did accumulate in the cells, but not in the vacuoles. The natural tendency for the lectin subunits to aggregate may have resulted in the retention of the protein in the endoplasmic reticulum (Chao & Etzler, 1994).

Although expressing PHA using the commercially available alpha prepro- leader sequence generally led to higher yields, incorrect processing by the yeast proteases resulted in the formation of different polypeptide species, a feature not uncommon in other reported cases (Sreekrishna *et al.*, 1997). Furthermore, since these proteins contained additional glycosylated N-terminal segments of the alpha-factor pro-region, the biological function of PHA was abolished probably due to the interfering N-linked carbohydrate side chains. However, this phenomena was not shared by alpha GNA since both species bound to and eluted from the affinity column.

GNA and an independent nonlectin, GFP, were both correctly processed in *Pichia* and secreted in culture medium when expressed from a construct containing the PHA-E leader sequence. It is assumed from these results that the PHA-E signal peptide could be used for proteins where secretion and processing using other signal peptides has proven problematic. This is important when considering the use of plant lectins for drug development or crop genetic engineering where it is imperative to maintain strict control on the quality of the recombinant protein. For instance, in order to produce an identical plant lectin to the one expressed in a transgenic crop, an efficient heterologous expression system, such as the one developed in this work, is highly desirable.

Expression of recombinant GNA-GFP in *Pichia* with a linker joining the two proteins which is sensitive to proteolysis is not necessarily disadvantageous. In nature, the snowdrop lectin exists as a homotetramer (Van Damme *et al.*, 1987) and although the GNA monomer has the potential to bind mannose monosaccharides in its three binding sites, sugar binding of more complex structures containing mannose residues, as those found on cell surface glycoproteins, is much stronger due to cooperation of the neighbouring subunits of the tetramer (Hester *et al.*, 1995; Hester & Wright, 1996; Wright & Hester, 1996). Thus, quaternary association of GNA-GFP into tetramers seemed to be facilitated by the presence of GNA, proteolytically cleaved from GNA-GFP fusion polypeptides. If the linker peptide was resistant to protease attack by *Pichia*, formation of a GNA-GFP tetramer might be disfavoured due to steric interference of the large GFP moiety(ies). In this case, in order to permit proper carbohydrate binding, supplementing the uncleaved GNA-GFP sample with free GNA molecules would probably be required to allow heterotetrameric formation. It would be interesting to carry out further experiments on quaternary association to verify the number of subunits that the chimaeric GNA-GFP contains.

The tomato moth (*Lacanobia oleracea*) was chosen for functional testing of the recombinant GNA chimaera due to its importance as a pest in agriculture and the facile handling of the caterpillars. GNA was shown to have a detrimental effect upon larval development, growth and consumption, with little effect on survival (Fitches & Gatehouse, 1998). It seems that the toxic effects attributable to the GNA lectin are ascribed to several factors. Binding to glycosylated gut receptors recognized by the GNA moiety would probably cause inhibition of nutrient absorption and/or midgut cell disruption in the insect (Pusztai, 1991). Incidentally, lectin toxicity towards

mammals depends largely on the binding of lectins to suitably glycosylated targets in the intestinal brush border membranes (Pusztai, 1991). In contrast to PHA, GNA lacks toxicity to mammals due to the relative scarcity of mannose-containing brush border glycans (Pusztai *et al.*, 1990). In the haemolymph of *Lacnobia*, Fitches *et al.* (1998) showed that GNA has the ability to bind to and reduce the number of haemocytes, which may have an effect on the immune system of the insect. This work complements the mode of action study on GNA's toxicity towards *Lacnobia* by showing that the GNA-GFP hybrid protein bound selectively to specific haemolymph cell-types. Binding of the GNA moiety to mannosylated gut receptors is presumed to be responsible for the initial stages of uptake by endocytosis of the chimaeric molecule on its way to the haemolymph.

Bacterially expressed rGNA105 and rGNA105(Y97F) were shown to bind differentially to the same specific BPH gut glycoprotein receptors *in vitro*. This difference in binding resulted in a decrease in toxicity of GNA towards BPH *in vivo*. This work complements earlier work by Powell *et al.* (1998) who reported GNA binding to the BPH gut tract. Thus, it would be desirable to further characterise the specific glycoprotein receptors concerned. Du *et al.* (2000) have characterised one such receptor, a ferritin, in detail.

In conclusion, the initial objectives of this research project have largely been achieved, i.e. the development of microbial heterologous expression systems to produce adequate functional plant lectins by (i) expressing snowdrop lectin in *E. coli* and carrying out site-directed mutagenesis of the carbohydrate-binding site, (ii) expressing snowdrop lectin and phytohaemagglutinin in *P. pastoris*, and (iii)

expressing a chimaeric GNA-GFP for functional studies. The expression of plant lectins in *Pichia* is preferred over *E. coli* due to the inefficient inherent bacterial problem of refolding/activation of insoluble protein from inclusion bodies. Also, the proteins expressed using the *Pichia* system developed throughout the project were 'clean cut', whereas those expressed by *E. coli* did not remove the initiation methionine although this did not compromise the activity of bacterial expressed snowdrop lectin. Yet, an expression system based on *Pichia* has now been developed to allow the production of a recombinant plant lectin (or protein), or protein derived therefrom (e.g. chimaeras), with the desired sequence.

Chapter 7

Appendix

A relationship between the no. of molecules of free GNA with the no. of molecules of GNA-GFP in the collected eluted fractions of affinity-purified recombinant GNA-GFP can be roughly estimated as follows:

variables

a = estimated total mass of GNA-GFP plus free GNA

b = estimated total mass of GFP

n = no. of molecules of GFP moiety and GNA-GFP

m = no. of molecules of free GNA

gna = mass of free GNA molecule

gfp = mass of GFP molecule

gnagfp = mass of GNA-GFP molecule

equations

$$(1) \quad a = n \cdot \text{gnagfp} + m \cdot \text{gna}$$

$$(2) \quad b = n \cdot \text{gfp}$$

$$(3) \quad \text{gnagfp} = \text{gna} + \text{gfp}$$

$$(2) \rightarrow (3)$$

$$(4) \quad n \cdot \text{gnagfp} = n \cdot \text{gna} + b$$

$$(4) \rightarrow (1)$$

$$(5) \quad a = n \cdot \text{gna} + b + m \cdot \text{gna}$$

$$(5) \quad a - b = \text{gna} (n + m)$$

$$(6) \quad (27/40) \cdot \text{gnagfp} = \text{gfp}$$

$$(7) \quad (14/40) \cdot \text{gnagfp} = \text{gna}$$

$$(6) \rightarrow (7)$$

$$(8) \quad (14/27) \cdot \text{gfp} = \text{gna}$$

$$(2) \rightarrow (8)$$

$$(9) \quad (14/27)(b/n) = \text{gna}$$

$$(9) \rightarrow (5)$$

$$(10) \quad \begin{aligned} a - b &= (14b/27n)(m + n) \\ (27n/14b)(a - b) &= m + n \end{aligned}$$

$$\underline{n[(27/14b)(a - b) - 1] = m}$$

thus, there exists a simple linear relationship between the no. of molecules of free GNA (m) with the no. of molecules of GNA-GFP (n).

e.g. for affinity-purified GNA-GFP fraction 4, $a = 120$ and $b = 60$, $m = n$; for fraction 5, $a = 82$ and $b = 24$, $m = 4n$.

References

- Abbot, W. S. (1925) A method of computing the effectiveness of an insecticide. *J. Econ. Entomol.* **18**, 266-267.
- Adar, R. & Sharon, N. (1996) Mutational studies of the amino acid residues in the combining site of *Erythrina corallodendron* lectin. *Eur. J. Biochem.* **239**, 668-674.
- Ammerer, G., Hunter, C. P., Rothman, J. H., Saari, G. C., Valls, L. A. & Stevens, T. H. (1986) *PEP4* gene of *Saccharomyces cerevisiae* encodes proteinase A, a vacuolar enzyme required for processing of vacuolar precursors. *Mol. Cell. Biol.* **6**, 2490-2499.
- Arango, R., Rodriguezarango, E., Adar, R., Belenky, D., Loontjens, F. G., Rozenblatt, S. & Sharon, N. (1993) Modification by site-directed mutagenesis of the specificity of *Erythrina corallodendron* lectin for galactose derivatives with bulky substituents at c-2. *FEBS Lett.* **330**, 133-136.
- Balzarini, J., Schols, D., Neyts, J., Van Damme, E. J. M., Peumans, W. J. & Declercq, F. (1991) Alpha-(1-3)-D-mannose-specific and alpha-(1-6)-D-mannose-specific plant lectins are markedly inhibitory to human immunodeficiency virus and cytomegalovirus infections in vitro. *Antimicrob. Agents Chemother.* **35**, 410-416.

-
- Ban, N. & McPherson, A. (1995) The structure of satellite panicum mosaic virus at 1.9 angstrom resolution. *Nat. Struct. Biol.* **2**, 882-890.
- Banerjee, R., Das, K., Ravishankar, R., Suguna, K., Surolia, A. & Vijayan, M. (1996) Conformation, protein-carbohydrate interactions and a novel subunit association in the refined structure of peanut lectin-lactose complex. *J. Mol. Biol.* **259**, 281-296.
- Barbieri, L., Batelli, G. B. & Stirpe, F. (1993) Ribosome-inactivating proteins from plants. *Biochim. Biophys. Acta* **1154**, 237-282.
- Bennett, M. J., Schlunegger, M. P. & Eisenberg, D. (1995) 3D domain swapping: a mechanism for oligomer assembly. *Protein Sci.* **4**, 2455-2468.
- Birnboim, H. C. & Doly, J. (1979) A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucleic Acids Res.* **7**, 1513-1523.
- Bohlool, B. B. & Schmidt, E. L. (1974) Lectins: a possible basis for specificity in the *Rhizobium*-legume root nodule symbiosis. *Science* **185**, 269-71.
- Bourne, Y., Roussel, A., Frey, M., Rouge, P., Fontecilla-Camps, J. C. & Cambillau, C. (1990) Three-dimensional structures of complexes of *Lathyrus ochrus* isolectin-I with glucose and mannose - fine specificity of the monosaccharide-binding site. *Proteins* **8**, 365-376.
- Bourne, Y., Mazurier, J., Legrand, D., Rouge, P., Montreuil, J., Spik, G. & Cambillau, C. (1994) Structures of a legume lectin complexed with the human lactotransferrin n2 fragment, and with an isolated biantennary glycopeptide - role of the fucose moiety. *Structure* **2**, 209-219.

-
- Boyd, W. C. (1963) The lectins: their present status. *Vox Sang.* **8**, 1-32.
- Brake, A. J., Merryweather, J. P., Coit, D. G., Heberlein, U. A., Masiarz, F. R., Mullenbach, G. T., Urdea, M. S., Valenzuela, P. & Barr, P. J. (1984) α -Factor-directed synthesis and secretion of mature foreign proteins in *S. cerevisiae*. *Proc. Natl. Acad. Sci. U. S. A.* **81**, 4642-4646.
- Brake, A. (1989) Secretion of heterologous proteins directed by the yeast alpha-factor leader. *Yeast Genetic Engineering*. P. J. Barr, A. J. Brake & P. Valenzuela. Stoneham, MA, Butterworth Publishers: 269-280.
- Broekaert, W. F., Lambrechts, D., Verbelen, J. P. & Peumans, W. J. (1988) *Datura stramonium* agglutinin - location in the seed and release upon imbibition. *Plant Physiol.* **86**, 569-574.
- Broekaert, W. F., Van Parijs, J., Leyns, F., Joos, H. & Peumans, W. J. (1989) A chitin-binding lectin from stinging nettle rhizomes with antifungal properties. *Science* **245**, 1100-1102.
- Cammue, B. P. A., Peeters, B. & Peumans, W. J. (1986) A new lectin from tulip (*Tulipa*) bulbs. *Planta* **169**, 583-588.
- Casaloungue, C. & Lezica, R. P. (1985) Potato lectin - a cell wall glycoprotein. *Plant Cell Physiol.* **26**, 1533-1539.
- Cereghino, J. L. & Cregg, J. M. (2000) Heterologous protein expression in the methylotrophic yeast *Pichia pastoris*. *FEMS Microbiol. Rev.* **24**, 45-66.

-
- Chandra, N. R., Ramachandraiah, G., Bachhawat, K., Dam, T. K., Surolia, A. & Vijayan, M. (1999) Crystal structure of a dimeric mannose-specific agglutinin from garlic: Quaternary association and carbohydrate specificity. *J. Mol. Biol.* **285**, 1157-1168.
- Chantalat, L., Wood, S. D., Rizkallah, P. & Reynolds, C. D. (1996) X-ray structure solution of amaryllis lectin by molecular replacement with only 4% of the total diffracting matter. *Acta Crystallogr. D.* **52**, 1146-1152.
- Chao, Q. & Etzler, M. E. (1994) Incorrect targeting of plant vacuolar lectins in yeast. *J. Biol. Chem.* **269**, 20866-20871.
- Chervenak, M. C. & Toone, E. J. (1994) A direct measure of the contribution of solvation to the thermodynamics of binding. *Biophys. J.* **66**, 359.
- Chervenak, M. C. & Toone, E. J. (1995) Calorimetric analysis of the binding of lectins with overlapping carbohydrate-binding ligand specificities. *Biochemistry* **34**, 5685-5695.
- Chrispeels, M. J. & Raikhel, N. V. (1991) Lectins, lectin genes, and their role in plant defense. *Plant Cell* **3**, 1-9.
- Clare, J., Rayment, F., Ballantine, S., Sreekrishna, K. & Romanos, M. (1991) High level expression of tetanus toxin fragment C in *Pichia pastoris* strains containing multiple tandem integrations of the gene. *Bio-Technol.* **9**, 455-460.
- Collinge, D. B., Kragh, K. M., Mikkelsen, J. D., Nielsen, K. K., Rasmussen, U. & Vad, K. (1993) Plant Chitinases. *Plant J.* **3**, 31-40.

-
- Cormack, B. P., Bertram, G., Egerton, M., Gow, N. A. R., Falkow, S. & Brown, A. J. P. (1997) Yeast-enhanced green fluorescent protein (yEGFP): A reporter of gene expression in *Candida albicans*. *Microbiol.-UK* **143**, 303-311.
- Cramer, A., Whitehorn, E. A., Tate, E. & Stemmer, W. P. C. (1996) Improved green fluorescent protein by molecular evolution using DNA shuffling. *Nat. Biotechnol.* **14**, 315-319.
- Cregg, J. M., Vedvick, T. S. & Raschke, W. C. (1993) Recent advances in the expression of foreign genes in *Pichia pastoris*. *Bio/Technol.* **11**, 905-910.
- Czapla, T. H. & Lang, B. A. (1990) Effect of plant lectins on the larval development of European Corn borer (Lepidoptera, Pyralidae) and Southern Corn Rootworm (Coleoptera, Chrysomelidae). *J. Econ. Entomol.* **83**, 2480-2485.
- Dao-Thi, M. H., Hamelryck, T. W., Poortmans, F., Voelker, T. A., Chrispeels, M. J. & Wyns, L. (1996) Crystallization of glycosylated and non-glycosylated phytohemagglutinin-L. *Proteins* **24**, 134-137.
- Dao-Thi, M. H., Rizkallah, P., Wyns, L., Poortmans, F. & Loris, R. (1998) Quaternary structure of UEA-II, the chitobiose specific lectin from gorse. *Acta Crystallogr. D.* **54**, 844-847.
- Debray, H., Decout, D., Strecker, G., Spik, G. & Montreuil, J. (1981) Specificity of 12 lectins towards oligosaccharides and glycopeptides related to N-glycosylproteins. *Eur. J. Biochem.* **117**, 41-55.
- Del Campillo, E., Howard, J. & Shannon, L. M. (1981) Interaction of soybean agglutinin with soybean callus cells. *Z. Pflanzenphysiol.* **104**, 97-102.

-
- Delbaere, L. T. J., Vandonselaar, M., Prasad, L., Quail, J. W., Wilson, K. S. & Dauter, Z. (1993) Structures of the lectin IV of *Griffonia simplicifolia* and its complex with the Lewis B human blood group determinant at 2.0 angstrom resolution. *J. Mol. Biol.* **230**, 950-965.
- Dessen, A., Gupta, D., Sabesan, S., Brewer, C. F. & Sacchettini, J. C. (1995) X-ray crystal-structure of the soybean agglutinin cross-linked with a biantennary analog of the blood-group-i carbohydrate antigen. *Biochemistry* **34**, 4933-4942.
- Dombrowski, J. E., Schroeder, M. R., Bednarek, S. Y. & Raikhel, N. V. (1993) Determination of the functional elements within the vacuolar targeting signal of barley lectin. *Plant Cell* **5**, 587-596.
- Drickamer, K. (1995) Multiplicity of lectin-carbohydrate interactions. *Nat. Struct. Biol.* **2**, 437-439.
- Drickamer, K. (1999) C-type lectin-like domains. *Curr. Opin. Struct. Biol.* **9**, 585-590.
- Edelman, G. M. & Wang, J. L. (1978) Binding and functional properties of concanavalin A and its derivatives. III. Interactions with indoleacetic acid and other hydrophobic ligands. *J. Biol. Chem.* **253**, 3016-22.
- Efimov, A. V. (1994) Common structural motifs in small proteins and domains. *FEBS Lett.* **355**, 213-219.

-
- Eisemann, C. H., Donaldson, R. A., Pearson, R. D., Cadagon, T., Vuocolo, T. & Tellam, R. L. (1994) Larvicidal activity of lectins on *Lucilia cuprina*: Mechanism of action. *Entomol. Exp. Appl.* **72**, 1-11.
- Elgavish, S. & Shaanan, B. (1997) Lectin-carbohydrate interactions: different folds, common recognition principles. *Trends Biochem. Sci.* **22**, 462-467.
- Elgavish, S. & Shaanan, B. (1998) Structures of the *Erythrina corallodendron* lectin and its complexes with mono- and disaccharides. *J. Mol. Biol.* **277**, 917-932.
- Emsley, J., White, H. E., Ohara, B. P., Oliva, G., Srinivasan, N., Tickle, I. J., Blundell, T. L., Pepys, M. B. & Wood, S. P. (1994) Structure of pentameric human serum amyloid-p component. *Nature* **367**, 338-345.
- Etzler, M. E. & Kabat, E. A. (1970) Purification and characterization of a lectin (plant hemagglutinin) with blood group A specificity from *Dolichos biflorus*. *Biochemistry* **9**, 869-877.
- Etzler, M. E. (1986) Distribution and function of plant lectins. The lectins: Properties, Functions, and Applications in Biology and Medicine. I. E. Liener, N. Sharon & I. J. Goldstein. Orlando, USA, Academic Press: 371-435.
- Ferrari, E., Lodi, T., Sorbi, R., Tirindelli, R. & Cavaggioni, A. (1997) Expression of a lipocalin in *Pichia pastoris*: secretion, purification, and binding activity of a recombinant mouse major urinary protein. *FEBS Lett.* **401**, 73-77.
- Fitches, E., Gatehouse, A. M. R. & Gatehouse, J. A. (1997) Effects of snowdrop lectin (GNA) delivered via artificial diet and transgenic plants on the

-
- development of tomato moth (*Lacanobia oleracea*) larvae in laboratory and glasshouse trials. *J. Insect Physiol.* **43**, 727-739.
- Fitches, E. & Gatehouse, J. A. (1998) A comparison of the short and long term effects of insecticidal lectins on the activities of soluble and brush border enzymes of tomato moth larvae (*Lacanobia oleracea*). *J. Insect Physiol.* **44**, 1213-1224.
- Frigerio, L., Vitale, A., Lord, J. M., Ceriotti, A. & Roberts, L. M. (1998) Free ricin A chain, proricin, and native toxin have different cellular fates when expressed in tobacco protoplasts. *J. Biol. Chem.* **273**, 14194-14199.
- Gatehouse, A. M. R., Dewey, F. M., Dove, J., Fenton, K. A. & Pusztai, A. (1984) Effect of seed lectins from *Phaseolus vulgaris* on the development of larvae of *Callosobruchus maculatus* - mechanism of toxicity. *J. Sci. Food Agric.* **35**, 373-380.
- Gatehouse, A. M. R., Hilder, V. A. & Gatehouse, J. A. (1992) Control of insect pests by plant genetic-engineering. *P. Roy. Soc. Edinb. B* **99**, 51-60.
- Gatehouse, A. M. R., Powell, K. S., Peumans, W. J., Van Damme, E. J. M. & Gatehouse, J. A. (1995) Insecticidal properties of plant lectins: their potential in plant protection. *Lectins: Biomedical Perspectives*. A. Pusztai & S. Bardocz. London, Taylor & Francis: 35-57.
- Gething, M. J. & Sambrook, J. (1992) Protein folding in the cell. *Nature* **355**, 33-45.
- Gietz, R. D. & Schiestl, R. H. (1995) Transforming yeast with DNA. *Methods Mol. Cell. Biol.* **5**, 255-269.

-
- Goldstein, I. J. & Hayes, C. E. (1978) The lectins: carbohydrate-binding proteins of plants and animals. *Adv. Carbohydr. Chem. Biochem.* **35**, 127-340.
- Goldstein, I. J. & Poretz, R. D. (1986) Isolation, physicochemical characterization, and carbohydrate-binding specificity of lectins. The Lectins. Properties, Functions, and Applications in Biology and Medicine. I. E. Liener, N. Sharon & I. J. Goldstein. London/Orlando, Academic: 35-247.
- Grinna, L. S. & Tschopp, J. F. (1989) Size distribution and general structural features of N-linked oligosaccharides from the methylotrophic yeast *Pichia pastoris*. *Yeast* **5**, 107-115.
- Haasz, D., Frey, R., Thiesen, M. & Kauss, H. (1981) Partial purification of a hemagglutinin associated with cell walls from hypocotyls of *Vigna radiata*. *Planta* **151**, 490-496.
- Habibi, J., Backus, E. A. & Czaplak, T. H. (1993) Plant lectins affect survival of the potato leaf hopper (Homoptera, Cicadellidae). *J. Econ. Entomol.* **86**, 945-51.
- Hamelryck, T. W., DaoThi, M. H., Poortmans, F., Chrispeels, M. J., Wyns, L. & Loris, R. (1996) The crystallographic structure of phytohemagglutinin-L. *J. Biol. Chem.* **271**, 20479-20485.
- Hamelryck, T. W., Loris, R., Bouckaert, J., DaoThi, M. H., Strecker, G., Imberty, A., Fernandez, E., Wyns, L. & Etzler, M. E. (1999) Carbohydrate binding, quaternary structure and a novel hydrophobic binding site in two legume lectin oligomers from *Dolichos biflorus*. *J. Mol. Biol.* **286**, 1161-1177.

-
- Hammarstrom, S., Hammarstrom, M. L., Sundblad, G., Arnarp, J. & Lonngren, J. (1982) Mitogenic leukoagglutinin from *Phaseolus vulgaris* binds to a pentasaccharide unit in N-acetyllactosamine-type glycoprotein glycans. *Proc. Natl. Acad. Sci. U. S. A.* **79**, 1611-1615.
- Harper, S. M., Crenshaw, R. W., Mullins, M. A. & Privalle, L. S. (1995) Lectin binding to insect brush border membranes. *J. Econ. Entomol.* **88**, 1197-1202.
- Harper, S. M., Hopkins, T. L. & Czapla, T. H. (1998) Effect of wheat germ agglutinin on formation and structure of the peritrophic membrane in European corn borer (*Ostrinia nubilalis*) larvae. *Tissue Cell* **30**, 166-176.
- Hazes, B. (1996) The (QxW)(3) domain: A flexible lectin scaffold. *Protein Sci.* **5**, 1490-1501.
- Hester, G., Kaku, H., Goldstein, I. J. & Wright, C. S. (1995) Structure of mannose-specific snowdrop (*Galanthus nivalis*) lectin is representative of a new plant lectin family. *Nat. Struct. Biol.* **2**, 472-479.
- Hester, G. & Wright, C. S. (1996) The Mannose-specific bulb lectin from *Galanthus nivalis* (snowdrop) binds mono- and dimannosides at distinct sites. Structure analysis of refined complexes at 2.3 angstrom and 3.0 angstrom resolution. *J. Mol. Biol.* **262**, 516-531.
- Higgins, D. R. & Cregg, J. M. (1998) *Pichia* Protocols. Totowa, New Jersey, Humana Press.
- Hilder, V. A., Powell, K. S., Gatehouse, A. M. R., Gatehouse, J. A., Gatehouse, L. N., Shi, Y., Hamilton, W. D. O., Merryweather, A., Newell, C. A., Timans, J. C.,

-
- Peumans, W. J., Van Damme, E. J. M. & Boulter, D. (1995) Expression of snowdrop lectin in transgenic tobacco plants results in added protection against aphids. *Transgenic Res.* **4**, 18-25.
- Hoffman, L. M. & Donaldson, D. D. (1985) Characterization of 2 *Phaseolus vulgaris* phytohemagglutinin genes closely linked on the chromosome. *EMBO J.* **4**, 883-889.
- Holwerda, B. C., Padgett, H. S. & Rogers, J. C. (1992) Proaleurain vacuolar targeting is mediated by short contiguous peptide interactions. *Plant Cell* **4**, 307-318.
- Howard, I. K., Sage, H. J. & Horton, C. B. (1972) Studies on the appearance and location of hemagglutinins from a common lentil during the life cycle of the plant. *Arch. Biochem. Biophys.* **149**, 323-26.
- Howard, J., Shannon, L., Oki, L. & Murashige, T. (1977) Soybean agglutinin. A mitogen for soybean callus cells. *Exp. Cell Res.* **107**, 448-50.
- Huesing, J. E., Murdoch, L. L. & Shade, R. E. (1991) Rice and stinging nettle lectins - insecticidal activity similar to wheat germ agglutinin. *Phytochemistry* **30**, 3565-3568.
- Huesing, J. E., Shade, R. E., Chrispeels, M. J. & Murdoch, L. L. (1991) Alpha amylase inhibitor, not phytohemagglutinin, explains resistance of common bean seeds to cowpea weevil. *Plant Physiol.* **96**, 993-996.
- Janzen, D. H., Juster, H. B. & Liener, I. E. (1976) Insecticidal action of the phytohemagglutinin in black bean on a bruchid beetle. *Science* **192**, 795-96.

-
- Juge, N., Andersen, J., Tull, D., Roepstorff, P. & Svensson, B. (1996) Overexpression, purification, and characterization of recombinant barley alpha-amylases 1 and 2 secreted by the methylotrophic yeast *Pichia pastoris*. *Protein Expr. Purif.* **8**, 204-214.
- Kabir, S. & Daar, A. S. (1994) The composition and properties of jacalin, a lectin of diverse applications obtained from the jackfruit (*Artocarpus heterophyllus*) seeds. *Immunol. Invest.* **23**, 167-188.
- Kaku, H. & Goldstein, I. J. (1989) Snowdrop lectin. *Methods Enzymol.* **179**, 327-331.
- Kaku, H., Goldstein, I. J. & Oscarson, S. (1991) Interactions of 5 D-mannose-specific lectins with a series of synthetic branched trisaccharides. *Carbohydr. Res.* **213**, 109-116.
- Kaku, H. & Goldstein, I. J. (1992) Interaction of linear manno-oligosaccharides with 3 mannose-specific bulb lectins - comparison with mannose glucose-binding lectins. *Carbohydr. Res.* **229**, 337-346.
- Kauss, H. & Glaser, C. (1974) Carbohydrate-binding proteins from plant cell walls and their possible involvement in extension growth. *FEBS Lett.* **45**, 304-7.
- Kilpatrick, D. C., Yeoman, M. M. & Gould, A. R. (1979) Tissue and subcellular distribution of the lectin from *Datura stramonium* (thorn apple). *Biochem. J.* **184**, 215-219.
- Kilpatrick, D. C. (1980) Isolation of a lectin from the pericarp of potato (*Solanum tuberosum*) fruits. *Biochem. J.* **191**, 273-275.

-
- Kinzy, W. R., Belenky, D. M., Loontjens, F. G. & Sharon, N. (1992) Synthesis of methyl alpha-N-dansyl-D-galactosaminides and beta-N-dansyl-D-galactosaminides, probes for the combining sites of N-acetyl-D-galactosamine-specific lectins. *Glycoconj. J.* **9**, 225-227.
- Kjeldsen, T., Hach, M., Balschmidt, P., Havelund, S., Pettersson, A. & Markussen, J. (1998) Prepro-leaders lacking N-linked glycosylation for secretory expression in the yeast *Saccharomyces cerevisiae*. *Protein Expr. Purif.* **14**, 309-316.
- Knibbs, R., Goldstein, I. J., Ratcliff, R. M. & Shibuya, M. D. (1991) Characterization of the carbohydrate binding specificity of the leukoagglutinating lectin from *Maackia amurensis*. *J. Biol. Chem.* **266**, 83-88.
- Knox, R. B., Clarke, A., Harrison, S., Smith, P. & Marchalonis, J. J. (1976) Cell recognition in plants: Determinants of the stigma surface and their pollen interactions. *Proc. Natl. Acad. Sci. U. S. A.* **73**.
- Kocourek, J. & Horejsi, V. (1983) A note on the recent discussion on definition of the term 'lectin'. *Lectins: Biology, Biochemistry and Clinical Biochemistry*. T. C. Bog-Hansen & G. A. Spengler. Berlin, Walter de Gruyter. **3**: 3-6.
- Kornfeld, K., Reitman, M. L. & Kornfeld, R. (1981) The carbohydrate-binding specificity of pea and lentil lectins - fucose is an important determinant. *J. Biol. Chem.* **256**, 6633-6640.
- Kozak, M. (1987) An analysis of 5'-noncoding sequences from 699 vertebrate messenger RNAs. *Nucleic Acids Res.* **15**, 8125-8148.

-
- Kozak, M. (1990) Downstream secondary structure facilitates recognition of initiator codons by eukaryotic ribosomes. *Proc. Natl. Acad. Sci. U. S. A.* **87**, 8301-8305.
- Kurjan, J. & Herskowitz, I. (1982) Structure of a yeast pheromone gene (*MF α*): a putative α -factor precursor contains four tandem repeats of mature α -factor. *Cell* **30**, 933-943.
- Laroche, Y., Storme, V., de Meutter, J., Messens, J. & Lauwreys, M. (1994) High level secretion and very efficient isotopic labelling of tick anticoagulant peptide (TAP) expressed in the methylotrophic yeast *Pichia pastoris*. *Bio/Technol.* **12**, 1119-1124.
- Lerouge, P., Roche, P., Faucher, C., Maillet, F., Truchet, G., Prome, J. C. & Denarie, J. (1990) Symbiotic host-specificity of *Rhizobium meliloti* is determined by a sulfated and acylated glucosamine oligosaccharide signal. *Nature* **344**, 781-784.
- Lerouge, P., Cabanes-Macheteau, M., Rayon, C., Fischette-Laine, A., Gomord, V. & Faye, L. (1998) N-Glycoprotein biosynthesis in plants: recent developments and future trends. *Plant Mol. Biol.* **38**, 31-48.
- Li, J. D., Carroll, J. & Ellar, D. J. (1991) Crystal structure of insecticidal delta-endotoxin from *Bacillus thuringiensis* at 2.5 angstrom resolution. *Nature* **353**, 815-821.
- Lis, H. & Sharon, N. (1977) Lectins: their chemistry and application to immunology. The Antigens. M. Sela. New York, Academic. **4**: 429-529.

-
- Lis, H. & Sharon, N. (1981) Lectins in higher plants. *The Biochemistry of Plants*. A. Marcus. New York, Academic: 372-447.
- Lis, H., Joubert, F. J. & Sharon, N. (1985) Isolation and properties of N-acetyllactosamine-specific lectins from 9 *Erythrina* species. *Phytochemistry* **24**, 2803-2809.
- Longstaff, M., Powell, K. S., Gatehouse, J. A., Raemaekers, R., Nevell, C. A. & Hamilton, W. D. O. (1998) Production and purification of active snowdrop lectin in *Escherichia coli*. *Eur. J. Biochem.* **252**, 59-65.
- Loris, R., Hamelryck, T., Bouckaert, J. & Wyns, L. (1998) Legume lectin structure. *B. B. A.-Protein Struct. M.* **1383**, 9-36.
- Machuka, J., Van Damme, E. J. M., Peumans, W. J. & Jackai, L. E. N. (1999) Effect of plant lectins on larval development of the legume pod borer, *Maruca vitrata*. *Entomol. Exp. Appl.* **93**, 179-187.
- Martinet, W., Saelens, X., Deroo, T., Neiryneck, S., Contreras, R., Min Jou, W. & Fiers, W. (1997) Protection of mice against a lethal influenza challenge by immunization with yeast-derived recombinant influenza neuraminidase. *Eur. J. Biochem.* **247**, 332-338.
- Martinoia, E., Heck, U., Boller, T. & Wiemken, A. (1979) Some properties of vacuoles isolated from *Neurospora crassa* slime variant. *Arch. Microbiol.* **120**, 31-34.

-
- Matsuoka, K. & Nakamura, K. (1991) Propeptide of a precursor to a plant vacuolar protein required for vacuolar targeting. *Proc. Natl. Acad. Sci. U. S. A.* **88**, 834-838.
- Matsuoka, K., Watanabe, N. & Nakamura, K. (1995) O-glycosylation of a precursor to a sweet potato vacuolar protein, sporamin, expressed in tobacco cells. *Plant J.* **8**, 877-889.
- Miller, J. B., Hsu, R., Heinrikson, R. & Yachnin, S. (1975) Extensive homology between the subunits of phytohemagglutinin mitogenic proteins derived from *Phaseolus vulgaris*. *Proc. Natl. Acad. Sci. U. S. A.* **72**, 1388-1391.
- Min, W., Dunn, A. J. & Jones, D. H. (1992) Nonglycosylated recombinant pro-concanavalin A is active without polypeptide cleavage. *EMBO J.* **11**, 1303-1307.
- Mirelman, D., Galun, E., Sharon, N. & Lotan, R. (1975) Inhibition of fungal growth by wheat germ agglutinin. *Nature* **256**, 414-16.
- Mirkov, T. E. & Chrispeels, M. J. (1993) Mutation of Asn(128) to Asp of *Phaseolus vulgaris* leucoagglutinin (PHA-I) eliminates carbohydrate-binding and biological activity. *Glycobiology* **3**, 581-587.
- Montesino, R., Garcia, R., Quintero, O. & Cremata, J. A. (1998) Variation in N-linked oligosaccharide structures on heterologous proteins secreted by the methylotrophic yeast *Pichia pastoris*. *Protein Expr. Purif.* **14**, 197-207.

-
- Morris, R. O., Bilyeu, K. D., Laskey, J. G. & Cheikh, N. N. (1999) Isolation of a gene encoding a glycosylated cytokinin oxidase from maize. *Biochem. Biophys. Res. Commun.* **255**, 328-333.
- Murdoch, L. L., Huesing, J. E., Nielsen, S. S., Pratt, R. C. & Shade, R. E. (1990) Biological effects of plant lectins on the cowpea weevil. *Phytochemistry* **29**, 85-89.
- Murzin, A. G., Lesk, A. M. & Chothia, C. (1992) Beta-trefoil fold patterns of structure and sequence in the kunitz inhibitors interleukins 1-beta and 1-alpha and fibroblast growth factors. *J. Mol. Biol.* **223**, 531-543.
- Nagahora, H., Ishikawa, K., Niwa, Y., Muraki, M. & Jigami, Y. (1992) Expression and secretion of wheat germ agglutinin by *Saccharaomyces cerevisiae*. *Eur. J. Biochem.* **210**, 989-997.
- Nagl, W. (1972) Phytohemagglutinin: transitory enhancement of growth in *Phaseolus* and *Allium*. *Planta* **106**, 269-72.
- Naismith, J. H., Emmerich, C., Habash, J., Harrop, S. J., Helliwell, J. R., Hunter, W. N., Raftery, J., Kalb, A. J. & Yariv, J. (1994) Refined structure of concanavalin A complexed with alpha-D-mannopyrannoside at 2.0 angstrom resolution and comparison with the saccharide-free structure. *Acta Crystallogr. D.* **50**, 847-858.
- Naismith, J. H. & Field, R. A. (1996) Structural basis of trimannoside recognition by concanavalin A. *J. Biol. Chem.* **271**, 972-976.

-
- Nakamura, K. & Matsuoka, K. (1993) Protein targeting to the vacuole in plant cells. *Plant Physiol.* **101**, 1-5.
- Neuhaus, J., Pietrzak, M. & Boller, T. (1994) Mutation analysis of the C-terminal vacuolar targeting peptide of tobacco chitinase: low specificity of the sorting system, and gradual transition between intracellular retention and secretion into the extracellular space. *Plant J.* **5**, 45-54.
- Neuhaus, J. M. & Rogers, J. C. (1998) Sorting of proteins to vacuoles in plant cells. *Plant Mol. Biol.* **38**, 127-144.
- Nishiguchi, M., Yoshida, K., Sumizono, T. & Tazaki, K. (1997) Studies by site-directed mutagenesis of the carbohydrate-binding properties of a bark lectin from *Robinia pseudoacacia*. *FEBS Lett.* **403**, 294-298.
- O'Donohue, M., Boissy, G., Huet, J. C., Nespoulus, C., Brunie, S. & Pernollet, J. C. (1996) Overexpression in *Pichia pastoris* and crystallization of an elicitor protein secreted by the phytopathogenic fungus, *Phytophthora cryptogea*. *Protein Expr. Purif.* **8**, 254-261.
- Peumans, W. J., Stinissen, H. M. & Carlier, A. R. (1983) Speculations about a physiological role of some plant lectins. Lectins: Biology, Biochemistry, and Clinical Biochemistry. T. C. Bog-Hansen & G. A. Spengler. Berlin, Walter de Gruyter. **3**: 583-92.
- Peumans, W. J., Deley, M. & Broekaert, W. F. (1984) Isolation and partial characterization of a small lectin from *Urtica dioica* rhizomes. *Arch. Int. Physiol. Bio.* **92**, B156-B157.

-
- Peumans, W. J., Nsimbalubaki, M., Peeters, B. & Broekaert, W. F. (1985) Isolation and partial characterization of a lectin from ground elder (*Aegopodium podagraria*) rhizomes. *Planta* **164**, 75-82.
- Peumans, W. J. & Van Damme, E. J. M. (1995) Lectins as plant defense proteins. *Plant Physiol.* **109**, 347-352.
- Peumans, W. J., Verhaert, P., Pfuller, U. & Van Damme, E. J. M. (1996) Isolation and partial characterization of a small chitin-binding lectin from mistletoe (*Viscum album*). *FEBS Lett.* **396**, 261-265.
- Powell, K. S., Gatehouse, A. M. R., Hilder, V. A. & Gatehouse, J. A. (1993) Antimetabolic effects of plant-lectins and plant and fungal enzymes on the nymphal stages of 2 important rice pests, *Nilaparvata lugens* and *Nephotettix cinciteps*. *Entomol. Exp. Appl.* **66**, 119-126.
- Powell, K. S., Gatehouse, A. M. R., Hilder, V. A. & Gatehouse, J. A. (1995a) Antifeedant effects of plant-lectins and an enzyme on the adult stage of the rice brown planthopper, *Nilaparvata lugens*. *Entomol. Exp. Appl.* **75**, 51-59.
- Powell, K. S., Gatehouse, A. M. R., Hilder, V. A., Van Damme, E. J. M., Peumans, W. J., Boonjawat, J., Horsham, K. & Gatehouse, J. A. (1995b) Different antimetabolic effects of related lectins towards nymphal stages of *Nilaparvata lugens*. *Entomol. Exp. Appl.* **75**, 61-65.
- Powell, K. S., Spence, J., Bharathi, M., Gatehouse, J. A. & Gatehouse, A. M. R. (1998) Immunohistochemical and developmental studies to elucidate the

-
- mechanism of action of the snowdrop lectin on the rice brown planthopper, *Nilaparvata lugens* (Stal). *J. Insect Physiol.* **44**, 529-539.
- Prabu, M. M., Sankaranarayanan, R., Puri, K. D., Sharma, V., Surolia, A., Vijayan, M. & Suguna, K. (1998) Carbohydrate specificity and quaternary association in basic winged bean lectin: X-ray analysis of the lectin at 2.5 angstrom resolution. *J. Mol. Biol.* **276**, 787-796.
- Pusztai, A., Ewen, S. W. B., Grant, G., Peumans, W. J., Van Damme, E. J. M., Rubio, L. & Bardocz, S. (1990) Relationship between survival and binding of plant-lectins during small intestinal passage and their effectiveness as growth factors. *Digestion* **46**, 308-316.
- Pusztai, A. (1991) *Plant Lectins*. Cambridge, UK, Cambridge University Press.
- Pusztai, A. & Bardocz, S. (1996) Biological effects of plant lectins on the gastrointestinal tract: Metabolic consequences and applications. *Trends Glycosci. Glyc.* **8**, 149-165.
- Rahbe, Y. & Febvay, G. (1993) Protein toxicity to aphids - an in vitro test on *Acyrtosiphon pisum*. *Entomol. Exp. Appl.* **67**, 149-160.
- Rahbe, Y., Sauvion, N., Febvay, G., Peumans, W. J. & Gatehouse, A. M. R. (1995) Toxicity of lectins and processing of ingested proteins in the pea aphid *Acyrtosiphon pisum*. *Entomol. Exp. Appl.* **76**, 143-155.
- Raikhel, N. V., Lee, H. I. & Broekaert, W. F. (1993) Structure and function of chitin-binding proteins. *Annu. Rev. Plant Phys.* **44**, 591-615.

-
- Raina, A. & Datta, A. (1992) Molecular cloning of a gene encoding a seed-specific protein with nutritionally balanced amino acid composition from *Amaranthus*. *Proc. Natl. Acad. Sci. U. S. A.* **89**, 11774-11778.
- Ravishankar, R., Ravindran, M., Suguna, K., Surolia, A. & Vijayan, M. (1997) Crystal structure of the peanut lectin-T-antigen complex. Carbohydrate specificity generated by water bridges. *Curr. Sci.* **72**, 855-861.
- Reverter, D., Ventura, S., Villegas, V., Vendrell, J. & Aviles, F. (1998) Overexpression of human procarboxypeptidase A2 in *Pichia pastoris* and detailed characterization of its activation pathway. *J. Biol. Chem.* **273**, 3535-3541.
- Ribeiro, C., Simoes, N. & Brehelin, M. (1996) Insect immunity: The haemocytes of the armyworm *Mythimna unipuncta* (Lepidoptera: Noctuidae) and their role in defense reactions. In vivo and in vitro studies. *J. Insect Physiol.* **42**, 815-822.
- Rinderle, S. J., Goldstein, I. J. & Remsen, E. E. (1990) Physicochemical properties of amaranthin, the lectin from *Amaranthus caudatus* seeds. *Biochemistry* **29**, 10555-10561.
- Rini, J. M., Hardman, K. D., Einspahr, H., Suddath, F. L. & Carver, J. P. (1993) X-ray crystal structure of a pea lectin - triamnoside complex at 2.6 angstrom resolution. *J. Biol. Chem.* **268**, 10126-10132.
- Rini, J. M. (1995) Lectin structure. *Annu. Rev. Biophys. Biomol. Struct.* **24**, 551-577.
- Rini, J. M. & Lobsanov, Y. D. (1999) New animal lectin structures. *Curr. Opin. Struct. Biol.* **9**, 578-584.

-
- Roberts, D. D. & Goldstein, I. J. (1983) Adenine binding sites of the lectin from lima beans (*Phaseolus lunatus*). *J. Biol. Chem.* **258**, 13820-24.
- Romanos, M. A., Scorer, C. A. & Clare, J. J. (1992) Foreign gene expression in yeast: a review. *Yeast* **8**, 423-488.
- Rosa, J. C., De Olivera, P. S. L., Garratt, R., Beltramini, L., Resing, K., Roque-Barreira, M. C. & Greene, L. J. (1999) KM+, a mannose-binding lectin from *Artocarpus integrifolia*: Amino acid sequence, predicted tertiary structure, carbohydrate recognition, and analysis of the beta-prism fold. *Protein Sci.* **8**, 13-24.
- Rothman, J. H., Yamashiro, C. T., Kane, P. M. & Stevens, T. H. (1989) Protein targeting to the yeast vacuole. *Trends Biochem. Sci.* **14**, 347-350.
- Rutenber, R. & Robertus, J. D. (1991) Structure of ricin B-chain at 2.5 angstrom resolution. *Proteins* **10**, 260-269.
- Sabnis, D. D. & Hart, J. W. (1978) The isolation and some properties of a lectin (hemagglutinin) from *Cucurbita phloem* exudate. *Planta* **142**, 97-101.
- Samaddar, M., Catterall, J. & Dighe, R. (1997) Expression of biologically active b subunit of bovine follicle-stimulating hormone in the methylotrophic yeast *Pichia pastoris*. *Protein Expr. Purif.* **10**, 345-355.
- Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989) Molecular cloning: a laboratory manual. Cold Spring Harbor NY, Cold Spring Harbor Laboratories.

-
- Sanger, F., Nicklen, S. & Coulsen, A. R. (1977) DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. U. S. A.* **74**, 5463-5467.
- Sankaranarayanan, R., Sekar, K., Banerjee, B., Sharma, V., Surolia, A. & Vijayan, M. (1996) A novel mode of carbohydrate recognition in jacalin, a *Moraceae* plant lectin with a beta-prism fold. *Nat. Struct. Biol.* **3**, 596-603.
- Sarkar, G. & Sommer, S. (1990) The megaprimer method of site-directed mutagenesis. *Biotechniques* **8**, 404-407.
- Sauerborn, M. K., Wright, L. M., Reynolds, C. D., Grossmann, J. G. & Rizkallah, P. J. (1999) Insights into carbohydrate recognition by *Narcissus pseudonarcissus* lectin: The crystal structure at 2 angstrom resolution in complex with alpha 1-3 manno-*bio*se. *J. Mol. Biol.* **290**, 185-199.
- Sauvion, N., Rahbe, Y., Peumans, W. J., Van Damme, E. J. M., Gatehouse, J. A. & Gatehouse, A. M. R. (1996) Effects of GNA and other mannose binding lectins on development and fecundity of the peach-potato aphid *Myzus persicae*. *Entomol. Exp. Appl.* **79**, 285-293.
- Scorer, C., Buckholz, R. G., Clare, J. & Romanos, M. (1993) The intracellular production and secretion of HIV-1 envelope protein in methylotrophic yeast *Pichia pastoris*. *Gene* **139**, 111-119.
- Shaanan, B., Lis, H. & Sharon, N. (1991) Structure of a legume lectin with an ordered N-linked carbohydrate in complex with lactose. *Science* **254**, 862-866.

-
- Sharma, V. & Surolia, A. (1997) Analyses of carbohydrate recognition by legume lectins: Size of the combining site loops and their primary specificity. *J. Mol. Biol.* **267**, 433-445.
- Sharon, N. & Lis, H. (1990) Legume lectins - a large family of homologous proteins. *FASEB J.* **4**, 3198-3208.
- Sharon, N. (1993) Lectin carbohydrate complexes of plant and animals - an atomic view. *Trends Biochem. Sci.* **18**, 221-226.
- Sheldon, P. & Bowles, D. (1992) The glycoprotein precursor of concanavalin A is converted to an active lectin by deglycosylation. *EMBO J.* **11**, 1297-1301.
- Shibuya, N., Goldstein, I. J., Broekaert, W. F., Nsimbalubaki, M., Peeters, B. & Peumans, W. J. (1987) The elderberry (*Sambucus nigra*) bark lectin recognizes the Neu5Ac(alpha-2-6)Gal-GalNAc sequence. *J. Biol. Chem.* **262**, 1596-1601.
- Shibuya, N., Goldstein, I. J., Van Damme, E. J. M. & Peumans, W. J. (1988) Binding properties of a mannose-specific lectin from the snowdrop (*Galanthus nivalis*) bulb. *J. Biol. Chem.* **263**, 728-734.
- Smeets, K., Van Damme, E. J. M. & Peumans, W. J. (1994) Comparative-study of the posttranslational processing of the mannose-binding lectins in the bulbs of garlic (*Allium sativum*) and ramsons (*Allium ursinum*). *Glycoconj. J.* **11**, 309-320.

-
- Smeets, K., Van Damme, E. J. M. & Peumans, W. J. (1997) Developmental regulation of lectin and alliinase synthesis in garlic bulbs and leaves. *Plant Physiol.* **113**, 765-771.
- Sreekrishna, K., Brankamp, R., Kropp, K., Blankenship, D., Tsay, J. T., Smith, P., Wierschke, J., Subramaniam, A. & Birkenberger, L. (1997) Strategies for the optimal synthesis and secretion of heterologous proteins in the methylotrophic yeast *Pichia pastoris*. *Gene* **190**, 55-62.
- Srinivasan, N., Rufino, S. D., Pepys, M. B., Wood, S. P. & Blundell, T. L. (1996) A superfamily of proteins with a lectin fold. *Chemtracts: Biochem. Mol. Biol.* **6**.
- Stevens, T. H., Rothman, J. H., Payne, G. S. & Schekman, R. (1986) Gene dosage-dependent secretion of yeast vacuolar carboxypeptidase Y. *J. Cell Biol.* **102**, 1551-1557.
- Studier, F. W. & Moffatt, B. A. (1986) Use of bacteriophage T7 RNA polymerase to direct selective high level expression of cloned genes. *J. Mol. Biol.* **189**, 113-130.
- Sturm, A. & Chrispeels, M. J. (1986) The high mannose oligosaccharide of phytohemagglutinin is attached to asparagine 12 and the modified oligosaccharide to asparagine 60. *Plant Physiol.* **81**, 320-322.
- Surolia, A., Sharon, N. & Schwarz, F. P. (1996) Thermodynamics of monosaccharide and disaccharide binding to *Erythrina corallodendron* lectin. *J. Biol. Chem.* **271**, 17697-17703.

-
- Swamy, M. J., Gupta, D., Mahanta, S. K. & Surolia, A. (1991) Further characterization of the saccharide specificity of peanut (*Arachis hypogaea*) agglutinin. *Carbohydr. Res.* **213**, 59-67.
- Tague, B. & Chrispeels, M. J. (1987) The plant vacuolar protein, phytohemagglutinin, is transported to the vacuole of transgenic yeast. *J. Cell Biol.* **105**, 1971-1979.
- Tahirov, T. H., Lu, T. H., Liaw, Y. C., Chen, Y. L. & Lin, J. Y. (1995) Crystal structure of abrin A at 2.14 angstrom. *J. Mol. Biol.* **250**, 354-367.
- Toone, E. J. (1994) Structure and energetics of protein carbohydrate complexes. *Curr. Opin. Struct. Biol.* **4**, 719-728.
- Tschopp, J. F., Sverlow, G., Kosson, R., Craig, W. & Grinna, L. (1987) High level secretion of glycosylated invertase in the methylotrophic yeast *Pichia pastoris*. *Bio/Technol.* **5**, 1305-1308.
- Tuite, M. (1991) Expression of heterologous genes. *Saccharomyces*. M. Tuite & S. Oliver. New York, Plenum Press: 169-212.
- Valls, L. A., Hunter, C. P., Rothman, J. H. & Stevens, T. H. (1987) Protein sorting in yeast: the localization of determinant of yeast vacuolar carboxypeptidase Y resides in the propeptide. *Cell* **48**, 887-897.
- Van Damme, E. J. M., Allen, A. K. & Peumans, W. J. (1987) Isolation and characterization of a lectin with exclusive specificity towards mannose from snowdrop (*Galanthus nivalis*) bulbs. *FEBS Lett.* **215**, 140-144.

-
- Van Damme, E. J. M. & Peumans, W. J. (1989) Developmental changes and tissue distribution of lectin in Tulipa. *Planta* **178**, 10-18.
- Van Damme, E. J. M. & Peumans, W. J. (1990) Developmental changes and tissue distribution of lectin in *Galanthus nivalis* and *Narcissus cv Carlton*. *Planta* **182**, 605-609.
- Van Damme, E. J. M., Declercq, N., Claessens, F., Hemschoote, K., Peeters, B. & Peumans, W. J. (1991a) Molecular cloning and characterization of multiple isoforms of the snowdrop (*Galanthus nivalis*) lectin. *Planta* **186**, 35-43.
- Van Damme, E.J.M., Kaku, H., Perini, F., Goldstein, I.J., Peeters, B., Yagi, F., Decock, B. & Peumans, W.J. (1991b) Biosynthesis, primary structure and molecular cloning of snowdrop (*Galanthus nivalis*) lectin. *Eur. J. Biochem.* **202**, 23-30
- Van Damme, E. J. M., Smeets, K., Torrekens, S., Van Leuven, F., Goldstein, I. J. & Peumans, W. J. (1992) The closely related homomeric and heterodimeric mannose-binding lectins from garlic are encoded by one-domain and 2-domain lectin genes, respectively. *Eur. J. Biochem.* **206**, 413-420.
- Van Damme, E. J. M., Smeets, K., Engelborghs, I., Aelbers, H., Balzarini, J., Pusztai, A., Van Leuven, F., Goldstein, I. J. & Peumans, W. J. (1993) Cloning and characterization of the lectin cDNA clones from onion, shallot and leek. *Plant Mol. Biol.* **23**, 365-376.
- Van Damme, E. J. M., Balzarini, J., Smeets, K., Van Leuven, F. & Peumans, W. J. (1994) The monomeric and dimeric mannose-binding proteins from the

-
- orchidaceae species *Listera ovata* and *Epipactis helleborine* - sequence homologies and differences in biological activities. *Glycoconj. J.* **11**, 321-332.
- Van Damme, E. J. M., Barre, A., Rouge, P., Van Leuven, F. & Peumans, W. J. (1995a) The seed lectins of black locust (*Robinia pseudoacacia*) are encoded by two genes which differ from the bark lectin genes. *Plant Mol. Biol.* **29**, 1197-1210.
- Van Damme, E. J. M., Goossens, K., Smeets, K., Van Leuven, F., Verhaert, P. & Peumans, W. J. (1995b) The major tuber storage protein of *Araceae* species is a lectin - characterization and molecular cloning of the lectin from *Arum maculatum*. *Plant Physiol.* **107**, 1147-1158.
- Van Damme, E. J. M., Brike, F., Winter, H. C., Van Leuven, F., Goldstein, I. J. & Peumans, W. J. (1996a) Molecular cloning of two different mannose-binding lectins from tulip bulbs. *Eur. J. Biochem.* **236**, 419-427.
- Van Damme, E. J. M., Barre, A., Rouge, P., Van Leuven, F. & Peumans, W. J. (1996b) Characterization and molecular cloning of *Sambucus nigra* agglutinin V (nigrin b), a GalNAc-specific type-2 ribosome-inactivating protein from the bark of elderberry (*Sambucus nigra*). *Eur. J. Biochem.* **237**, 505-513.
- Van Damme, E. J. M., Barre, A., Rouge, P., Van Leuven, F. & Peumans, W. J. (1996c) The NeuAc(α -2,6)Gal/GalNAc-binding lectin from elderberry (*Sambucus nigra*) bark, a type-2 ribosome-inactivating protein with an unusual specificity and structure. *Eur. J. Biochem.* **235**, 128-137.

-
- Van Damme, E. J. M., Peumans, W. J., Pusztai, A. & Bardocz, S. (1998a) Handbook of Plant Lectins: Properties and Biomedical Applications, John Wiley & Sons, Chichester, U.K.
- Van Damme, E. J. M., Peumans, W. J., Barre, A. & Rouge, P. (1998b) Plant lectins: A composite of several distinct families of structurally and evolutionary related proteins with diverse biological roles. *Crit. Rev. Plant Sci.* **17**, 575-692.
- Van Eijsden, R. R., Hoedemaeker, F. J., Diaz, C. L., Lugtenberg, B. J. J., Depater, B. S. & Kijne, J. W. (1992) Mutational analysis of pea lectin - substitution of Asn125 for Asp in the monosaccharide-binding site eliminates mannose glucose binding activity. *Plant Mol. Biol.* **20**, 1049-1058.
- Van Eijsden, R. R., Depater, B. S. & Kijne, J. W. (1994) Mutational analysis of the sugar-binding site of pea lectin. *Glycoconj. J.* **11**, 375-380.
- Van Parijs, J., Broekaert, W. F., Goldstein, I. J. & Peumans, W. J. (1991) Hevein - an antifungal protein from rubber tree (*Hevea brasiliensis*) latex. *Planta* **183**, 258-264.
- Van Parijs, J., Joosen, H. M., Peumans, W. J., Geuns, J. M. & Van Laere, A. J. (1992) Effect of the lectin UDA (*Urtica dioica* agglutinin) on germination and cell wall formation of *Phycomyces blakesleeanus*. *Burgeff Arch Microbiol* **158**.
- Vasil, I. K. & Hubbell, D. H. (1977) The effect of lectins on cell division in tissue cultures of soybean and tobacco. *Cell Wall Biochemistry Related to*

-
- Specificity in Host-Plant Interactions. B. Solheim. Tromsø/Oslo/Bergen, Universitetsforlaget: 361-67.
- Vijayan, M. & Chandra, N. (1999) Lectins. *Curr. Opin. Struct. Biol.* **9**, 707-714.
- Vitale, A. & Raikhel, N. V. (1999) What do proteins need to reach different vacuoles? *Trends Plant Sci.* **4**, 149-155.
- von Heijne, G. (1994) Membrane proteins - from sequence to structure. *Annu. Rev. Biophys. Biomol. Struct.* **23**, 167-192.
- Von Schauwen, A. & Chrispeels, M. J. (1993) Identification of vacuolar sorting information in phytohemagglutinin, an unprocessed vacuolar protein. *J. Exp. Bot.* **44**, 339-342.
- Vuorela, A., Myllyharju, J., Nissi, R., Pihlaniemi, T. & Kivirikko, K. (1997) Assembly of human prolyl 4-hydroxylase and type III collagen in the yeast *Pichia pastoris*: formation of a stable enzyme tetramer requires coexpression with collagen and assembly of a stable collagen requires coexpression with prolyl 4-hydroxylase. *EMBO J.* **16**, 6702-6712.
- Wang, M. B., Boulter, D. & Gatehouse, J. A. (1994) Characterization and sequencing of cDNA clone encoding the phloem protein pp2 of *Cucurbita pepo*. *Plant Mol. Biol.* **24**, 159-170.
- Weiss, H., Haase, W., Michel, H. & Reilander, H. (1995) Expression of functional mouse 5-HT_{5A} serotonin receptor in the methylotrophic yeast *Pichia pastoris*: pharmacological characterization and localization. *FEBS Lett.* **377**, 21-25.

-
- Weis, W. I. & Drickamer, K. (1996) Structural basis of lectin-carbohydrate recognition. *Annu. Rev. Biochem.* **65**, 441-473.
- Weis, W. I. (1997) Cell-surface carbohydrate recognition by animal and viral lectins. *Curr. Opin. Struct. Biol.* **7**, 624-630.
- Wood, S. D., Wright, L. M., Reynolds, C. D., Rizkallah, P. J., Allen, A. K., Peumans, W. J. & Van Damme, E. J. M. (1999) Structure of the native (unligated) mannose-specific bulb lectin from *Scilla campanulata* (bluebell) at 1.7 angstrom resolution. *Acta Crystallogr. D.* **55**, 1264-1272.
- Wright, C. S. (1989) Comparison of the refined crystal structures of 2 wheat germ isolectins. *J. Mol. Biol.* **209**, 475-487.
- Wright, C. S. (1990) 2.2 angstrom resolution structure analysis of 2 refined N-acetylneuraminyllactose-wheat-germ-agglutinin-isolectin complexes. *J. Mol. Biol.* **215**, 635-651.
- Wright, C. S. (1992) Crystal structure of a wheat germ agglutinin glycoporphin-sialoglycopeptide receptor complex - structural basis for cooperative lectin-cell binding. *J. Biol. Chem.* **267**, 14345-14352.
- Wright, C. S. & Jaeger, J. (1993) Crystallographic refinement and structure-analysis of the complex of wheat germ agglutinin with a bivalent sialoglycopeptide from glycoporphin-A. *J. Mol. Biol.* **232**, 620-638.
- Wright, C. S. & Hester, G. (1996) The 2.0 angstrom structure of a cross-linked complex between snowdrop lectin and a branched mannopentaose: Evidence for two unique binding modes. *Structure* **4**, 1339-1352.

-
- Wright, H. T., Sandrasegaram, G. & Wright, C. S. (1991) Evolution of a family of N-acetylglucosamine binding proteins containing the disulfide-rich domain of wheat germ agglutinin. *J. Mol. Evol.* **33**, 283-294.
- Young, N. M., Watson, D. C. & Thibault, P. (1995) Mass spectrometric analysis of genetic and posttranslational heterogeneity in the lectins jacalin and *Maclura pomifera* agglutinin. *Glycoconj. J.* **12**, 135-141.
- Zhu, K. Y., Bressan, R. A., Hasegawa, P. M. & Murdock, L. L. (1996) Identification of N-acetylglucosamine binding residues in *Griffonia simplicifolia* lectin II. *FEBS Lett.* **390**, 271-274.
- Zhu-Salzman, K., Shade, R. E., Koiwa, H., Salzman, R. A., Narasimhan, M., Bressan, R. A., Hasegawa, P. M. & Murdock, L. L. (1998) Carbohydrate binding and resistance to proteolysis control insecticidal activity of *Griffonia simplicifolia* lectin II. *Proc. Natl. Acad. Sci. U. S. A.* **95**, 15123-15128.
- Zsebo, K., Lu, H. S., Fieschko, J., Goldstein, L., Davis, J., Duker, K., Suggs, S., Lai, P. H. & Bitter, G. (1986) Protein secretion from *Saccharomyces cerevisiae* directed by the prepro-alpha-factor leader region. *J. Biol. Chem.* **261**, 5858-5865.

Functional phytohemagglutinin (PHA) and *Galanthus nivalis* agglutinin (GNA) expressed in *Pichia pastoris*

Correct N-terminal processing and secretion of heterologous proteins expressed using the PHA-E signal peptide

Maarten J. M. Raemaekers, Laura de Muro*, John A. Gatehouse and Anthony P. Fordham-Skelton

Department of Biological Sciences, University of Durham, South Road, Durham, DH1 3LE, UK

Phytohemagglutinin (*Phaseolus vulgaris* agglutinin; PHA; E- and L-forms) and snowdrop lectin (*Galanthus nivalis* agglutinin; GNA) were expressed in *Pichia pastoris* using native signal peptides, or the *Saccharomyces* α -factor preprosequence, to direct proteins into the secretory pathway. PHA and GNA were present as soluble, functional proteins in culture supernatants when expressed from constructs containing the α -factor preprosequence. The recombinant lectins, purified by affinity chromatography, agglutinated rabbit erythrocytes at concentrations similar to the respective native lectins. However, incomplete processing of the signal sequence resulted in PHA-E, PHA-L and GNA with heterogeneous N-termini, with the majority of the protein containing N-terminal extensions derived from the α -factor preprosequence. Polypeptides in which most of the α -factor preprosequence was present were also glycosylated. Inclusion of Glu-Ala repeats at the C-terminal end of the α -factor preprosequence led to efficient processing N-terminal to the Glu-Ala sequence, but inefficient removal of the repeats themselves, resulting in polypeptides with heterogeneous N-termini still containing N-terminal extensions. In contrast, PHA expressed with the native signal peptide was secreted, correctly processed, and also fully functional. No expression of GNA from a construct containing the native GNA signal peptide was observed. The PHA-E signal peptide directed correct processing and secretion of both GNA and green fluorescent protein (GFP) when used in expression constructs, and is suggested to have general utility for synthesis of correctly processed proteins in *Pichia*.

Keywords: *Pichia*; lectin; signal peptide; processing; glycosylation; GFP

Extensive study of the properties of many plant lectins has been hampered by their occurrence as heterogeneous mixtures of isoforms *in planta* (reviewed in [1]) as, for example, the lectins from *Robinia pseudoacacia* [2]. These isoforms occur as a result of a number of factors. Firstly, in some cases imprecise processing of the C-termini of mature polypeptides can give rise to heterogeneity in lectin sequences ([3], and references therein). Secondly, most lectins are encoded by multigene families, members of which can be expressed in the same tissue, or differentially expressed in different tissues and/or developmental stages. Thirdly, allelic variation in lectin sequences between plant lines is well-established, and can contribute further to the heterogeneity of lectin preparations where the source material is not genetically defined or uniform. The difficulty of separating multiple lectin isoforms by conventional techniques has led to functional properties and biological activities being defined for protein mixtures, and many of the naturally occurring lectin variants probably

differ little in functional properties, in other cases the differences in sequence between isoforms causes significant effects on the biological activity of the molecule. For example, two of the isoforms of the *Phaseolus vulgaris* lectin family, PHA-E and PHA-L, differ in their agglutination activity on human blood cells; at low concentrations the former agglutinates erythrocytes whereas the latter agglutinates, and is mitogenic towards, leucocytes [4–6].

Expression of plant lectins in heterologous systems offers a means of producing proteins of defined amino acid sequence, and allows sequence–function relationships in lectins to be explored through site-directed mutagenesis. However, attempts to express plant lectins in bacterial and yeast hosts have met with mixed success and is very much case-dependent. Whilst some plant lectins can be produced as functional proteins in *Escherichia coli*, albeit nonglycosylated [7–9], many form insoluble inclusion bodies, for example; snowdrop lectin (GNA) [10], *Erythrina* lectin (ECoRL) [11] and soybean agglutinin [12]. In these cases, lectins must be solubilized by denaturation-renaturation steps, and recovery of active lectin is often poor, and the possibility that the lectin does not fully recover its proper native conformation cannot be ruled out. *E. coli* is also incapable of glycosylating eukaryotic proteins expressed in this host. Many lectins are glycosylated, and the presence or absence of the carbohydrate side chains can have significant effects on the functional and/or physical properties of the recombinant proteins. For example, concanavalin A is activated by deglycosylation *in planta* [13,14] and

*Correspondence to: J. A. Gatehouse, Department of Biological Sciences, University of Durham, South Road, Durham, DH1 3LE, UK.

Phone: +44 191 3742417, E-mail: J.A.Gatehouse@durham.ac.uk

Abbreviations: PHA, *Phaseolus vulgaris* phytohemagglutinin (E- and L-forms); GNA, *Galanthus nivalis* agglutinin; ECL, enhanced cyanine fluorescence; PVDF, poly(vinylidene difluoride).

Contract grant sponsor: DISMA, Università degli Studi di Milano, via Celoria 2, 20133 Milan, Italy.

Received 25 May 1999, accepted 27 July 1999

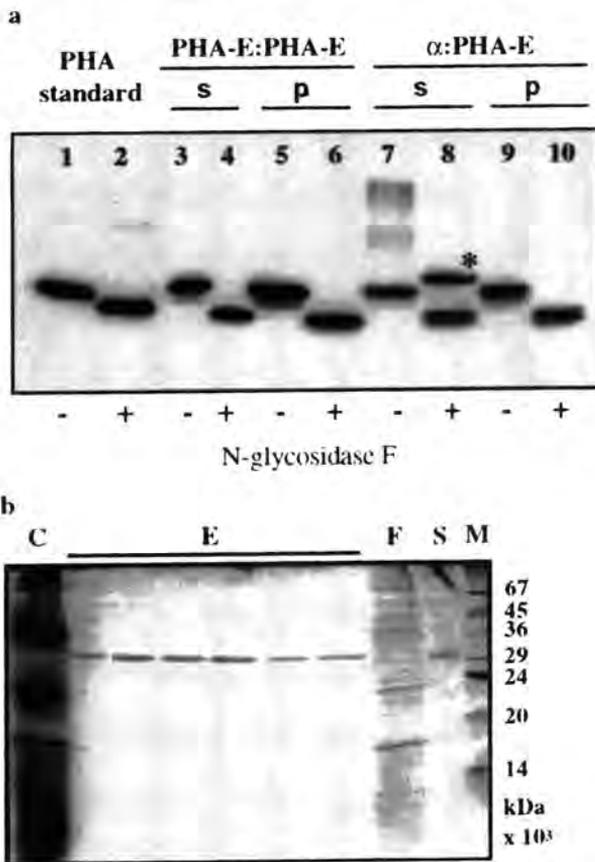


Fig. 1. (a) Summary of expression and purification of PHA in *Pichia*, and (b) Silver stained gel showing the affinity purification of recombinant PHA-L on a thyroglobulin-agarose column, expressed with the native signal peptide. (a) Data for PHA-E are shown but PHA-L gave essentially identical results. Western analysis of PHA-E in culture supernatants (s) or affinity purified (p) samples expressed using either the native signal peptide or the prepro α -factor sequence. Samples were treated (\pm) with N-glycosidase F to resolve differences in glycosylated forms. The conspicuous 'smear' present in the α -factor supernatant is resolved to a single deglycosylated higher molecular mass form of PHA-E (asterisk) not present in the affinity purified sample. (b) Silver stained gel, with: (C) ammonium sulphate precipitate of induced culture supernatant; (E) eluted fractions; (F) column flowthrough; (S) PHA standard. Molecular mass markers were SDS-7 (BioRad).

PHA, estimates of the relative amounts of protein in the cells and supernatant suggested that at least 50% of the protein was secreted (not shown). The proportion of GNA inside cells could not be estimated as GNA also binds to cell wall mannose residues.

Both PHA and GNA were purified from culture supernatants by affinity chromatography, using the functional activity of the lectins, on thyroglobulin- and mannan-agarose columns, respectively. All four forms of PHA (from the PHA-E:PHA-E, α :PHA-E, PHA-L:PHA-L and α :PHA-L constructs) and GNA (from the α :GNA construct) bound to, and could be eluted from, the corresponding affinity column to yield proteins free of contaminants. A representative purification of PHA-L, expressed using the native signal peptide sequence, is shown in Fig. 1b, indicating the protein is essentially pure as judged by silver staining. Similar results were obtained for the other PHA constructs. Recombinant PHA-E (from both the PHA-E:PHA-E and α :PHA-E constructs) and GNA agglutinated rabbit erythrocytes at concentrations similar to those of native lectins

assayed under the same conditions (Table 2), and which consistent with concentrations reported for the corresponding commercially available plant-derived lectins. Sugar inhibition of haemagglutination was also consistent with the results of native lectins (data not presented). PHA-L does not agglutinate erythrocytes at these concentrations, and thus was not assayed. The data from haemagglutination assays and affinity purification shows that the lectins produced in *Pichia* are functional proteins.

Processing of proteins expressed using the α -factor preprosequence

Although the constructs in which the α -factor preprosequence was used as a signal peptide led to the secretion of functional lectins from *Pichia*, analysis of polypeptides present in culture supernatants showed that each construct gave rise to multiple bands detected by antilectin antibodies, suggesting that processing was not occurring correctly. For both PHA and GNA, although a band of similar mobility to the 'native' protein was present in the culture supernatant, diffuse bands and smears at higher molecular masses were also present (Fig. 1a; lane 7; Fig. 2, lanes 1 and 3). The diffuse bands were absent from α :PHA constructs after purification by affinity chromatography, showing that these polypeptides does not bind to the affinity column (compare lanes 7 and 9 of Fig. 1a). For GNA, diffuse higher molecular mass bands were present both before and after purification, although a smear of immunoreactive material of very low mobility on SDS/PAGE was not present in the eluted fraction after affinity chromatography, but was present in the flow-through fraction. The diffuse band

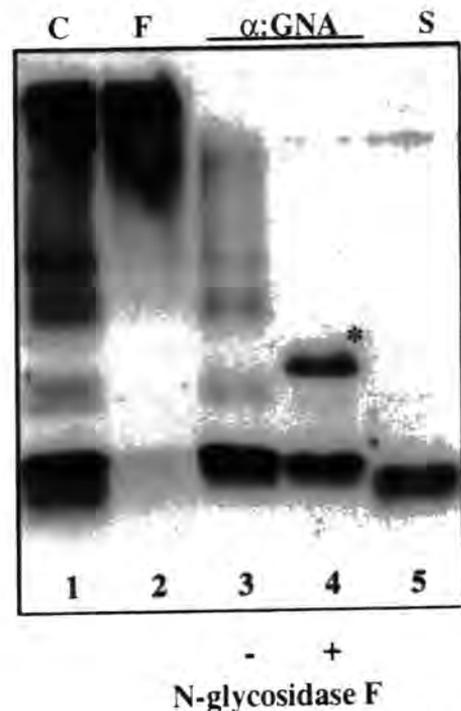


Fig. 2. Summary of expression and purification of GNA in *Pichia* expressed using the α -factor preprosequence. Western analysis of GNA in culture supernatants (C), column flowthrough (F), affinity purified (α :GNA) and standard native GNA (S). Affinity purified GNA was treated (\pm) with N-glycosidase F to resolve differences in glycosylated forms. Asterisk indicates the higher molecular mass form of GNA containing residues of α -factor pro sequence.

Lectins, standard proteins and haemagglutination assays

Vegetable lectins purified from plant sources were obtained from Pharmacia (PHA E/L) and W. Peumans (GNA; University of Leuven, Belgium). Standard recombinant GFP expressed in *E. coli* was obtained from Clontech. Native and recombinant protein concentrations were estimated by measuring absorbance at 280 nm.

Haemagglutination assays were carried out in round-bottomed microtitre plates. A total volume of 100 μ L was used in each assay: 50 μ L aliquots of serial twofold dilutions of the lectin in NaCl/P_i and 50 μ L of 2% erythrocyte suspension in NaCl/P_i. In a microtitre plate with the 100 μ L erythrocyte suspension containing the serial double dilutions of the sample lectin was incubated for 1 h at room temperature. The lowest concentration required to completely agglutinate the red blood cells was determined visually.

Affinity-purification of PHA and GNA

Recombinant PHA isoforms expressed in *Pichia* were affinity-purified on thyroglobulin-agarose essentially as described [28]. Briefly, proteins from *Pichia* culture supernatants (50 mL) were precipitated by the addition of solid (NH₄)₂SO₄ to 80% saturation and incubated overnight at 4 °C. Precipitates were collected by centrifugation at 3000 g for 40 min, resuspended in 10 mL NaCl/P_i and dialysed against 4 L NaCl/P_i overnight at 4 °C. Dialysed samples were loaded onto a thyroglobulin-agarose column (Sigma; 5-mL volume), followed by washing in NaCl/P_i until all nonbound material had eluted. The column was then washed with nonbuffered saline (1 column volume), and eluted in 0.5 M NaCl, 50 mM glycine (pH 3.0). The pH of eluted samples was brought to 7.4 by the addition of Tris/HCl to 0.1 M and samples were dialysed against NaCl/P_i overnight at 4 °C.

Recombinant GNA was purified from *Pichia* cultures (50 mL) by binding to a mannan-agarose (Sigma) affinity column. Briefly, proteins from culture supernatants were precipitated with (NH₄)₂SO₄ as described for the PHA purification. After resuspending precipitates in NaCl/P_i, the whole fraction was loaded directly onto a column (10 mL) containing mannan-agarose (Sigma). The column was washed with NaCl/P_i to remove any unbound proteins until the absorbance reached background level. The bound fraction was subsequently eluted isocratically with 20 mM diammonium citrate and eluted proteins were monitored by online measurement of absorbance at 280 nm. Fractions containing recombinant GNA were pooled, dialysed against NH₄HCO₃, lyophilized and dissolved in NaCl/P_i.

Glycosylation of protein with N-glycosidase F

Reaction samples were denatured by boiling for 10 min in the presence of 1.0% (w/v) SDS and 8.0% 2-mercaptoethanol. Samples were diluted to 0.1% SDS, 0.8% 2-mercaptoethanol, 50 mM sodium phosphate (pH 7.4), 25 mM EDTA and 2% Triton X-100. Recombinant N-glycosidase F (Boehringer) (0.6 U) was added and samples were incubated overnight at 37 °C. Control reactions omitted the enzyme. Following SDS-PAGE, samples were analysed by Western blotting as described above.

N-terminal sequencing

N-terminal sequencing was carried out on affinity purified proteins or protein bands blotted onto Problot poly(vinylidene difluoride) (PVDF) membrane after separation by SDS/PAGE, as described (<http://www.bio.cam.ac.uk/proj/adr/PNAC/blotguide.html>), using an Applied Biosystems model 477 amino acid sequencer. A standard procedure for identifying N-terminal sequences was followed, where the machine was run for six cycles and the resulting sequence data was analysed for agreement with the amino acid sequences predicted by the nucleotide sequences of the expression constructs. Any ambiguities were resolved by carrying out further sequencing.

RESULTS

PHA and GNA expressed in *Pichia* are functional secreted proteins

Initially, two sets of constructs for the expression of PHA (E- and L-forms) and GNA were prepared; in the first, the lectin coding sequence contained the 'native' signal peptide sequence and in the second, the signal peptide sequence was replaced by the *Saccharomyces* α -factor preprosequence (extending as far as the processing site of the Kex2p protease, i.e. omitting the 'Glu-Ala' repeats). All construct designations and corresponding sequences are shown in Table 1 and Fig. 4. Coding sequences were placed under the control of the methanol-inducible *AOX1* promoter present in the pPICZ series of *Pichia* expression vectors (Invitrogen). Constructs were transformed into *Pichia* strains GS115 or KM71 and the supernatants from methanol-induced cultures were analysed for the presence of the corresponding lectin.

Figures 1 (PHA) and 2 (GNA) summarize results for the expression and purification studies of lectins in *Pichia*. Proteins which reacted with the appropriate antilectin antibody were present in supernatants of induced cultures for all the PHA expression constructs, both with the native signal peptide (PHA-E:PHA-E and PHA-L:PHA-L;) and the α -factor preprosequence (α :PHA-E and α :PHA-L) (Fig. 1a, lanes 3 and 7). Results for E- and L-forms of PHA were essentially the same and a representative analysis of PHA-E is shown. In contrast to PHA, 8 independent transformants containing the complete GNA coding sequence (GNA:GNA) failed to produce GNA, although GNA expression could be detected readily in *Pichia* clones containing the construct where the native GNA signal peptide was replaced by the α -factor preprosequence (α :GNA) (Fig. 2, lane 1). Transformation into either strain GS115 or KM71 did not produce any significant differences in levels of lectin secretion (not shown).

Analysis by SDS/PAGE and Western blotting showed that polypeptides of similar molecular masses to appropriate lectin standards, recognized by anti-PHA or anti-GNA antibodies, were present. However, the polypeptide profiles of the bands recognized by antilectin antibodies after Western blotting differed from the 'native' lectins. In the case of PHA, the bands recognized by anti-PHA antibodies also differed between constructs, depending on whether the α -factor preprosequence or native signal peptide sequence was used. The differences in the polypeptide profiles derived from different constructs are due to differences in processing between native signal peptides and the α -factor preprosequence, and are described in more detail in the following section. Levels of PHA and GNA in the culture supernatants were estimated from Western blots as being in the range 0.4–1.0 and 1–2 mg·L⁻¹, respectively. For

Table 1. Oligonucleotide sequences of primers used to prepare *Pichia* expression constructs. The PHA-E:GNA and PHA-E:GFP constructs produced by a two-step process; the product of amplification of the first pair of primers was used as a 'Megaprimer' in the second amplification step. M13RP1 = M13 reverse sequencing primer 1 (5'-CACACAGGAAACAGCTATGAC-3'); 5' AOX1 = *Pichia* alcohol oxidase promoter primer 1 (5'-GACTGGTTCCAATTGACAAGC-3')

Construct	PCR Primers: N-terminal (5'-3') C-terminal (5'-3')	Vector	Cloning sites	Coding sequence(s)
PHA-E:PHA-E	GCGAATTCACCATGGCTTCCTCCAACCTACTC M13RP1	pPICZ B	<i>EcoRI</i> / <i>XbaI</i>	PHA-E signal peptide Mature PHA-E
PHA-L:PHA-L	GCGAATTCACCATGGCTTCCTCCAAGTTCTTC M13RP1	pPICZ B	<i>EcoRI</i> / <i>XbaI</i>	PHA-L signal peptide Mature PHA-L
α :PHA-E	GCCTCGAGAAAAGAGCCAGCCAAACCTCCTTCAGC M13RP1	pPICZ α A	<i>XhoI</i> / <i>XbaI</i>	α -factor prepro- : Mature PHA-E
α :PHA-L	GCCTCGAGAAAAGAAGCAACGATATCTACTTCAAC M13RP1	pPICZ α A	<i>XhoI</i> / <i>XbaI</i>	α -factor prepro- : Mature PHA-L
GNA:GNA	ATTACTCGAGAAAATGGCTAAGGCAAGTC TAATTCTAGATTACTTTGCCGTCACAAGC	pPICZ B	<i>XhoI</i> / <i>XbaI</i>	GNA signal peptide Mature GNA
α :GNA	ATTACTCGAGAAAAGAGACAATATTTGTAC TAATTCTAGATCATCCGGTGTGAGTTCC	pPICZ α A	<i>XhoI</i> / <i>XbaI</i>	α -factor prepro- : Mature GNA
α EA:GNA	ATTACTCGAGAAAAGAGAGCTGAAGCTGACAATATTTT GTACTCCTAATTCTAGATCATCCGGTGTGAGTTCC	pPICZ α A	<i>XhoI</i> / <i>XbaI</i>	α -factor prepro- / E repeats : Mature GNA
PHA-E:GNA	1. 5' AOX1 GGAGTACAAAATATTGTCTGAGTTGCGTGGGTGAG 2. Megaprimer from above TAATTCTAGATCATCCGGTGTGAGTTCC	pPICZ B	<i>EcoRI</i> / <i>XbaI</i>	PHA-E signal peptide Mature GNA
α EA:GFP	ATTACTGCAGCAAGTAAAGGAGAAGAAGCTTTTC TAATTCTAGAATTCATTATTTGTAGAGTCCATC	pPICZ α B	<i>PstI</i> / <i>XbaI</i>	α -factor prepro- / EAEA repeats : GFP
PHA-E:GFP	1. 5' AOX1 CCAGTGAAAAGTTCTTCTCCTTTACTTGAGTTGCGTGGGTGAG 2. Megaprimer from above TAATTCTAGAATTCATTATTTGTAGAGTCCATC	pPICZ B	<i>EcoRI</i> / <i>XbaI</i>	PHA-E signal peptide GFP

Glu-Ala repeats of pPICZ α B, resulting in an extra two alanine residues between the Glu-Ala repeats and the N-terminus of GFP (methionine initiation codon omitted). Construct PHA-E:GFP (fusion of PHA-E signal peptide to GFP) was constructed by the 'Megaprimer' method. The first round of amplification used PHA-E:PHA-E in pPICZ B as template with the 5' AOX1 vector primer and a primer corresponding to the last six residues of the PHA-E signal peptide and the first six residues of GFP (omitting the methionine initiation codon). The second amplification used pGFPuv as template, the purified megaprimer obtained above and the GFP C-terminal primer.

Pichia transformation and induction of expression of recombinant lectins

Plasmids, linearized by restriction with *Bst*XI, were transformed into *Pichia* strains GS115 or KM71 by electroporation or using the 'EasyComp' chemical transformation method as described in the Invitrogen users manual. Transformants were selected on YPD-sorbitol plates containing 100 μ g·mL⁻¹ Zeocin (Invitrogen) incubated at 30 °C. Single colonies were grown overnight in 10 mL BMGY medium, spun down, resuspended in 1.0 mL distilled water and inoculated into fresh BMMY medium (50 mL). Cultures were shaken at 300 r.p.m. and fresh methanol was added daily to 0.5% (v/v) at 9 am and 6 pm during the course of induction (7 days).

SDS/PAGE and Western analysis

Samples were analysed using denaturing SDS/PAGE with 15% gels run under reducing conditions. The bands were

visualized by either Coomassie Brilliant Blue R250 or silver staining. Samples were treated with 10% 2-mercaptoethanol and heated in boiling water for 10 min before loading on a gel. Molecular mass markers (SDS7; Sigma) were used to calibrate the gels.

Electrophoretic transfer of proteins from gels to nitrocellulose membranes (Schleicher and Schuell, BA85) was carried out by the semidry blotting technique [27], and was followed by immunodetection. The membrane was incubated in phosphate buffered saline (NaCl/P_i; 8 g NaCl, 0.2 g KCl, 1 g Na₂HPO₄, 0.2 g KH₂PO₄ per litre, pH 7.4) containing 5% (w/v) nonfat milk powder (Marvel) and 1% Tween 20 for 1 h at room temperature with gentle shaking. Polyclonal rabbit anti-GNA serum (produced by R. R. D. Croy and L. N. Gatehouse, University of Durham), rabbit anti-PHA serum (Vector Laboratories), mouse anti-GFP serum (Clontech) were diluted (1 : 10 000, 1 : 20 000 or 1 : 1000, respectively) with NaCl/P_i containing 5% nonfat milk powder and 0.1% Tween 20. Membranes were incubated in primary antibody solution at room temperature for 1.5 h with gentle shaking. These were then washed with antibody dilution buffer for 3 \times 5 min at room temperature. Secondary antibodies were IgG horseradish peroxidase conjugates (BioRad) diluted 1 : 10 000, as above, and incubation was for 1.5 h at room temperature followed by 3 \times 5 minute washes in antibody dilution buffer and a final wash in distilled water. Enhanced chemiluminescence reagents (Amersham) were used to detect specifically bound secondary antibodies as instructed by the manufacturer, and bands were visualized by exposure to Fuji film.

glycosylated PHA-L expressed in transgenic plants gave high quality crystals under conditions where glycosylated PHA-L was crystallized successfully [15].

In a similar manner, the expression of lectins in *Saccharomyces cerevisiae* also has associated problems. PHA-L expressed in *Saccharomyces* accumulates mainly in the vacuole [16], due to the presence of 'cryptic' vacuolar targeting signals [17], and only about 1% is secreted. In addition, a significant proportion of PHA-L expressed in this host was not correctly processed, approximately half of the PHA-L accumulating in the vacuole appeared to contain the uncleaved signal peptide, and all of the secreted PHA-L was in this unprocessed form [16]. A second related lectin, from *Dolichos biflorus*, could be directed into a secretory pathway when expressed at low levels, but accumulated in the cells (not in the vacuole) when expressed at high levels [18]. In both cases, the functional properties of the recombinant lectin were not reported. The only report of the correct processing and secretion of a plant lectin in *Saccharomyces* is that of wheat germ agglutinin which was secreted and exhibited sugar binding activity, however, yields of protein were relatively low, of the order of $1 \mu\text{g}\cdot\text{L}^{-1}$ [19]. Thus it would appear that *Saccharomyces* is not a satisfactory host for lectin expression.

The development of relatively facile expression methodologies for the methylotrophic yeast *Pichia pastoris*, especially the secretion of recombinant proteins, offers an alternative to expressing plant lectins [20,21]. In this study, two plant lectins were selected for expression in *Pichia*, *Phaseolus vulgaris* agglutinin (PHA; both the E- and L-forms) and soybean lectin (GNA). The rationale for this selection was (a) PHA-L is incorrectly processed and accumulates in the vacuole of *Saccharomyces* and (b) GNA accumulates in inclusion bodies when expressed in *E. coli*. In addition, PHA and GNA are members of two very different lectin families, the concanavalin and the monocot. mannose-specific lectins, respectively. These families are unrelated in sequence and we would therefore predict that, if these lectins could be expressed, *Pichia* may become a useful system for the expression of a wide range of plant lectins. In this paper we demonstrate that *Pichia* expresses functional PHA and GNA, which are secreted into the culture medium. We also show that the PHA-E signal peptide directed the secretion of proteins which are correctly processed at the N-termini, whereas proteins secreted under the control of the *Saccharomyces* prepro α -factor sequence have unprocessed N-terminal extensions, and suggest that the PHA-E signal peptide may have a wider utility in the production of recombinant proteins in *Pichia*.

EXPERIMENTAL PROCEDURES

Sequence analysis

Genomic DNA templates were sequenced using fluorescently labeled dye terminators and an AmpliTaq cycle sequencing kit (Applied Biosystems, Warrington, Cheshire, UK). Reaction products were analysed on a PE ABI 373 stretch or 377 DNA sequencer. All constructs described were sequenced to verify that no mutations had occurred during PCR amplification.

Expression constructs for PHA-E and PHA-L

Genomic DNA isolated from *Phaseolus vulgaris* cv. Tendergreen was used as a template for PCR amplification of PHA-E and PHA-L sequences. Primer sequences were designed using the published sequences of PHA-E (EMBL X02408) and PHA-L

(EMBL X02409) [22]. PCR primers corresponded to the first six N-terminal codons of the PHA signal peptide sequence and the last six codons (including the stop codon) of the mature coding sequence and a further 5 bp of the 3' UTR. A *Bam*HI site was included at the 5' end of each primer. Primer sequences were:

PHA-E N-terminal primer: 5'-CGGATCCATGGCTTCCTC-CAACTTAC-3';

PHA-E C-terminal primer: 5'-CGGATCCTGGAGTCTAGAGGATTTGGTTGAG-3';

PHA-L N-terminal primer: 5'-CGGATCCCATGGCTTCCTC-CCAAGTTC-3';

PHA-E C-terminal primer: 5'-CGGATCCTGGAGTCTAGAGGATTTGGTTGAG-3'.

After amplification, PCR products were blunt-end cloned into pUC18 and sequenced to check for absence of PCR errors. To produce *Pichia* expression constructs (see Table 1), modified N-terminal oligonucleotides were used to amplify PHA template DNA, in combination with a generic M13 sequencing primer (which amplifies across the *Xba*I site containing a stop codon in the PHA coding sequence), using *Pfu* polymerase (Stratagene). Primers for expression using the native signal peptide sequence contained a consensus ATG initiation codon (ANNATGG; [23,24]). Primers for expression using the α -factor preprosequence needed no initiation codon and the mature coding sequence was fused flush to the Kex2p protease cleavage site (..EKR;). PCR products were cloned into pCR-Script (Stratagene) and subsequently cloned into the respective pPIC plasmid vector using restriction sites incorporated into the primer sequences (Invitrogen; Table 1).

GNA expression constructs

A cDNA sequence encoding LECGNA2 in pUC19, one of several GNA isoforms [25], was used as a template for the amplification of the mature GNA coding sequence (109 residues) to yield the constructs described below and summarized in Table 1. The construct for the expression of GNA using the native signal peptide sequence (GNA:GNA) contained a consensus ATG initiation codon. GNA was fused to the α -factor preprosequence flush with the Kex2p protease cleavage site to yield construct α :GNA. A second α -factor construct contained two Glu-Ala repeats (EAEA) between the Kex2p protease cleavage site and mature GNA.

Mature GNA fused to the PHA-E signal sequence was created by the megaprimer method [26] using two sequential polymerase chain reactions. Firstly, the PHA-E:PHA-E construct (in pPICZ B) was used as a template for amplification of the PHA-E native leader sequence using the 5' *AOX1* primer (Invitrogen) and a C-terminal primer which introduces the first six residues of mature GNA (omitting the methionine initiation codon) immediately after the last six residues of the PHA-E signal peptide. The resulting PCR product was purified and used as a 'megaprimer' in combination with a GNA C-terminal primer to produce the PHA-E:GNA fusion using the LECGNA2 cDNA sequence as a template. In all cases above, PCR products were cloned directly into the appropriate pPIC vector.

Green fluorescent protein (GFP) expression constructs

Primers used for the construction of GFP expression constructs are shown in Table 1. For construct α EA:GFP (fusion of GFP to α -factor containing Glu-Ala repeats) a modified GFP coding sequence was amplified by PCR from pGFPuv (Clontech) introducing a *Pst*I restriction site for 'in-frame-cloning' to the

Table 2. Haemagglutination activity of rabbit erythrocytes by recombinant PHA-E and GNA; lowest concentrations of lectins to give magglutination are given.

Protein	Concentration ($\mu\text{g}\cdot\text{mL}^{-1}$)
PHA standard	0.49
α -E:PHA-E	0.41
PHA-E	0.38
PHA standard	6.3
GNA	3.1
α -E:GNA	3.1

SDS/PAGE are diagnostic of protein glycosylation. PHA (both α - and L-forms) contains potential endogenous N-linked glycosylation sites, and is glycosylated *in planta* [29]; the diffuse bands could thus reflect hyperglycosylation occurring *in vitro*. However, mature GNA is nonglycosylated, and does not contain any potential N-linked glycosylation sites. The diffuse bands in this case therefore cannot be due to glycosylation of the mature GNA sequence.

To resolve differences in apparent molecular mass on SDS/PAGE polypeptides were deglycosylated by treatment with N-glycosidase F. Prior to deglycosylation, affinity-purified PHA produced by the α :GNA construct contains a major polypeptide band at a molecular mass slightly larger than 'native' GNA, and a series of diffuse bands at higher molecular masses (Fig. 2, lane 3). In the deglycosylated protein, the intensity of the major polypeptide band is unchanged (Fig. 2, lane 4). However, after deglycosylation the higher molecular mass diffuse bands resolve to a single polypeptide, of molecular mass approximately 6 kDa greater than native GNA (Fig. 2, lane 4, asterisk). The higher molecular mass diffuse bands must represent glycosylated GNA, and as the glycosylation cannot be taking place on residues present within mature GNA, it must be occurring on incompletely processed α -factor prosequence. In agreement with this conclusion, GNA contained two distinct N-terminal sequences (Fig. 4) with extensions of 9 and 56 amino acids corresponding to part of the α -factor prosequence. The 56 amino acid N-terminal

extension on the larger polypeptide contains two of the three potential N-glycosylation sites present in the α -factor prosequence. Taken together, this demonstrates that GNA expressed using the α -factor prosequence is incompletely processed and highly glycosylated on the partially cleaved α -factor prosequence.

Analysis of the results obtained for PHA expressed from constructs containing the α -factor prosequence produced similar results to GNA. Deglycosylation decreases the indicated molecular mass of 'native', glycosylated PHA (Fig. 1, lanes 1 and 2) due to the removal of the N-linked carbohydrate side chains from the protein, although both before and after N-glycosidase treatment the native protein (E- or L-form) gives a single band after SDS/PAGE and Western blotting. For PHA expressed from the construct containing the α -factor prosequence, culture supernatants contain a major polypeptide detected by anti-PHA antibodies at a similar molecular mass to glycosylated native PHA (Fig. 1, lanes 7 and 1), and a series of diffuse bands at higher molecular mass. A faint band at slightly lower molecular mass is also present. N-glycosidase treatment of culture supernatants results in three bands being observed in culture supernatants on Western blots (Fig. 1, lane 8). Two bands form a close doublet, of similar molecular mass to deglycosylated 'native' PHA. N-glycosidase treatment removes the high molecular mass diffuse bands observed in the culture supernatant on SDS/PAGE, which resolve to a single band of higher indicated molecular mass (approx. 6 kDa) than deglycosylated 'native' PHA (Fig. 1, lane 8, asterisk). This band is not observed in α :PHA samples after purification and deglycosylation (compare lanes 8 and 10 of Fig. 1), showing that it must have come from the glycopolypeptides which produce the diffuse bands on SDS/PAGE, as these do not bind to the affinity column (see above). N-terminal sequencing of PHA expressed from the α :PHA constructs after affinity purification (i.e. corresponding to Fig. 1, lane 10) gave two N-terminal sequences, corresponding to the two closely spaced polypeptides observed in the deglycosylated recombinant protein. The lower molecular mass band of the doublet had the same N-terminal sequence as mature, native PHA produced *in planta*, thus showing that some correct processing had taken place. However, the higher molecular mass band of the doublet had a

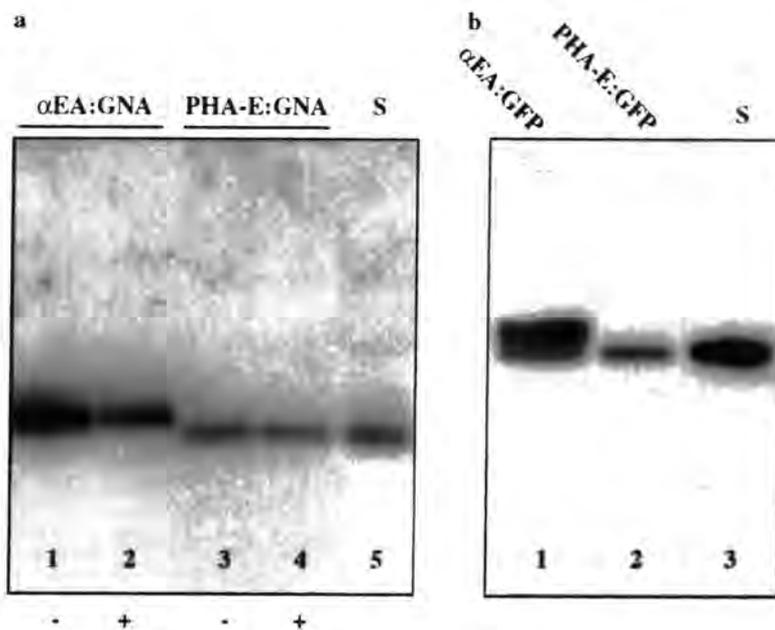


Fig. 2. GNA/GFP expression. (a) GNA expressed in *Pichia* with Glu Ala repeats between the α -factor prosequence (α EA:GNA) or with the PHA-E signal peptide (PHA-E:GNA). (b) GFP expressed in *Pichia* using the α -factor prosequence with Glu-Ala repeats (α EA:GFP) or with the PHA-E signal peptide (PHA-E:GFP). Western analysis of GFP in culture supernatants (S) is standard recombinant protein obtained from Clontech.

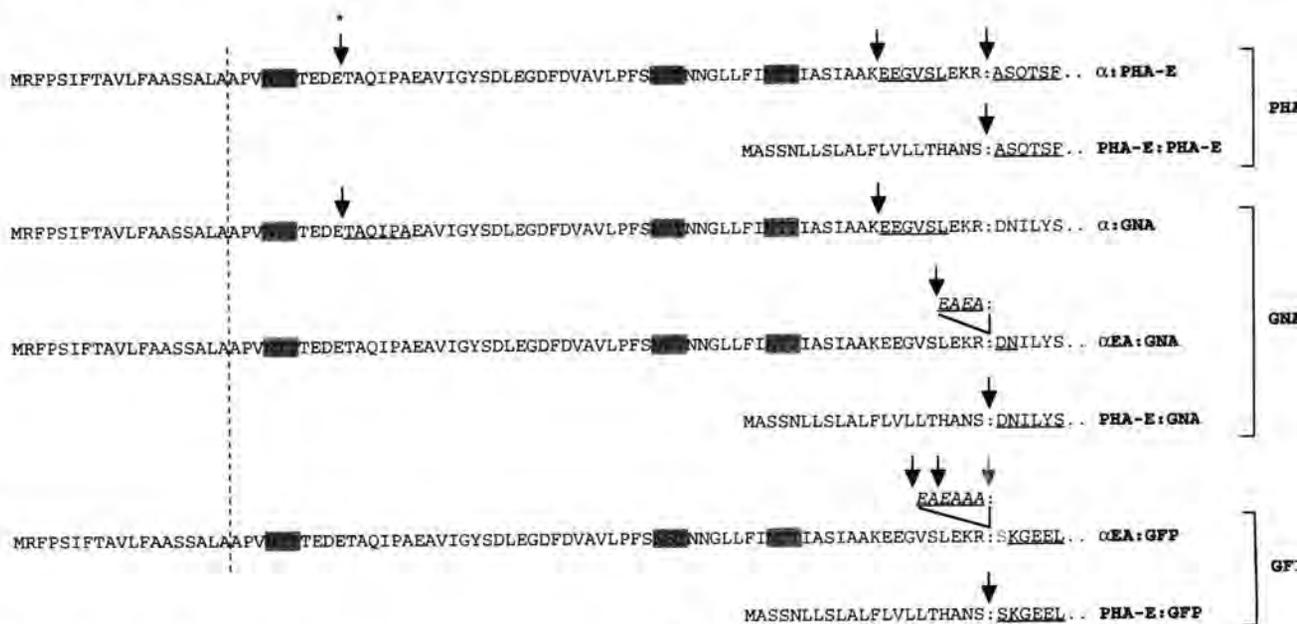


Fig. 4. Summary of processing and N-terminal sequences of lectin and GFP constructs expressed in *Pichia*. Construct designations are as for Table 1. Colon (:) denotes the start of mature protein coding sequences. Full arrows show the sites of processing determined by N-terminal sequencing and the results are underlined. For α EA:GFP, the grey arrow denotes a minor component for which the first residue (in grey) could not be determined. Consensus N-linked glycosylation sites are shown on a grey background. A dashed grey line shows the site of cleavage of the signal peptide of the *Saccharomyces cerevisiae* factor preprosequence. For α :PHA-E, the arrow with an asterisk (*) denotes processing by analogy with α :GNA as an N-terminal sequence was not obtained.

sequence consistent with an N-terminal extension of 9 amino acid residues derived from the C-terminal region of the α -factor preprosequence (Fig. 4). Thus this polypeptide (present in comparable amount to the correctly processed form) was incompletely processed, as observed for the equivalent GNA construct described above. By analogy with the results obtained for GNA (see above) the highly glycosylated forms of PHA most probably contain the 56 amino acid N-terminal extension derived from the α -factor preprosequence. In contrast to GNA, this incompletely processed form does not bind to carbohydrates.

Processing of proteins expressed from constructs containing a modified α -factor preprosequence containing Glu-Ala repeats

As the goal of this work was to produce functional, correctly processed, recombinant lectins, a second GNA construct with the addition of Glu-Ala repeats between the α -factor preprosequence and the GNA mature N-terminus (α EA:GNA; Fig. 4) was prepared. The Glu-Ala repeats have been used to enhance the processing of secreted proteins when the α -factor preprosequence is used, improving the efficiency of cleavage by the Kex2p protease; the Glu-Ala repeats are then cleaved by the Ste13p protease [21,30]. This construct produced functional GNA in the culture supernatant, which on examination by SDS/PAGE and Western blotting, proved to contain a homogeneous polypeptide of molecular mass slightly larger than 'native' GNA (Fig. 3a, lanes 1 and 5). No evidence for the presence of GNA polypeptides with large N-terminal extensions was observed, nor was the protein glycosylated as judged by the lack of a shift in mobility after N-glycosidase F treatment (Fig. 3a, lanes 1 and 2). However, although the inclusion of the Glu-Ala repeats clearly improved processing of the α -factor preprosequence, the protein still contained a

four amino acid residue extension, EAEA, to the correct GNA N-terminus, showing that processing of the Glu-Ala repeats themselves was inefficient.

Similar results were obtained with a construct in which the α -factor preprosequence plus Glu-Ala repeats was fused to a GFP coding sequence (α EA:GFP). Expression in *P. pastoris* resulted in the production of functional protein, which was secreted into the culture supernatant. Analysis of the secreted protein by SDS/PAGE and Western blotting showed that it contained several polypeptides, showing a slight size heterogeneity (Fig. 3b, lane 1). In agreement with the results of SDS-PAGE and electrophoresis, the protein had a heterogeneous N-terminal sequence. The sequence data were consistent with most of the polypeptides containing either a 4 or 6 amino acid N-terminal extension (EAAA or EAEAAA, respectively), although a small fraction of the protein did show evidence of correct processing (Fig. 4). Thus, although the use of a modified α -factor preprosequence including the Glu-Ala repeats can improve processing efficiency in *Pichia*, we conclude that in the case of the proteins studied here the *Saccharomyces cerevisiae* α -factor preprosequence is not optimal for obtaining correctly processed protein.

The PHA signal peptide sequence can be used to produce correctly processed heterologous proteins

As PHA could be successfully expressed in *Pichia pastoris* from constructs containing the protein's 'native' signal peptide sequence, the processing of the protein was examined in more detail (Fig. 1a). PHA expressed from constructs containing the native signal peptide (both E- and L-forms) gave a major band and one minor band on SDS/PAGE both before and after affinity purification; the major band was of a similar molecular mass to native PHA (compare lanes

5 of Fig. 1a). When the recombinant protein was treated with N-glycosidase F, the two bands were resolved to a single deglycosylated polypeptide which migrated at a rate equivalent to deglycosylated native PHA (compare lanes 2, 4 and 6 of Fig. 1a). Finally, the N-terminal sequence of affinity purified recombinant PHA (both E- and L-forms) was determined. A single, homogeneous N-terminal sequence was present, which responded exactly to the N-terminal sequence of mature native PHA as purified from plant sources (Fig. 4). The PHA forms expressed in *Pichia* from constructs containing the 'native' signal peptides are thus correctly processed.

PHA expressed from the constructs containing the native signal peptides did not contain diffuse bands at a high molecular mass on SDS/PAGE, although the protein was deglycosylated. In this case, the glycosylation that took place in *Pichia* resulted in carbohydrate side chains comparable in size to those observed when the protein is synthesized *in planta*. However, the composition of the carbohydrate side chains on PHA synthesized in *Pichia* is likely to differ from that of native PHA, due to differences in the respective glycosylation systems [20]. PHA synthesized in *Pichia* contains α -1,3 or 1,6 linked mannose residues, as GNA (specific for mannose residues) binds to recombinant PHA (result not shown).

The correct processing of PHA expressed from constructs containing the complete preprotein suggested that this signal sequence might be useful for expression of secreted proteins in *Pichia*. In the absence of a readily available alternative, we tested if the PHA signal peptide could be used to (a) obtain correctly processed GNA and (b) direct the secretion and correct processing of an unrelated protein, in this case GFP. The corresponding expression constructs were (a) PHA (E-terminus)-signal peptide fused to mature coding sequences of GNA (PHA-E:GNA) and (b) PHA-E signal peptide fused to GFP (PHA-E:GNA) (Table 1/Fig. 4).

Expression of the PHA-E:GNA construct in *Pichia* resulted in the presence of soluble nonglycosylated protein, reactive with anti-GNA antibodies, in the culture supernatant (Fig. 3a, lanes 3 and 4). This contrasts with attempts to express GNA from constructs containing the native GNA signal peptide, where no expression could be detected. GNA produced from the PHA-E:GNA construct was functional, as shown by affinity purification through binding to a mannan-agarose column, and by hemagglutination assays using a purified protein (Table 2). The purified recombinant GNA contained a single N-terminal sequence, identical to the N-terminal sequence of native GNA (Fig. 4), showing correct processing of the signal peptide.

Similar results were obtained with the PHA-E:GFP construct, where soluble functional GFP accumulated in the culture supernatant. The culture supernatant, after concentration by ultrafiltration, was fluorescent (with no fluorescence observed in the control supernatant), and analysis by SDS/PAGE and Western blotting with anti-GFP antibodies showed a single protein band migrating at a rate equivalent to a GFP standard (Fig. 3b, lane 2). The GFP protein band was blotted and the N-terminal sequence corresponded to the N-terminal sequence of mature GFP (Fig. 4), again demonstrating correct processing of the heterologous signal peptide. These results demonstrate that the PHA-E signal peptide is correctly processed from two heterologous proteins which are directed to a secretory pathway.

DISCUSSION

To our knowledge, this is the first description of the use of a signal peptide from *Pichia pastoris* to produce secreted, functional plant lectins

which can be easily purified from culture supernatants by one-step affinity chromatography. However, there are particular problems in expressing a mannose-specific lectin like GNA in *Pichia pastoris*, due to the presence of high-mannose carbohydrates on the surface of the cells, and secreted in the medium. As is suggested by the gel blot shown in Fig. 2, significant amounts of GNA may be present as complexes with soluble carbohydrates in the culture supernatants. These complexes give rise to the intense smear at, and near, the origin on SDS/PAGE gels, due to incomplete denaturation, and are not retained on mannan-agarose affinity columns. GNA-carbohydrate complexes can be precipitated from the culture medium by increasing the pH, but the protein cannot be recovered from the precipitate without denaturation (result not presented). Thus, although GNA can be purified from culture supernatants, further research will be necessary to optimize the recovery of active GNA from *Pichia*. Interestingly, α :GNA species containing most of the prosequence were present in the protein after affinity purification on the mannan-agarose column, showing that GNA was still functional even with a highly glycosylated N-terminal extension. In contrast, PHA molecules containing the glycosylated N-terminal extension derived from the α -factor prosequence were not retained on an affinity column. This may be due to the carbohydrate side chains interfering with the sugar-binding site in PHA, or altering the structure of the molecule to abolish its activity, or blocking access to the binding site.

N-linked oligosaccharide side chains produced on glycoproteins expressed in *Pichia* have been shown to be mainly of the 'high-mannose' type, where a core structure of NAcglu-(man)₂ is modified by the addition of (mannose)_n branches to the terminal mannose residue. However, the vast majority of these oligosaccharides in *Pichia* contain 8–14 mannose residues, and thus would be expected to increase the molecular mass of a glycoprotein by approximately 2 kDa [20]. The decrease in molecular mass observed when PHA expressed in *Pichia* from the PHA-E:PHA-E construct is deglycosylated is consistent with the molecule containing 1 or 2 carbohydrate side chains of this type, and we have observed that GNA binds strongly to PHA expressed in *Pichia* (result not shown). The mature PHA-E sequence contains three potential N-linked glycosylation sites, although only two of these are utilized *in planta* [29]. The glycosylation that is present on both PHA and GNA expressed in *Pichia* from constructs containing the α -factor preprosequence (α :PHA-E and α :GNA) is different in nature, and appears to involve long carbohydrate side chains resulting from glycosylation on the uncleaved α -factor prosequence. This is likely to be a result of inefficient processing which has previously been shown to result in the secretion of hyperglycosylated unprocessed proteins in *Saccharomyces* ([31] and references therein).

While the yield of lectins obtained from *Pichia* cultures in the present work were fairly low, no attempt was made to optimize yields by selection of clones with multiple inserts [32], or to optimize culture conditions. Consequently, the potential yields of recombinant lectins produced in *Pichia* are likely to be at least an order of magnitude greater than those reported here as, after optimization, yields of > 1 g·L⁻¹ for HIV-1 gp120 [33] and > 10 g·L⁻¹ for a tetanus toxin fragment [32] have been reported.

Many heterologous signal peptides do not function in yeast, either to direct protein secretion or to generate the 'natural' N-terminus of the protein (reviewed in [34]). Early reports (e.g. [35]), demonstrated the secretion and correct processing of

proteins when fused to the α -factor preprosequence. As a consequence, the α -factor preprosequence has been widely used in the synthesis of recombinant proteins in *Saccharomyces* and in *Pichia*, where, as with *Saccharomyces*, it has been shown to direct secretion and correct N-terminal processing in a number of examples [21]. Processing of the α -factor preprosequence in *Saccharomyces* involves several distinct proteolytic cleavage steps, carried out by different enzymes. The 19 amino acid signal peptide (presequence) is removed by a signal peptidase system; further processing of the prosequence then involves the action of an endopeptidase encoded by the *KEX2* gene, which cleaves C-terminally to a specific Lys-Arg sequence, and a dipeptidyl aminopeptidase encoded by the *STE13* gene, which removes N-terminal Glu-Ala repeats [30]. The presence of the Glu-Ala repeats enhances the activity of the Kex2 protease, but subsequent processing of these repeats by the Ste13 aminopeptidase has been found to be inefficient in many cases [34].

The correct processing of the α -factor preprosequence is very much case-dependent and numerous examples have been reported in both *Saccharomyces* and *Pichia* where proteins have failed to give correct processing to yield the mature N-terminus [21]. In *Pichia*, the propeptide is not removed from protein disulphide isomerase and remains, hyperglycosylated, on a proportion of recombinant protein [36]; the pro-peptide is only partially processed from human procarboxypeptidase A2, resulting in heterogeneity [37]; expression of influenza neuraminidase (in *Saccharomyces*), or the b subunit of bovine follicle-stimulating hormone, gave products where the Glu-Ala repeats were not removed [38,39]. The results obtained in the present paper agree with and extend these previous observations. Both PHA and GNA show inefficient processing by the *Pichia* equivalent of the Kex2 protease. Addition of Glu-Ala repeats to the α :GNA construct resulted in efficient cleavage by this enzyme but the products of both the α EA:GNA and α EA:GFP constructs are inefficiently processed by the *Pichia* equivalent of the Ste13 aminopeptidase, resulting in proteins with N-terminal extensions.

Given the problems sometimes associated with the α -factor preprosequence, efforts have been made to develop alternative signal peptides for use in *Pichia* [21]; for example, the *Pichia* acid phosphatase PHO1 signal peptide. This signal peptide fused to mature protein sequences can lead to secretion of either correctly processed protein [40], or protein containing N-terminal heterogeneity [41], again showing case-dependency. A hybrid PHO1 signal peptide, containing a Kex2p cleavage site to improve processing efficiency, has also been used [42]. Some preproteins expressed in *Pichia* produce secreted and correctly processed protein [21,43,44], but it is not known if the corresponding signal peptides can confer correct processing on heterologous proteins. This paper describes such an analysis and demonstrates the secretion and correct N-terminal processing of two unrelated proteins (GNA and GFP) using the PHA-E signal peptide. This is unexpected in view of the failure of this signal peptide to function correctly in *Saccharomyces* [16], whether this reflects a difference in signal peptide recognition between *Saccharomyces* and *Pichia* or some other factor(s) remains to be determined. The amount of protein accumulating in the culture supernatant appears to be slightly less with constructs using the PHA-E signal peptide compared to equivalent constructs using the α -factor preprosequence (unpublished observations). Notwithstanding, these results suggest the PHA-E signal peptide could be used for proteins where secretion and processing using other signal peptides has proven problematic.

ACKNOWLEDGEMENTS

Funding from the Scottish Office Agriculture, Environment and Directorate (Programme FF821) is gratefully acknowledged.

REFERENCES

- van Damme, E.J.M., Peumans, W., Pustzai, A. & Bardocz, S. (1990) *Handbook of Plant Lectins: Properties and Biomedical Applications*. John Wiley and Sons Ltd, Chichester, UK.
- van Damme, E.J.M., Barre, A., Smeets, K., Torrekens, S., van Leeuwen, F., Rouge, P. & Peumans, W. (1995) The bark of *Rosa pseudoacacia* contains a complex mixture of lectins. *Plant Physiol.* **107**, 833–843.
- Young, N., Watson, D., Yaguchi, M., Adar, R., Arango, R., Rodriguez-Arango, E., Sharon, N., Blay, P. & Thibault, P. (1995) C-terminal post-translational proteolysis of plant lectins and their recombinant forms expressed in *Escherichia coli*. *J. Biol. Chem.* **270**, 2563–2568.
- Miller, J., Hsu, R., Henrikson, R. & Yachnin, S. (1975) Extensive homology between the subunits of the phytohemagglutinin mitogenic proteins derived from *Phaseolus vulgaris*. *Proc. Natl. Acad. Sci. USA* **72**, 1388–1391.
- Feldsted, R., Egorin, M., Leavitt, R. & Bachur, N. (1977) Recombinant subunits of *Phaseolus vulgaris* isolectins. *J. Biol. Chem.* **252**, 2967–2971.
- Leavitt, R., Feldsted, R. & Bachur, N. (1977) Biological and biochemical properties of *Phaseolus vulgaris* isolectins. *J. Biol. Chem.* **252**, 2961–2966.
- Hoffman, L. & Donaldson, D. (1987) Synthesis of mitogenic phytohemagglutinin-L in *Escherichia coli*. *Bio/Technol.* **5**, 157–160.
- Rodriguez-Arango, E., Arango, R., Adar, R., Galili, G. & Sharon, N. (1992) Cloning, sequence analysis and expression in *Escherichia coli* of the cDNA encoding a precursor of peanut agglutinin. *FEBS Lett.* **307**, 185–189.
- Zhu, K., Huesing, J., Shade, R., Bressan, R., Hasegawa, P. & Mu, L. (1996) An insecticidal *N*-acetylglucosamine-specific lectin from *Griffonia simplicifolia* (Leguminosae). *Plant Physiology* **110**, 195–202.
- Longstaff, M., Powell, K.S., Gatehouse, J.A., Raemaekers, R., Mowbray, C.A. & Hamilton, W.D.O. (1998) Production and purification of active snowdrop lectin in *Escherichia coli*. *Eur. J. Biochem.* **215**, 59–65.
- Arango, R., Adar, R., Rozenblatt, S. & Sharon, N. (1992) Expression of *Erythrina corallodendron* lectin in *Escherichia coli*. *Eur. J. Biochem.* **205**, 575–581.
- Adar, R., Streicher, H., Rozenblatt, S. & Sharon, N. (1997) Synthesis of soybean agglutinin in bacterial and mammalian cells. *Eur. J. Biochem.* **249**, 684–689.
- Sheldon, P. & Bowles, D. (1992) The glycoprotein precursor concanavalin A is converted to an active lectin by deglycosylation. *EMBO J.* **11**, 1297–1301.
- Min, W., Dunn, A. & Jones, D. (1992) Non-glycosylated recombinant pro-concanavalin A is active without polypeptide cleavage. *EMBO J.* **11**, 1303–1307.
- Dao-Thi, M.-H., Hamelryck, T., Poortmans, F., Voelker, T., Christmann, M. & Wyns, L. (1996) Crystallisation of glycosylated and nonglycosylated phytohaemagglutinin-L. *Proteins* **24**, 134–138.
- Tague, B. & Chrispeels, M. (1987) The plant vacuolar phytohemagglutinin, is transported to the vacuole of transgenic tobacco cells. *J. Cell Biol.* **105**, 1971–1979.
- von Schaeuwen, A. & Chrispeels, M. (1993) Identification of vacuolar sorting information in phytohemagglutinin, an unprocessed vacuolar protein. *J. Exp. Bot.* **44** (Suppl.), 339–342.
- Chao, Q. & Etzler, M. (1994) Incorrect targeting of plant vacuolar lectins in yeast. *J. Biol. Chem.* **269**, 20866–20871.
- Nagahora, H., Ishikawa, K., Niwa, Y., Muraki, M. & Jigami, Y. (1994) Expression and secretion of wheat germ agglutinin by *Saccharomyces cerevisiae*. *Eur. J. Biochem.* **210**, 989–997.
- Cregg, J., Vedvick, T. & Raschke, W. (1993) Recent advances

- expression of foreign genes in *Pichia pastoris*. *BioTechnol.* **11**, 905–910.
- Sreekrishna, K., Brankamp, R., Kropp, K., Blankenship, D., Tsay, J.-T., Smith, P., Wierschke, J., Subramaniam, A. & Birkenberger, L. (1997) Strategies for the optimal synthesis and secretion of heterologous proteins in the methylotrophic yeast *Pichia pastoris*. *Gene* **190**, 55–62.
- Hoffman, L. & Donaldson, D. (1985) Characterization of two *Phaseolus vulgaris* phytohemagglutinin genes closely linked on the chromosome. *EMBO J.* **4**, 883–889.
- Kozak, M. (1987) An analysis of 5'-noncoding sequences from 699 vertebrate messenger RNAs. *Nucleic Acids Res.* **15**, 8125–8148.
- Kozak, M. (1990) Downstream secondary structure facilitates recognition of initiator codons by eukaryotic ribosomes. *Proc. Natl Acad. Sci. USA* **87**, 8301–8305.
- van Damme, E.J.M., Declercq, N., Claessens, F., Hemschoote, K., Peeters, B. & Peumans, W.J. (1991) Molecular-cloning and characterization of multiple isoforms of the snowdrop (*Galanthus nivalis* L.) lectin. *Planta* **186**, 35–43.
- Sarkar, G. & Sommer, S. (1990) The megaprimer method of site-directed mutagenesis. *Biotechniques* **8**, 404–407.
- Khyse-Andersen, J. (1974) Electrophoretic blotting of multiple gels: a simple apparatus without buffer tank for rapid transfer of proteins from polyacrylamide to nitrocellulose. *J. Biochem. Biophys. Meth.* **10**, 203–209.
- Osborn, T., Manen, J.-F., Brown, J. & Bliss, F. (1984) Bean lectins IV: genetic variation in the non-denatured structure of lectins from different *Phaseolus vulgaris* L. cultivars. *Theor. Appl. Genet.* **67**, 547–552.
- Sturm, A. & Chrispeels, M. (1986) The high mannose oligosaccharide of phytohemagglutinin is attached to asparagine-12 and the modified oligosaccharide to asparagine-60. *Plant Physiol.* **80**, 320–322.
- Brake, A. (1989) Secretion of heterologous proteins directed by the yeast α -factor leader. In *Yeast Genetic Engineering* (Barr, P.J., Brake, A.J. & Valenzuela, P., eds), pp. 269–280. Butterworth Publishers, Stoneham, MA, USA.
- Kjeldsen, T., Hach, M., Balschmidt, P., Havelund, S., Pettersson, A. & Markussen, J. (1998) Prepro-leaders lacking N-linked glycosylation for secretory expression in the yeast *Saccharomyces cerevisiae*. *Prot. Exp. Purific.* **14**, 309–316.
- Clare, J., Rayment, F., Ballantine, S., Sreekrishna, K. & Romanos, M. (1991) High-level expression of tetanus toxin fragment C in *Pichia pastoris* strains containing multiple tandem integrations of the gene. *BioTechnol* **9**, 455–460.
- Scorer, C., Buckholz, R., Clare, J. & Romanos, M. (1993) The intracellular production and secretion of HIV-1 envelope protein in the methylotrophic yeast *Pichia pastoris*. *Gene* **136**, 111–119.
34. Tuite, M. (1991) Expression of heterologous genes. In *Saccharomyces* (Tuite, M. & Oliver, S., eds), pp. 169–212, Plenum Press, New York, USA.
35. Zsebo, K., Lu, H.-S., Fieschko, J., Goldstein, L., Davis, J., Duker, K., Suggs, S., Lai, P.-H. & Bitter, G. (1986) Protein secretion from *Saccharomyces cerevisiae* directed by the prepro- α -factor leader region. *J. Biol. Chem.* **261**, 5858–5865.
36. Vuorela, A., Myllyharju, J., Nissi, R., Pihljaniemi, T. & Kivirikko, K. (1997) Assembly of human prolyl 4-hydroxylase and type III collagen in the yeast *Pichia pastoris*: formation of a stable enzyme tetramer requires coexpression with collagen and assembly of a stable collagen requires coexpression with prolyl 4-hydroxylase. *EMBO J.* **16**, 6702–6712.
37. Reverter, D., Ventura, S., Villegas, V., Vendrell, J. & Aviles, F. (1998) Overexpression of human procaryboxypeptidase A2 in *Pichia pastoris* and detailed characterization of its activation pathway. *J. Biol. Chem.* **273**, 3535–3541.
38. Martinet, W., Saelens, X., Deroo, T., Neiryneck, S., Contreras, R., Min Jou, W. & Fiers, W. (1997) Protection of mice against a lethal influenza challenge by immunization with yeast-derived recombinant influenza neuraminidase. *Eur. J. Biochem.* **247**, 332–338.
39. Samaddar, M., Catterall, J. & Dighe, R. (1997) Expression of biologically active b subunit of bovine follicle-stimulating hormone in the methylotrophic yeast *Pichia pastoris*. *Prot. Exp. Purif.* **10**, 345–355.
40. Weiss, H., Haase, W., Michel, H. & Reilander, H. (1995) Expression of functional mouse 5-HT_{5A} serotonin receptor in the methylotrophic yeast *Pichia pastoris*: pharmacological characterization and localization. *FEBS Lett.* **377**, 21–25.
41. O'Donohue, M., Boissy, G., Huet, J.-C., Nespoulus, C., Brunie, S. & Pernellet, J.-C. (1996) Overexpression in *Pichia pastoris* and crystallization of an elicitor protein secreted by the phytopathogenic fungus, *Phytophthora cryptogea*. *Prot. Exp. Purif.* **8**, 254–261.
42. Laroche, Y., Storme, V., de Meutter, J., Messens, J. & Lauwreys, M. (1994) High-level secretion and very efficient isotopic labelling of tick anticoagulant peptide (TAP) expressed in the methylotrophic yeast, *Pichia pastoris*. *BioTechnol.* **12**, 1119–1124.
43. Juge, N., Andersen, J., Tull, D., Roepstorff, P. & Svensson, B. (1996) Overexpression, purification, and characterization of recombinant barley α -amylases 1 and 2 secreted by the methylotrophic yeast *Pichia pastoris*. *Prot. Exp. Purif.* **8**, 204–214.
44. Ferrari, E., Lodi, T., Sorbi, R., Tirindelli, R. & Cavaggioni, A. (1997) Expression of a lipocalin in *Pichia pastoris*: secretion, purification and binding activity of a recombinant mouse major urinary protein. *FEBS Lett.* **401**, 73–77.

