

**An improved reporter system
based on a novel β -glucuronidase (GUS)
from *Staphylococcus* sp.**

by

Tuan Anh Nguyen, *BSc (Cantho), MSc (London)*

A thesis submitted for the degree of Doctor of Philosophy
of the Australian National University

Center for the Application of Molecular Biology
to International Agriculture (CAMBIA)
and
Research School of Biological Sciences



THE AUSTRALIAN
NATIONAL UNIVERSITY

September 2002

Table of contents

Declaration	i
Acknowledgements.....	ii
Abstract	iii
Table of contents	iv

1 Chapter I: General Introduction..... 1

1.1 Reporter genes as tools to study regulation of gene expression.....	1
1.2 Plant reporter systems	4
1.2.1 CAT and NPTII	4
1.2.2 Luciferase	5
1.2.3 GFP	7
1.2.4 GUS.....	9
1.2.5 Other less popular reporter systems.....	18
1.3 Limitations and common pitfalls in the use of reporter genes.....	19
1.4 Protein secretion in eukaryotes	21
1.4.1 Targeting to and translocation across the ER membrane.....	21
1.4.2 In the lumen of the ER	22
1.4.3 Transport from ER to Golgi.....	26
1.4.4 Transport within the Golgi apparatus.....	27
1.4.5 Transport from <i>Trans</i> Golgi Network to the cell surface.....	31
1.5 Previous work on the development of better or secreted GUS	32
1.6 Aims of the project.....	34

2 Chapter II: Construction of the synthetic *gusA^{Ssp}* gene and its variants..... 36

2.1 Introduction	36
2.1.1 Strategies for synthetic gene construction	36
2.1.2 Construction strategy for the <i>gusA^{Ssp}</i> gene	38
2.2 Materials and Methods	39
2.2.1 Oligonucleotides	39
2.2.2 Enzymes, chemicals, plasmids and bacterial strains	39
2.2.3 General methods.....	39
2.2.4 Gene construction	39

2.2.5	Site-directed mutagenesis	41
2.2.6	Quantitative GUS assay on permeabilized <i>E. coli</i> cells.....	42
2.2.7	Hexahistidine-tagged protein purification	42
2.3	Results and Discussions	43
2.3.1	A codon-optimized version of the <i>gusA</i> ^{Ssp} gene constructed.....	43
2.3.2	Lessons learned from the construction.....	48
2.3.3	Non-silent changes in the coding sequence reversed using site-directed mutagenesis	49
2.3.4	Several 5' end variations of the synthetic gene created.....	52
2.3.5	Cysteine-free, N-glycosylation-free, and cysteine-and-N-glycosylation-free GUS ^{Ssp} variants created with reasonable GUS activity retained	58

3 Chapter III: Purification of β -glucuronidases and preparation of their antibodies 61

3.1	Introduction	61
3.2	Materials and Methods	63
3.2.1	Chemicals, plasmids and bacterial strains	63
3.2.2	Affinity purification of β -glucuronidases using thiophenyl-glucuronide.....	64
3.2.3	Affinity purification of β -glucuronidases using saccharolactone	64
3.2.4	Affinity purification of 6xhis-tagged β -glucuronidases using Ni-NTA resin	64
3.2.5	SDS-PAGE	65
3.2.6	Production of rabbit polyclonal antibodies against GUS ^{Eco} and GUS ^{Ssp}	65
3.2.7	Affinity purification and titering of GUS ^{Eco} and GUS ^{Ssp} antibodies	65
3.2.8	Immunoblotting and immunodetection.....	66
3.2.9	Protein extractions	66
3.3	Results and Discussions	68
3.3.1	Affinity purification of β -glucuronidases with thiophenyl-glucuronide	68
3.3.2	Affinity purification of β -glucuronidases with saccharolactone.....	70
3.3.3	Affinity purification of 6xhis-tagged GUS ^{Eco} and GUS ^{Ssp} with Ni-NTA resin.....	73
3.3.4	Production, affinity purification, and titering of rabbit polyclonal antibodies against GUS ^{Eco} and GUS ^{Ssp}	75
3.3.5	Optimization of immunodetection methods	75

4 Chapter IV: Expression and secretion studies of GUS^{Eco} and GUS^{Ssp} in yeast..... 79

4.1	Introduction	79
4.2	Materials and Methods	81
4.2.1	Yeast strain and transformation protocols.....	81

4.2.2	DNA constructs.....	81
4.2.3	Media and growth conditions.....	82
4.2.4	Indicator plates	82
4.2.5	Hot-agarose overlay assay.....	82
4.2.6	GUS secretion assay	82
4.3	Results and Discussions	83
4.3.1	Establishment of yeast indicator plate assay	83
4.3.2	Both GUS ^{Eco} and GUS ^{Ssp} are efficiently secreted with the invertase signal sequence	88
4.3.3	C499A and N118Q variants of GUS ^{Ssp} are also efficiently secreted	91
5	Chapter V: Expression and secretion studies of GUS^{Eco} and GUS^{Ssp} in plants	93
5.1	Introduction	93
5.2	Materials and Methods	96
5.2.1	DNA constructs.....	96
5.2.2	Rice apoplastic protein extraction	96
5.2.3	GUS quantitative and histochemical assays	97
5.2.4	Detection of post-translational modifications of GUS ^{Eco} and GUS ^{Ssp}	97
5.3	Results and Discussions	98
5.3.1	SP-GUS ^{Ssp} plants have strong histochemical GUS staining despite their relatively low total GUS activity	98
5.3.2	SP-GUS ^{Ssp} plants have GUS activity in the apoplastic fluid extract.....	102
5.3.3	SP-GUS ^{Ssp} reveals cell-wall localized GUS activity.....	105
5.3.4	Detection of post-translational modifications of GUS ^{Eco} and GUS ^{Ssp}	107
6	Chapter VI: Investigation of some applications based on GUS^{Ssp}	111
6.1	Introduction	111
6.1.1	GUS ^{Ssp} as an improved reporter	111
6.1.2	GUS ^{Ssp} as non-destructive reporter	112
6.1.3	GUS ^{Ssp} in positive selection schemes	113
6.1.4	GUS ^{Ssp} in tissue-specific manipulation systems.....	114
6.2	Materials and Methods	115
6.2.1	DNA constructs.....	115
6.2.2	Transformation protocols for rice, tobacco and <i>Arabidopsis</i>	115
6.2.3	Effects of fixation on histochemical GUS assay.....	115
6.2.4	Non-destructive histochemical GUS assays	116
6.2.5	DNA southern blot.....	116

6.3 Results and Discussions	117
6.3.1 An intron introduced to the synthetic <i>gusA^{Ssp}</i> gene	117
6.3.2 Demonstration of improved histochemical assays with GUS ^{Ssp}	119
6.3.3 Demonstration of non-destructive GUS staining in various plant tissues	125
6.3.4 Demonstration of potential use of GUS ^{Ssp} as a tool for the manipulation of specific cells and tissues	130

7 Chapter VII: Conclusion and prospects	133
--	------------

Literature cited

Appendix 1

Appendix 2

Declaration

The contents of this thesis are the results of my original work, except where otherwise acknowledged. No material in this thesis has been submitted for any degree or diploma at any other university or institution.

Some of the work in the thesis is in the process of submission to scientific journals, or has been published as conference posters or proceedings:

- ♣ Nguyen, Tuan, Peter Wenzl, Selvameena Rajagopal, Murray Badger, Andrzej Kilian, Jorge Mayer, and Richard Jefferson. Biochemical characterization and preliminary secretion analysis of a new microbial beta-glucuronidase produced from a synthetic gene. *Poster, International Program on Rice Biotechnology meeting, 20-24 Sept 1999, Phuket, Thailand.*
- ♣ Nguyen, Tuan, Selvameena Rajagopal, Murray Badger, Andrzej Kilian, Paul Keese, and Richard Jefferson. A synthetic reporter gene encoding a microbial beta-glucuronidase: design, construction, and expression in heterologous systems. *Abstract and poster, International Program on Rice Biotechnology meeting, 20-24 Sept 1999, Phuket, Thailand.*
- ♣ Nguyen, Tuan, Peter Wenzl, Selvameena Rajagopal, Andrzej Kilian, Jorge Mayer, Paul Keese, Murray Badger, and Richard Jefferson. An improved plant reporter system based on a novel microbial beta-glucuronidase (GUS). *To be submitted to The Plant Journal.*

Tuan A. Nguyen

September 2002

Acknowledgements

Firstly, I would like to thank my main supervisor, Dr. Richard A. Jefferson, for his guidance and encouragement, both in and outside science. I thank him for giving me the opportunity to be one of his grad students. His strong leadership, remarkable knowledge and admirable philosophy have always been inspirational to me personally. I certainly look forward to our continued close contact in the future, RAJ!

I would also like to thank Prof. Murray Badger, who chairs my supervisory panel, for his scientific advice, and for helping me with the administrative work. Special thanks are due to my other supervisors, Drs. Jorge Mayer and Dean Price, for their support and understanding. I should especially thank Jorge for sharing his amazing wealth of knowledge with me on a daily basis.

Secondly, I would like to express my appreciation to all past and present members of the CAMBIA team. In particular:

Drs. Andrzej Kilian and Paul Keese, for their excellent mentoring, and guidance in my research.

Drs. Carol Nottenburg, Eric Huttner, and Peter Wenzl, for their expert help and advice.

My fellow PhD students: Leon Smith, Satya Nugroho, Wei Yang, Thach Tran, Fu Xiqin, Damien Jaccoud, Sujin Patarapuwadol and Sri Koerniati, for all the discussion and stuff we shared.

Selvameena Rajagopal, Mazz Irwin, Emma Hely, and Lijun Tian, for their help in tissue culture and handling of transgenic plants in the glasshouse. Liz Deaves and Chris Pratt for the best human resource support I can get.

Thirdly, I would like to thank friends and people in RSBS, and CSIRO Entomology & Plant Industry, who have always been helpful when help is needed. I am also grateful to many of my new friends in Australia: Tuong, Erica, Minh, Gia, Sharon, the Tequila team... Their camaraderie has made my time here in Canberra very enjoyable.

I am very thankful to the Rockefeller Foundation for providing me with the PhD scholarship. I wish to extend my appreciation to the Foundation's administrators for being so supportive and efficient in helping their scholars Down Under.

I wish to thank my bosses and friends from my home institute in Vietnam (CLRRI) for their kind support and encouragement during my study in Australia.

Finally, I would like to dedicate this work to my family: my parents, brother, and especially my fiancé, Shigeko Kurokawa. Without their love and support, I would not have been here writing these lines. I owe them the biggest THANK YOU!

Abstract

The GUS reporter system, based on the enzyme beta-glucuronidase (GUS) from *E. coli*, has been the most widely used tool in plant molecular biology in the last decade. This thesis reports various studies on a different GUS with superior biochemical characteristics, isolated from a soil *Staphylococcus* sp. bacteria.

Because of the high AT content of the native coding gene, a codon-optimized version of it was constructed and tested in *E. coli*, yeast and plants. The protein was expressed in *E. coli* and purified to high homogeneity for biochemical studies and antibody production.

In yeast, both *E. coli* GUS and *Staphylococcus* GUS, when provided with the invertase signal peptide, were efficiently secreted, and mostly localized in the periplasmic space. This is the first report of efficient secretion of *E. coli* GUS in yeast. All *Staphylococcus* GUS variants (wildtype, N118Q, and C499A mutants) were also secreted with high efficiency.

In plants, only *Staphylococcus* GUS was secreted when provided with either the GRP or extensin signal peptide. *E. coli* GUS was not secreted, consistent with earlier reports. The use of an intron-containing version of the synthetic *Staphylococcus gusA* as an improved reporter in plants was validated. *Staphylococcus* GUS was shown to offer faster and more sensitive histochemical detection. Its activity was also better detected after tissue fixation, allowing more precise histochemical localization of the enzyme. With secreted *Staphylococcus* GUS, non-destructive GUS assays were possible with tissues such as calli and roots. Tissue specific manipulation system using secreted *Staphylococcus* GUS was also demonstrated, with the example of the secreted *Staphylococcus* GUS driven by a pollen-specific promoter ntp303.

1 Chapter I: General Introduction

1.1 Reporter genes as tools to study regulation of gene expression

Regulation of gene expression is a dominant theme in modern molecular biology. Among various tools available for the study of gene expression, the use of reporter genes has become indispensable. Broadly defined, reporter genes are genes whose products can be easily monitored or assayed without undue background interference from the host. Once developed, reporter assays in a given system can be readily adapted to almost any gene of interest, without the need to develop separate assays for individual gene products, which are sometimes very difficult or laborious.

Typically, a reporter gene is fused, or associated, with a regulatory sequence in either a controlled or arbitrary manner, e.g. in artificial DNA constructs, or by random insertion into the genome. The results of such engineering can be divided into two main classes: *transcriptional fusions*, if the protein encoded consists only of amino acids derived from the reporter; or *translational fusions*, if the protein is encoded by both reporter and 'regulator' sequences (e.g. fusion of signal peptide sequence to a reporter gene to monitor protein transport).

There are a variety of ways in which reporter genes can be used to study gene expression. Fusions of reporter genes to constitutive promoters, for example, have been instrumental in the development of many transformation methods, as well as the routine monitoring of transformation experiments. While such use of reporter genes *per se* is not directly related to gene expression studies, it is a prerequisite, since the use of reporter genes is only possible in systems where transformation is available.

Provided that transformation is available, reporter genes can facilitate and expand the ways in which regulation of gene expression can be studied. The precise fusion of a

reporter gene to a regulatory sequence provides many insights to the complex gene regulation process that are very difficult or laborious to obtain otherwise. For example, fusions to reporter genes can be designed to help characterize the specific signals that exert controls on the initiation of transcription (e.g. promoters, enhancers, repressor binding sequences); stability, transport or processing of mRNAs (e.g. 5'- and 3'- untranslated regions, introns and alternative mRNA splicing, or polyadenylation signals); translation efficiency (e.g. the context surrounding translation initiation site, or specific mRNA secondary structures); protein degradation (e.g. stabilizing and destabilizing N-terminal amino acids).

In broader terms, studies on regulation of gene expression can be extended to the subcellular localization and/or subcellular-specific processing and degradation of a gene product of interest. Precise translational fusion of a reporter to various targeting signal sequences, or to the protein of interest itself, has been a powerful tool to elucidate these complex processes.

Random, or at least arbitrary, insertion of a reporter gene into the genome has become a very popular strategy in high-throughput screening and identification of gene regulatory sequences (and of genes, for that matter) at the whole-genome scale. This is most widely seen in the examples of 'enhancer trap' strategies. Enhancer traps can be either primary (direct) or secondary (indirect). In the former case, a reporter gene is "pre-disposed" to express by linkage with (e.g.) a minimal promoter (Sundaresan et al. 1995). In the latter case, the primary enhancer trap (or responder) is actually a transcriptional activator, which can act – in *trans* – on a reporter gene that has been manipulated to respond to only that activator (Croston et al. 1992). The availability of "libraries" of such enhancer trap lines (Klimyuk et al. 1995; Campisi et al. 1999; Kiegle et al. 2000), that have harnessed regulatory sequences, allows gain- and loss-of-function mutations to be created at will, even without prior information on the exact DNA sequence, and prior information on how these sequences actually work. This

represents a powerful approach to create novel variants, to be studied and used with conventional forward genetics, and allows quick and convenient reverse genetic analysis if need arises. At the application level, this allows researchers to utilize the existing complex and elegant gene regulation machinery in living organisms, to turn genes on and off virtually where and when they want. In addition, a reporter such as GUS (β -glucuronidase) can also act as an *effector*, due to its ability to convert non-active precursors to bioactive chemicals (section 1.2.4.3), so it could be used in combination with these identified regulatory sequences to manipulate certain cells and tissues at the researchers disposal. This approach for large-scale identification of gene regulation sequences is an essential part of functional genomics, and will become increasingly important in the “post-genomics” era.

Reporter genes are also invaluable tools to understand the regulation of individual members of multi-gene families, whose products are very similar (therefore nearly impossible to assay separately by other techniques such as RNA analysis or immunological detection), but can be regulated differentially during development. Some reporter genes offer extreme sensitivity, and therefore can be used to study regulation of genes whose products are of low abundance, and thus not measurable by other means.

The development of *in vivo* reporters, such as GFP (green fluorescent protein), has made it possible to monitor gene expression live in its more or less natural cellular context, rather than in homogenized, lysed, permeabilized, frozen or otherwise abused tissues. The unique power of an *in vivo* reporter is that it allows non-destructive sampling not possible by using other reporters, and in fact, by any other means, so that the material can be further studied or propagated.

In brief, reporter genes have become powerful tools that have not only facilitated, but also expanded the ways gene expression can be studied. In the future, new and improved reporters and assay methods that allow increased sensitivity of detection will

certainly contribute to a higher-resolution view of gene expression not seen before, whether one is considering the regulation of a single gene, or the bigger picture of gene regulation networks in general. Developing new and improved reporters that offer *in vivo* and highly sensitive analysis will clearly be the focus of future work. Currently, the intrinsic low level of detection sensitivity is the main factor preventing GFP to become the “holy grail” reporter. Reporters with enzymatic signal amplification capability, such as GUS or luciferase, could overcome such limitations, provided that non-destructive and non-interfering ways to bring the substrate and enzyme together is addressed. In this aspect, many possible paths can be explored: use of novel membrane permeable substrates; use of permeases to actively bring substrate inside the cell; use of reporter enzymes that secrete to the extracellular space of the cell, possibly with an anchoring sequence to retain them within the secreting cell. Effective delivery of substrates to the cells or tissues of interest is also a key issue to be addressed. For example, at a whole organism level, uptake and systemic transport of substrates needs further studies. Lessons from the delivery of drugs, herbicides, and pesticides in various systems will provide a good starting point for such studies.

1.2 Plant reporter systems

In higher plants, more than a dozen reporters have been used. Here I describe some of the popular systems, based on chloramphenicol acetyltransferase, neomycin phosphotransferase, luciferase, green fluorescent protein, and β -glucuronidase. Some of the historical and less popular reporters are very briefly described.

1.2.1 CAT and NPTII

The bacterial *cat* gene from transposon Tn9, encoding chloramphenicol acetyltransferase (CAT), is an early reporter used in plant biology. CAT detoxifies chloramphenicol by transferring an acetyl group from acetyl-coenzyme A to the free

hydroxyl of chloramphenicol. For a recent review on CAT fusions, see Bullock & Gorman (2000).

The bacterial *nptII* (*neo*) gene from transposon Tn5, encoding neomycin phosphotransferase (NPTII), is mainly known as a selectable marker, but has also been used as a plant reporter. NPTII phosphorylates kanamycin to confer kanamycin resistance. For a recent review on the NPTII assay on plants, see Curtis et al. (1995).

CAT and NPTII have specificities not normally found in plant tissues (Fraley et al. 1983; Bevan et al. 1983; Herrera-Estrella et al. 1992; Davis et al. 1992). Additionally, NPTII can tolerate N-terminal fusions, therefore is useful for studying organelle transport in plants (Van den et al. 1985; Barnes 1990). The main disadvantage of the CAT and NPTII systems is their relatively difficult, tedious and expensive assays (radioisotopes and TLC or HPLC), and the inability to readily localize activity to particular cells or tissues *in situ*. Besides, plants often possess some endogenous enzymes with broad substrate specificity that could cause some background activity. This limits the sensitivity and quantitative validity of the assay (Gorman et al. 1982; Reiss et al. 1984; Davey et al. 1995; Curtis et al. 1995). Neither CAT nor NPTII are widely used in plant molecular biology, having been superseded in the late eighties by GUS.

1.2.2 Luciferase

Luciferases are a class of enzymes that generate visible light through catalyzing chemical transformations of a number of substrates. Luciferases have been isolated from a number of organisms, for example, beetle (firefly – *Photinus pyralis*, click beetle – *Pyrophorus plagiophthalmus*), coelenterate (sea pansy – *Renilla reniformis*), ostracod (sea firefly – *Vargula hilgendorffii*), luminous marine and terrestrial bacteria (*Vibrio fischeri*, *Vibrio harveyi*) (Thompson et al. 1989; Meighen 1993; Baldwin et al. 1995; Fisher et al. 1995; Langridge & Szalay 1998).

The light-producing reactions by luciferases are complex and different between classes of organisms. For example, bacterial luciferase uses FMNH₂, O₂, and a ten-carbon aldehyde decanal, whereas beetle and coelenterate luciferases use ATP, Mg²⁺, O₂, and the substrate luciferin (Hori et al. 1973; McElroy & DeLuca 1983; Thompson et al. 1989).

Luciferase offers rapid and sensitive assays. The rapid turnover of luciferases makes them particularly suitable for monitoring dynamics of gene transcription (White et al. 1995; Brandes et al. 1996). Furthermore, different natural or mutant luciferases and different substrates are available, to produce different emitted colors from blue to red (Wood et al. 1989; Branchini et al. 1989; Kajiyama & Nakano 1991; Branchini et al. 1999). More stable variants with improved thermal stability (Kajiyama & Nakano 1994), and increased resistance to proteases (Thompson et al. 1997) have been developed. A dual-luciferase assay system, using green-light-emitting firefly luciferase and blue-light-emitting *Renilla* luciferase, has been developed (Grentzmann et al. 1998), and is available commercially (Promega Corp., WI, USA).

There have been various strategies developed for non-destructive luciferase assays: the use of membrane permeant and photolysable firefly luciferin esters (Craig et al. 1991; Yang & Thomason 1993); the use of *Renilla* luciferase which uses the membrane permeant coelenterazine substrate (Lorenz et al. 1991); and the production of the aldehyde substrate inside bacterial cell (Manen et al. 1997). In particular, the use of secreted luciferases has been quite successful. *Renilla* luciferase has been engineered to be a secreted reporter used in mammalian cells (Liu et al. 1997; Liu & Escher 1999; Liu et al. 2000). More interestingly, *Vargula* luciferase is naturally secreted (Thompson et al. 1989). It has been used quite successfully as a secreted reporter to monitor gene expression in mammalian cells (Thompson et al. 1990; Inouye et al. 1992; Thompson et al. 1995; Tanahashi et al. 2001). This enzyme has also been shown to tolerate N-terminal fusions (Maeda et al. 1996; Maeda et al. 1997).

Although bacterial luciferase (*luxAB* genes) has been used as reporter in plants (Langridge et al. 1994; Langridge & Szalay 1998), firefly luciferase (*luc* gene) is perhaps more popularly used as a plant reporter (Gould & Subramani 1988; Barnes 1990; Schneider et al. 1990; Luehrsen et al. 1992; Nass & Scheel 2001). In plants, luciferase offers rapid and sensitive assay with minimal background, since plants contain little or no endogenous luminescence. That said, the many chromophores and other light-absorbing compounds in plants could cause a substantial absorption and thus diminution of both sensitivity and reliability.

Most of the development of improved luciferase variants, and non-destructive luciferase assays cited above, has been carried out in animal systems. To my knowledge, there has been no report on the use of non-destructive luciferase assay in plants. Nevertheless, the successful use of secreted luciferases in animal systems should validate the proof-of-concept, and provide useful lessons for the development of a secreted GUS reporter system in plants.

On the negative side, luciferases are labile and have short half-life (Pazzagli et al. 1992). The enzymatic reactions are relatively complicated and depend on the provision of multiple reagents, therefore data normalization is difficult (DeLuca et al. 1979). For the same reasons, the use of secreted luciferase as a non-destructive reporter is complicated. Luciferase also offers very little potential for meaningful histochemical analysis, or fusion genetics. In plants, luciferase cannot match GUS in terms of ease, convenience, cost, and flexibility.

1.2.3 GFP

Green fluorescent proteins (GFPs) have been isolated from many organisms, including jellyfish (*Aequorea victoria*), and sea pansy (*Renilla reniformis*). GFP from jellyfish is one of the most commonly used reporters in plants. The ability to fluoresce without the need of exogenously added substrates makes GFP a true *in vivo* reporter. Coupled with its N- and C-termini fusion tolerance (Kain et al. 1995), GFP has been an

invaluable tool in monitoring trafficking and subcellular localization of proteins, offering real-time visualization of dynamics *in situ*, avoiding potential fixation and staining artifacts. For some recent reviews of the GFP system, see Gerdes & Kaether (1996); Leffel et al. (1997); Satiat-Jeunemaitre et al. (1999); Margolin (2000); Phillips (2001); van Roessel & Brand (2002).

The increased popularity of fluorescent proteins as reporters has been greatly facilitated by genetic engineering of GFP. This has led to improved versions, like mGFP5-ER – a modified GFP with a cryptic plant intron in the coding sequence removed, and is targeted to the ER (Haseloff et al. 1997); and brighter, more soluble, red- and blue-shifted GFP variants – EGFP, YFP, CFP, to name just a few (Reichel et al. 1996; Davis & Vierstra 1998).

New fluorescent proteins from other sources have also been discovered, for example the red fluorescent protein (RFP) from *Discosoma* sp. (Matz et al. 1999; Fradkov et al. 2000; Bevis & Glick 2002), which has been shown to work well in plants (Jach et al. 2001; Dietrich & Maiss 2002).

The color variants of fluorescent proteins allow dual-reporter experiments, and when combined with fluorescence resonance energy transfer (FRET) technique, allow identifying complex protein interactions in their natural environment within the living cell (Ellenberg et al. 1999; Haseloff 1999; Day et al. 2001; Hanson & Kohler 2001; Falk & Lauf 2001). Besides, pH-sensitive GFP variants that alter their fluorescent properties in acidic environments have also been developed and used extensively as pH indicators of various organelles (Kneen et al. 1998; Llopis et al. 1998; Miesenbock et al. 1998; Allen et al. 1999; Nakanishi et al. 2001).

GFP system, however, offers lower sensitivity compared to other systems, due to its intrinsic limit of one-photon-of-light-per-one-GFP-molecule (no enzymatic signal amplification). In addition, the system requires relatively expensive detection facilities. This restricts the use of GFP to a relatively well-equipped laboratory environment, and

hinders its use under field or “low-tech” conditions. Besides, the assay is not totally non-interfering. UV-induced toxicity and photo-bleaching is common, limiting long observation/analysis. In addition, GFP aggregates are cytotoxic (Haseloff & Amos 1995; Cramer et al. 1996). Unstable expression due to high toxicity of certain GFP variants is common. GFP-induced apoptosis has also been recorded (Liu et al. 1999).

1.2.4 GUS

1.2.4.1 Biochemistry, natural occurrence and biology of various β -glucuronidases

β -glucuronidase (E.C. 3.2.2.31; GUS) is a hydrolase that catalyzes the cleavage of a wide variety of β -glucuronides. It also cleaves some β -galacturonides with much lower efficiency. The enzyme from most sources, including *E. coli*, is a homotetramer with a monomer molecular weight of about 68-70 kDa. Almost any aglycone conjugated in a hemiacetal linkage to the C1 hydroxyl of a free D-glucuronic acid in the β configuration serves as a GUS substrate. As an exo-hydrolase, GUS does not cleave glucuronides in internal positions within polymers. In addition, it does not cleave β -glucosides, β -galactosides, β -manosides, or glycosides in the α -configuration. Thus GUS is apparently inactive against uronic acids in plant cell walls and mucilages, as these are generally in the α -configuration, and situated in internal position of polymers (Jefferson et al. 1986).

β -glucuronidases are present in all mammals and some microbes. To date, dozens of genes coding for β -glucuronidases have been identified from mammals: rat (Nishimura et al. 1986), human (Oshima et al. 1987), mouse (Gallagher et al. 1988), dog (Ray et al. 1998), cat (Fyfe et al. 1999), horse (Caetano et al. 1999)...

A number of microbial GUS genes have also been identified from, for example, *Thermotoga maritima* (Nelson et al. 1999), *Lactobacillus gasseri* (Russell & Klaenhammer 2001), *Clostridium perfringens* (Shimizu et al. 2002). Many other GUS

genes of microbial and fungal origin have been isolated in CAMBIA (R. Jefferson et al., unpublished).

Most plants, algae, bacteria and fungi do not have GUS activity. These include many bacteria of economic and agricultural importance: *Rhizobium*, *Bradyrhizobium*, *Agrobacterium*, *Pseudomonas* and *Azospirillum*; many fungi such as *Saccharomyces*, *Schizosaccharomyces*, *Aspergillus*, *Neurospora*, and *Ustilago* (Wilson et al. 1995).

GUS from *E. coli* is a natural cytoplasmic enzyme. In its native environment, it is not subjected to any post-translational modification. However, it has one potential N-glycosylation site, and nine cysteine residues. Therefore, modifications like glycosylation and disulfide bond formation are possible when the protein is expressed in a heterologous environment that allows these modifications to occur. More details on these potential modifications, and their implications on the enzyme's activity and transport, are presented in section 1.5.

E. coli has evolved to live in the vertebrate's large intestine, an environment very rich in glucuronidated compounds. These compounds are the results of glucuronidation, a principal xenobiotic detoxification and excretion pathway of all vertebrates. The *gus* operon, responsible for the ability of *E. coli* to metabolize glucuronides, has been characterized. Besides *gusA*, it includes *gusB*, a glucuronide-specific H⁺ symporting permease, and *gusC*, whose function is not yet known but speculated to be an outer membrane porin facilitating access of glucuronides to the periplasmic spaces. Transcription of the operon is repressed by the product of the *gusR*, which lies immediately upstream of the *gus* operon. This *gusR* gene is specifically induced by a wide variety of glucuronides (R. Jefferson et al., unpublished).

Mammalian GUSes are lysosomal acid hydrolases. They have N-terminal signal peptides that target them into the ER, as well as C-terminal lysosomal targeting propeptides. In the ER, they are retained by specific interaction with egasyn, a mammalian-specific ER-resident protein (Zhen et al. 1995; Islam et al. 1999). They are

glycosylated at various N-glycosylation sites, and subsequently transported to the lysosome via the man6P mechanism (Shipley et al. 1993).

Mammalian GUSes are responsible for the lysosomal degradation of glucuronic acid-containing glycoaminoglycans. Deficiency of the enzyme in humans results in the genetic disorder mucopolysaccharidosis type VII, or Sly syndrome, which is characterized by the lysosomal accumulation of glycosaminoglycans in all tissues, and affects a wide range of organs, including the central nervous system (Sly et al. 1973).

In terms of enzymatic characteristics, mammalian GUSes have acidic pH optimum and narrow pH response curve. They are also generally less robust compared to the *E. coli* enzyme. Their mammalian origin would also make them less suitable for use as reporters, due to public concerns.

1.2.4.2 GUS as a plant reporter system

GUS from *E. coli* (referred herein as GUS^{E_{co}}) has clearly been the most widely used reporter enzyme in plant molecular biology since its introduction to the scientific community in 1987 (Jefferson et al. 1987; Jefferson 1987). There have been more than 5000 citations to its use in the primary literature, and over 1000 field releases of transgenic plants containing GUS worldwide.

The enzyme has many characteristics of an ideal reporter. It is remarkably stable, has no cofactors, and will tolerate many commonly used chemicals and assay conditions (temperature, pH...). It is not inhibited by the presence of substrates or end products. Its assay is simple and sensitive, with a wide range of available substrates that accommodate different assay types and formats (section 1.2.4.3). Besides, it can tolerate large N-terminal and C-terminal fusions without loss of enzyme activity, therefore, has been used in many fusion and transport studies (Kavanagh et al. 1988; Iturriaga et al. 1989; Schmitz & Lonsdale 1989).

With the above characteristics, GUS has become a very popular reporter in systems lacking intrinsic GUS activity, such as most plants, and some bacteria and fungi. In plants, there have been various reports about endogenous GUS-like activities (Plegt & Bino 1989; Hu et al. 1990; Alwen et al. 1992; Hodal et al. 1992; Thomasset et al. 1996), but very few studies on plant GUS have been carried out. Some plants in which GUS has been studied are rye (Schulz & Weissenböck 1987), tobacco (Alwen et al. 1992; Thomasset et al. 1996), and skullcap -- *Scutellaria baicalensis* (Levy 1954; Ikegami et al. 1995; Sasaki et al. 2000). Some of these are clearly substrate-specific (e.g. rye and skullcap) and are thus questionably classified as glucuronidases.

In plants, no other endogenous enzyme or substrate that could interfere with GUS assay has been recorded. In addition, GUS does not seem to interfere with plant physiology and metabolism. It is also regarded as safe for the environment and consumers (Gilissen et al. 1998).

The major drawback of the current GUS system is its destructive assay: the cell membrane must be disrupted so that the cytosol-localized enzyme can come into contact with its substrate. It has long been envisaged that a secretable GUS enzyme would eliminate the need to kill the host cell, and therefore, open the way to a non-destructive reporter system and other novel applications (Jefferson et al. 1987).

Compared to the GFP and luciferase systems, the GUS system has been subjected to relatively fewer modifications since first introduced. This is due partly to the various characteristics of the system capable of meeting most requirements for plant reporter analyses, and the comprehensive techniques published on its first release. Most of the progress has been in the development of new substrates, rather than better or new enzyme(s), or assay procedures. Various GUS substrates are covered in section 1.2.4.3. Development of better or new GUSes with better enzymatic characteristics and secretability was recognized very early on (Jefferson 1989), and has become the focus of recent work, presented in section 1.5.

1.2.4.3 GUS substrates: the companions of the GUS system

The GUS system has been greatly facilitated by the development of a wide range of substrates that are suitable for different types and formats of spectrophotometric, fluorimetric, chemiluminescent, histochemical reporter analyses.

It should be noted that glucuronidation is a common detoxification process in vertebrates. As a result, thousands of natural glucuronides (detoxified excretion forms of xenobiotic and endogenous compounds) have been described in the literature. Review of those glucuronides is found elsewhere (e.g. Dutton 1980). Here, I will describe only the common substrates used for GUS reporter assays, and some known pro-active substrates for GUS to be used as an effector. Most substrates described here are commercially available through companies like Biosynth, Molecular Probes, Sigma, Tropix, ICN Pharmaceuticals...

1.2.4.3.1 Common substrates used for GUS reporter assay

5-bromo-4-chloro-3-indoxyl- β -D-glucuronide (X-glcA) and other 3-indoxyl derivatives

X-glcA (often incorrectly abbreviated as X-gluc or X-glc and occasionally X-glcU) is the most commonly used substrate for GUS histochemical analysis. This substrate is hydrolysed in the presence of GUS to yield a colorless soluble intermediate that undergoes oxidative dimerization to form the final blue precipitate of diX-indigo (for details see Guivarc'h et al. (1996). Other 3-indoxyl derivatives (commercially known as magenta-, salmon-, lapis- β -D-glucuronides...) undergo similar reactions to give final precipitates of different colors. These precipitated crystals are stable through histochemical procedures for light microscopy. The presence of Br and Cl atoms also makes these crystals identifiable by electron microscopy (Caissard et al. 1992).

The main drawback of these substrates is the diffusion of the soluble intermediate to other places beside the site of GUS activity, leading to histochemical artifacts

(Jefferson et al. 1987; De Block & Debrouwer 1992; De Block & Van Lijsebettens 1998). The addition of oxidative catalyst, like potassium ferricyanide/ferrocyanide mixture (Lojda 1970), is often used to minimize such possible diffusion.

4-methylumbelliferyl- β -D-glucuronide (MUG) and other coumarin glucuronides

MUG is the most popular fluorogenic substrate for quantitative GUS activity assay. This substrate when cleaved by GUS releases 4-methyl umbelliferone (7-hydroxy-4-methylcoumarin) which is highly fluorescent at pH>8.0 (excitation=365nm, emission=455nm). Being a fluorogenic substrate, MUG offers very sensitive assay, with high signal-to-noise ratio, compared to colorimetric substrates like pNPG (Jefferson 1987).

Some other coumarin glucuronides even give higher extinction coefficient, better fluorescence at neutral pH, or longer emission wavelength, than MUG. For example, trifluoromethylumbelliferyl glucuronide (TFMUG), 3-cyano-7-hydroxycoumarin glucuronide, carboxymethylumbelliferyl glucuronide, 6,8-difluoro-7-hydroxy-4-methylcoumarin glucuronide (DiFMUG) (Jefferson 1987; Haugland 2001).

p-nitrophenyl- β -D-glucuronide (pNPG)

pNPG is commonly used for colorimetric assay of GUS. The cleaved product, p-nitrophenol, is bright yellow (absorbance=415nm) at neutral to alkaline pH. It offers cheap and simple assay without sophisticated instrumentation. Its limitations are the intrinsic lack of sensitivity (of methods based on absorption of light), and the problems caused by light absorption by pigments in extracts. Sensitivity can be somewhat enhanced by using long assays, made possible due to the remarkable stability of GUS (Jefferson 1987).

EFL-97- β -D-glucuronide

ELF-97-glcA is an excellent alternative to X-glcA for histochemical analysis. Upon hydrolysis, this fluorogenic substrate produces a bright yellow-green-fluorescent precipitate (excitation=365nm, emission=500-580nm) at the site of enzymatic activity.

This fluorescent precipitate has some unique spectral characteristics, including an extremely large Stokes shift, that make it easily distinguishable from the endogenous fluorescent components commonly found in plants (Haugland 2001).

This substrate has also been used for in-gel zymography to detect GUS activity (Zhou et al. 1996; Steinberg et al. 2000; Kemper et al. 2001), immunoassays on protein microarrays (Arenkov et al. 2000), and flow cytometric analysis (Matsumura et al. 2001).

Note that EFL-97 fluorescence is quenched in acidic environment, therefore, simultaneous staining with other dyes that have maximal fluorescence in acidic pH (like the popular nuclei counterstain DAPI) is less than optimal (personal observation).

Fluorescein diglucuronide and 5-dodecanoylamino fluorescein di- β -D-glucuronide (ImaGene Green)

Fluorescein diglucuronide is hydrolyzed to the monoglucuronide and then to the highly fluorescent fluorescein (Jefferson 1985; Haugland 2001). It has also been used for flow cytometric assay of individual mammalian cells (Lorincz et al. 1996).

ImaGene Green is a fluorogenic analogue of fluorescein di- β -D-glucuronide (Haugland 2001). The lipophilic substrate can diffuse across cell membrane under physiological conditions. The hydrolyzed product is retained inside the cell in which it is produced (no intercellular diffusion), where it can be visualized as a yellow-green fluorescence (Fleming et al. 1996; Guivarc'h et al. 1996; Herd et al. 1997).

Both substrates are not very stable, and cannot be used for permanent histochemical slides.

Resorufin- β -D-glucuronide

This is another fluorogenic substrate (Jefferson 1985; Haugland 2001). The cleaved product, resorufin, has high extinction coefficient and quantum efficiency, and fluoresces maximally at neutral pH (excitation=570nm, emission=590nm, a range

where plant tissues do not heavily absorb or fluoresce). Note that resorufin undergoes reduction to a non-fluorescent form in the presence of reducing agents, therefore, DTT should be omitted from the assay buffer when this substrate is used (Jefferson 1987).

Naphthol ASBI- β -D-glucuronide

Naphthol ASBI- β -D-glucuronide is an alternative colorigenic histochemical substrate. The cleaved product, free naphthol ASBI, must be coupled to a diazo dye (Fishman & Goldmans 1965). The coupling reaction is not totally specific to naphthol ASBI, therefore, suitable controls must be performed.

Phenolphthalein- β -D-glucuronide

This is an alternative substrate to pNPG for colorimetric assay. Phenolphthalein gives deep red color under alkaline conditions. This substrate has low V_{max} for GUS -- about 30 times lower than that for pNPG (Tomasic & Keglevic 1973).

1,2-dioxetane- β -D-glucuronide (Glucuron, or GUS-Light)

Glucuron is a chemiluminescent GUS substrate for quantitative assay. The assay is very sensitive, simple, rapid, and does not suffer from issues related to fluorogenic assay, like auto-fluorescence and quenching. It provides excellent alternative to the luciferase reporter system for plants (Bronstein et al. 1989; Bronstein et al. 1994; Bronstein et al. 1996; Olesen et al. 2000).

Some other substrates that are less used are: sudan II-glucuronide (Terry et al. 1993), DDAO glucuronide (Haugland 2001), 8-hydroxyquinoline- β -D-glucuronide (James & Yeoman 1988).

1.2.4.3.2 Pro-active GUS substrates for novel applications

Pro-active GUS substrates are non-bioactive glucuronides that can be cleaved by GUS to release the bioactive components (hormones, nutrients, toxins...) capable of acting upon the cell or tissue that is expressing GUS. Such pro-active glucuronides have been

used in a medical application, known as Antibody-Directed Enzyme Prodrug Therapy – ADEPT. Examples are anthracycline-based prodrugs, like epirubicin- and doxorubicin-glucuronides (Haisma et al. 1994; Houba et al. 1996; Houba et al. 2001). Details of these GUS substrates can be found in the references cited.

In plant molecular biology, pro-active glucuronides could be utilized in a range of exciting applications. For example, the development of “positive selection” schemes, where the active compounds are the sole sugar or hormone source, and only GUS-transformed cells can use this sugar or hormone source to proliferate. Another example is selective ablation methods (where the active compounds are cytotoxic), which could be used to kill pollen grains for the production of male sterile plants.

In theory, many of these pro-active substrates can be synthesized either chemically, or biosynthetically using vertebrate’s glucuronidation process. Among those, cytokinin-glucuronides and glucose-glucuronide are the most promising.

Cytokinin-glucuronides

Some glucuronide derivatives of cytokinins have been synthesized and tested for release of active cytokinins upon hydrolysis by GUS. Among successful candidates are N6-benzyladenine and isopentenyladenine coupled to glucuronic acid at the N-3 position. A positive selection scheme using these substrates and non-secreted GUS was moderately successful (Joersbo & Okkels 1996; Okkels et al. 1997). It is expected that a secreted GUS can greatly improve the scheme by improving substrate-enzyme contact.

Cellobiouronic Acid (CBA)

CBA is a glucose-conjugated glucuronide. This substrate can be obtained by acid hydrolysis of the gel-forming polysaccharide gellan gum (also known as phytigel or Gelrite). Gellan gum is a complex heteropolysaccharide with repeating units of tetrasaccharides. These tetrasaccharides are linked internally together by pyranosidic

bonds, except the uronosidic bonds between glucuronic acid and glucose (Giavasis et al. 2000). Upon hydrolysis with strong acid, most pyranosidic bonds are cleaved, while the stronger uronosidic bonds remain intact. CBA can then be purified from the hydrolysate by ion exchange chromatography (P. Wenzl and R. Jefferson, unpublished).

Unfortunately, this substrate is not efficiently cleaved by GUS^{Eco} and GUS^{Ssp} (P. Wenzl and R. Jefferson, unpublished). Recently, however, a novel microbial GUS has been shown to be able to hydrolyse CBA efficiently (J. Mayer and R. Jefferson, unpublished). It is, therefore, expected that *in vitro* evolution by DNA shuffling with all the available GUSes will result in new GUS(es) that cleaves CBA efficiently while maintains other desirable characteristics.

1.2.5 Other less popular reporter systems

LacZ: The *lacZ* gene of *E. coli*, coding for β -galactosidase, is historically the first reporter gene ever used. It has been a very useful reporter in microbial and animal systems. It offers simple assays (similar to the GUS system), good fusion tolerance, and excellent histochemical detection due to β -galactosidase's remarkable tolerance to glutaraldehyde (Teeri et al. 1989). However, in plants, the usefulness of *lacZ* fusions (Helmer et al. 1984; Matsumoto et al. 1988) is limited mainly by high endogenous β -galactosidase activity in most plants and most tissues (Jefferson 1987).

NOS and OCS: The two genes *nos* and *ocs*, coding for nopaline synthase and octopine synthase, respectively, are Ti-plasmid encoded genes of *Agrobacterium tumefaciens* (Depicker et al. 1982; De Greve et al. 1982; Bevan et al. 1983). Their products, opines, are not normally found in plant cells. However, these genes are no longer used as reporters, mainly because their assays are cumbersome and difficult to quantify (Otten & Schilperoort 1978).

InaZ: The ice nucleation gene *inaZ* of the plant pathogen *Pseudomonas syringae* has been used as reporter in bacteria (Lindgren et al. 1989; Arvanitis et al. 1995), and some plants, but its use in plants is very limited, mainly due to its low temperature assay condition (van Zee et al. 1996).

A few other genes have been occasionally used as reporter in plants, for example the *xylE* gene of *Pseudomonas putida*, coding for catechol dioxygenase (Buell & Anderson 1993), or the *pat/bar* genes of *Streptomyces*, coding for phosphinothricin acetyltransferase (PAT) (Botterman et al. 1991).

1.3 Limitations and common pitfalls in the use of reporter genes

Although common and powerful for gene expression studies, use of a reporter gene has its limitations and pitfalls, which need to be considered so that experiments are designed and results are interpreted correctly. First, and perhaps obviously, the use of reporter genes is only possible in systems where transformation or gene transfer of some sort is available. Secondly, very often in many “gene expression” studies, the simplistic view of directly relating *reporter activity* to *gene expression* is presented. While this might be arguably regarded as “practically adequate” in some studies, it is certainly very superficial, because it omits the whole cascade of regulation events that leads to the fine tuning of gene expression in living organisms: the initiation of transcription and translation; the processing, transport, and degradation of mRNA or protein. This simplistic view, therefore, can often lead to the omission and/or incorrect interpretation of biologically meaningful results.

A reporter assay, in any format and at its best, only measures the end result: the steady-state level of the reporter protein being examined, and in the case of enzymes, the steady state level of *enzymatically active* reporter protein. It does not directly reveal previous, or even concurrent regulation events. Studies on regulation of gene

expression, therefore, require a variety of additional tools/methodologies in combination with the use of reporter gene. For example, southern blot and PCR to confirm the presence and correctness of the DNA sequence, northern blot or RT-PCR to confirm the level of steady-state level of mRNA, western blot and other protein modification detection methods to confirm the level and possible processing of the protein, or immunohistochemistry to verify the presence of reporter product against a variety of artifacts/background, etc. Emerging powerful tools such as DNA and protein microarrays, in some cases, will be complementary to reporter analyses.

Many limitations and pitfalls of reporter genes are intrinsic to the reporters' properties. For example, the relatively small size of GFP could lead to possible unwanted cell-to-cell or long distance trafficking (Imlau et al. 1999), the remarkable stability of GUS could make it unsuitable for the study of dynamics of gene transcription (De Block & Debrouwer 1992). Other limitations and pitfalls are intrinsic to various assay methods, for example, possible diffusion of signal when using the GUS substrate X-glcA, or strong autofluorescence in certain tissues when using GFP, etc. Readers are referred to the substantial literature for more examples of pitfalls and artifacts in reporter assay analyses.

1.4 Protein secretion in eukaryotes

Protein secretion is a broad topic of which there are many good reviews, including several chapters in classic molecular biology textbooks (Alberts et al. 1994; Lodish et al. 2000; Raikhel & Chrispeels 2000). Here, the topic is very briefly reviewed, with some emphasis on yeast and plants, and on aspects relevant to the GUS secretion engineering work.

Newly synthesized proteins can be delivered to the exterior of eukaryotic cell via the so-called *secretory pathway*, which can be dissected into distinct steps described hereinbelow.

1.4.1 Targeting to and translocation across the ER membrane

Secretory proteins (plus ER-, Golgi-resident proteins, membrane proteins, and other proteins destined for lysosome, endosome, vacuole) carry in themselves a *signal peptide*, normally found in the N-termini of the proteins (Blobel & Dobberstein 1975a; Blobel & Dobberstein 1975b; Blobel et al. 1979). This peptide serves as a molecular “postcode”, which tells the cellular machinery to target the protein to the ER, the first compartment in the secretory pathway.

Targeting and translocation can occur via co- or post-translational mechanisms. Most proteins are targeted and translocated cotranslationally (i.e. during translation). When the signal peptide emerges from the ribosome, a *signal recognition particle* (SRP) (Siegel & Walter 1988) binds to this peptide and the ribosome, pausing translation. This ribosome-SRP complex then binds to the *SRP receptor* in the ER membrane. The SRP is subsequently released, leaving the ribosome attached to the *translocon* (a protein-conducting channel also known in yeast and mammals as the Sec61p-complex) in the ER membrane. Translation resumes, and as the nascent polypeptide extends, the translocon opens to form an aqueous pore, allowing the protein to enter the ER lumen.

For good reviews on cotranslational translocation, see Crowley et al. (1993); Andrews & Johnson (1996); Rapoport et al. (1996); Wilkinson et al. (1997); Menetret et al. (2000). The signal peptide is removed from most soluble proteins by *signal peptidases* (Dalbey & Von Heijne 1992), but remains uncleaved for most membrane proteins (Wilkinson et al. 1997).

Translocation and membrane integration of transmembrane proteins is more complex, and requires a combination of start- and stop-transfer signal peptides. This combination determines the topology of the integration (Hartmann et al. 1989; High & Dobberstein 1992).

Post-translational ER targeting and translocation has recently been elucidated in yeast. In this case, the protein is fully synthesized in the cytosol, preserved in a translocation competent conformation by cytosolic chaperones (folding-assisting proteins), and translocated via the translocon. The question of how the protein is initially targeted to the ER remains largely unanswered. Interactions at the ER membrane are also not completely understood, although evidence suggested that the protein interacts with the Sec-61p complex (the translocon) and Sec62p/Sec63p complex (Wilkinson et al. 1997; Plath et al. 1998; Rapoport et al. 1999; Plath & Rapoport 2000).

1.4.2 In the lumen of the ER

In the ER, protein folding, assembly and modification occur. The ER maintains a high concentration of chaperones and an environment optimal for protein folding and assembly (Hammond & Helenius 1995). Misfolded and assembly-defective proteins are not able to continue their passage through the secretory pathway, and are instead targeted for proteolytic degradation before they reach their usual destination. Most transport-incompetent intermediates are retained in the ER, and then retro-translocated (via the translocon) back to the cytosol to be degraded by cytosolic ubiquitin/proteasome pathway (Pilon et al. 1997; Lord et al. 2000; Mancini et al. 2000; Vashist et al. 2001). Those that escape the ER are diverted to lysosomes/lytic vacuoles

for degradation (Chang & Fink 1995; Hong et al. 1996). Degradation in the ER was previously proposed (Klausner & Sitia 1990; Bonifacino & Lippincott-Schwartz 1991), but proteolytic enzymes were not found in the ER (Lodish et al. 2000). This sorting mechanism, ensuring that only correctly folded, assembled and modified proteins are transported along the secretory pathway, is known as ER quality control (Hammond & Helenius 1995; Kopito 1997; Ellgaard et al. 1999; Ellgaard & Helenius 2001).

Post-translational modifications of proteins in the ER include asparagine-linked (N-linked) glycosylation and the formation/rearrangement of disulfide bonds. Most proteins made in the ER are glycosylated by the addition of a common 14-unit oligosaccharide ($\text{Glc}_3\text{Man}_9\text{GlcNAc}_2$) to an asparagine residue within the motif Asn-Xaa-Ser/Thr where Xaa is any amino acid except proline (Abeijon & Hirschberg 1992; Hart 1992). The oligosaccharide is added to the target asparagine as the nascent polypeptide is being translated and translocated at the ribosome-translocon complex. It is subsequently trimmed in the ER, and mostly in the Golgi apparatus, resulting in the diversified N-linked oligosaccharide structures found on mature glycoproteins. It should be noted that not all Asn-Xaa-Ser/Thr configurations in a given protein are necessarily glycosylated.

N-linked oligosaccharide is an important signal for various quality control and targeting mechanisms. For example, it allows binding to lectins (proteins that bind tightly to specific sugars) like calnexin and calreticulin, therefore, prevents exit of immature glycoproteins, promotes correct folding, inhibits aggregation, blocks premature oligomerization, and regulates ER associated degradation (Helenius 1994; Ellgaard et al. 1999; Parodi 2000). Another example is the Man6P lysosomal targeting motif, where specific mannose residues on the oligosaccharide chain get phosphorylated in the *cis* Golgi network, and become markers for the Man6P-receptor in the Golgi. Receptor-bound enzymes are therefore segregated from secretory enzymes, and translocated to lysosomes. This is a common targeting motif for many mammalian

lysosomal acid hydrolases, including mammalian β -glucuronidase (von Figura & Hasilik 1986; Gonzalez-Noriega et al. 2001).

Another posttranslational modification of proteins in the ER is the oxidation of freely-exposed sulfhydryl (SH) groups of cysteines to form disulfide (S-S) bonds. This reaction is favorable only in the ER (and not the cytosol) due to its oxidative redox environment. Thus disulfide bonds are common in secretory proteins and exoplasmic domains of membrane proteins, but are absent from soluble cytosolic proteins. Formation and rearrangement of disulfide bonds is the basis for the so-called thiol-mediated retention mechanism. This mechanism is well-described in the retention of unassembled immunoglobulin (Ig) chains, where they form intermolecular disulfide bonds with ER-resident thiol reductases such as PDI and ERp72 (Sitia et al. 1990; Fra et al. 1993; Reddy et al. 1996; Isidoro et al. 1996; Reddy & Corley 1998; Ellgaard et al. 1999). Thiol-mediated retention has also been shown for unassembled subunits of acetylcholinesterase (Kerem et al. 1993). Aspects of glycosylation and disulfide bond formation in relation to the secretion engineering of GUS are presented in section 1.5.

The ER has its own set of residential proteins. Most of them are molecular chaperones and other folding-assisting enzymes, or enzymes involved in N-linked oligosaccharide trimming. For example, **BiP** (binding protein), also known as GRP87 (glucose regulated protein), is an ER-specific hsp70-class heat shock protein that binds to hydrophobic patches on the protein surface (Munro & Pelham 1986; Gething 1999). **GRP94** (also referred to as 100K, hsp108, endoplasmin, or ERp99) is an hsp90-class heat shock protein that binds to oxidized intermediates (Munro & Pelham 1987; Melnick et al. 1992; Melnick et al. 1994). **Calnexin** and **calreticulin** are lectin chaperones that assist folding, assembly and quality control of glycoproteins (Rodan et al. 1996; Trombetta & Helenius 1998; High et al. 2000). **PDI** (protein disulfide isomerase), **ERp57**, and **ERp72**, are all multifunctional proteins which act as chaperones, as well as thiol oxidoreductases, catalyzing the formation and rearrangement of disulfide bonds (Wang

& Tsou 1993; Puig et al. 1994; Sun et al. 2000). **PPI** (Peptidyl-prolyl isomerase), also referred to as cyclophilin, is a class of enzymes that accelerates the rotation about peptidyl-prolyl bonds in unfolded segments of a polypeptide (Stamnes & Zuker 1990; Price et al. 1991; Colley et al. 1991; Frigerio & Pelham 1993). **Glucosidase I, glucosidase II, manosidase I, and manosidase II** are enzymes involved in N-linked oligosaccharide trimming (Herscovics 1999).

The ER uses retention and retrieval as mechanisms to maintain its own set of residential proteins, while allowing a constant flow of transported proteins *en route* to other compartments. Most soluble ER-resident proteins mentioned above, and type II membrane proteins (proteins with N-termini in the cytoplasm) have a C-terminal KDEL/HDEL sequence which retains them in the ER, or retrieves them from the cis-Golgi network by the KDEL receptor, ERD2p – an integral membrane protein found in *cis* Golgi and ER-Golgi vesicles (Munro & Pelham 1987; Pelham 1989; Semenza et al. 1990). Similarly, type I membrane proteins (proteins with C-termini in the cytoplasm) have a C-terminal KKXX sequence which helps retrieve them from post-ER compartments by COPI-mediated retrieval (Jackson et al. 1990; Pelham 1995; Teasdale & Jackson 1996).

With various quality control mechanisms in the ER, it is by no mean self-evident that hererologous proteins will be secreted efficiently when engineered. The proteins could be abnormally folded/assembled, and retained in the ER, either because they form aggregates or because they are permanently bound to ER-retained chaperones. The addition of of N-linked oligosaccharides could compromise protein folding or activity, mediate degradation, or mistarget the protein to other compartments like lysosomes or vacuoles. Formation of inappropriate inter- or intra-molecular disulfide bonds could interfere with proper folding, and/or mediate thiol retention. Besides those “primary” quality control mechanisms, which apply to all proteins, various “secondary” quality control mechanisms, which are protein-specific, also exist (for details see Ellgaard et

al. (1999). Inappropriate interactions via these “secondary” mechanisms would also lead to retention/retrieval/degradation of heterologous proteins.

1.4.3 Transport from ER to Golgi

The passage from the ER via the Golgi apparatus to the cell surface is often referred to as the *default pathway*. Any protein that enters the ER (and folds and assembles properly) will automatically be transported through the Golgi apparatus to the cell surface if it does not carry any additional targeting signal (Alberts et al. 1994).

Proteins leave the ER as ‘cargo’ captured inside COPII (coat protein) vesicles (Barlowe et al. 1994). These vesicles bud from the ER ‘exit sites’, and fuse to the ER-Golgi intermediate compartment (ERGIC, also called the vesiculotubular clusters-VTCs) (Hauri & Schweizer 1992). From there, the proteins are transported to the Golgi. Proteins with ER retention signals like KDEL or KKXX are retro-transported, from ERGIC or Golgi, back to the ER via COPI vesicles (Griffiths et al. 1994). For recent reviews on biogenesis and components of these COP-coated vesicles, see Schekman & Orci (1996); Kirchhausen (2000); Wakeham et al. (2000); Gorelick & Shugrue (2001).

Transport from ER to Golgi, and all later steps on the secretory pathway (intra-Golgi, Golgi-cell surface) are all vesicle-mediated, which involve vesicle’s budding off donor membrane, and fusing to acceptor membrane. If the ER and Golgi are to maintain their membrane and constituents integrity, the anterograde (forward) flux of vesicles must be matched by a retrograde (backward) flux of equal magnitude. Vesicle transport from ER to Golgi (COPII/COPI-mediated), and within Golgi (COPI-mediated – see later in 1.4.4) is, therefore, bi-directional. In contrast, vesicle transport to final destinations like the plasma membrane or lysosome/vacuole is uni-directional. The implied *de novo* content generation (vacuole), and recycling of content (plasma membrane) by various means, is not discussed here.

It is obvious that vesicles must dock and fuse with the correct acceptor membrane. This requires “compatible” signals present on the target membrane (t-SNARE), and on the vesicles (i.e. donor membrane) themselves (v-SNARE). Both SNAREs, meaning SNAP receptor because they both bind to another protein called SNAP, are integral membrane proteins (Sollner et al. 1993a; Sollner et al. 1993b). Different isoforms of v-SNAREs and t-SNAREs are present in various cell membranes and vesicles, where they may provide specificity to the docking event. Examples are SNAREs for ER-Golgi transport (Hay et al. 1997; McNew et al. 1997; Paek et al. 1997), for later transport within the Golgi (Nagahama et al. 1996; Lowe et al. 1997; McNew et al. 1998), for vacuolar transport (Nichols et al. 1997; Boeddinghaus et al. 2002), or for plasma membrane/secretion (Timmers et al. 1996; Bajjalieh 2001).

1.4.4 Transport within the Golgi apparatus

The Golgi apparatus consists of an ordered series of compartments: *cis* Golgi network, Golgi stack (includes *cis*, *medial* and *trans* cisternae), and *trans* Golgi network (Alberts et al. 1994; Lodish et al. 2000). As the proteins move along these compartments in the *cis-to-trans* direction, the N-linked oligosaccharide chains are processed in multiple steps: trimming off existing sugars (resulting in high-mannose glycans), and with or without adding new sugars like galactose, sialic acid, fucose, xylose (resulting in complex glycans). Exact composition and configuration of these sugars on mammalian glycans are different from those of plants. For example, xylose is only specific to plant glycans (likely explanation for why plant glycoproteins are highly immunogenic to mammals). The question of why some glycans are modified and other are not, remains unclear (Raikhel & Chrispeels 2000).

In mammals, each glycan processing step happens only in specific sub-compartment(s) of the Golgi. For example, removal of mannoses happens in *cis* and *medial* cisterna, addition of N-acetylglucosamines in *media* cisterna, addition of galactoses in *trans* cisterna, and addition of sialic acids in the *trans* Golgi network.

Because these steps are highly ordered, the presence of certain carbohydrate residues on proteins can provide useful markers for following their movement from the ER and through the Golgi cisternae (Alberts et al. 1994; Lodish et al. 2000). However, in plants, such distinct modifications in specific Golgi compartments by specifically localized enzymes remains to be demonstrated (Raikhel & Chrispeels 2000).

Also happens in the Golgi is oxygen-linked (O-linked) glycosylation, where oligosaccharides are linked to the hydroxyl group on the side chain of serine, threonine, or hydroxylysine, hydroxyproline. O-linked sugars are added one at a time, and the chain are generally short, often containing only one to four sugar residues (Lodish et al. 2000; Kieliszewski 2001). Examples of well-known O-linked glycoproteins are collagen, and glycophorin. In plants, hydroxyproline-rich glycoproteins (ubiquitous architectural components of the growing plant cell wall), such as extensin and gum arabic glycoproteins, are heavily O-glycosylated (Kieliszewski & Lamport 1994; Goodrum et al. 2000). The very diverse roles of N- and O-linked glycans are not discussed.

As mentioned earlier, transport within the Golgi is bi-directional, mediated by COPI-coated vesicles shuttling both ways (Orci et al. 1986; Orci et al. 1997; Orci et al. 2000b). Golgi resident proteins, mostly carbohydrate modifying enzymes: glycosidases and glycosyltransferases, are maintained by retrograde transport (Munro 1998; Opat et al. 2001). Recently, however, the role of anterograde vesicles in cargo transport has been under debate, when another model called “cisternal progression/maturation” has become prominent. According to this model, anterograde transport results from movement of the cisternae themselves; transport vesicles would have no role in anterograde transport, but rather would function exclusively in retrograde transport (Glick et al. 1997; Bonfanti et al. 1998; Allan & Balch 1999). Evidence supporting the “vesicular transport” model (Volchuk et al. 2000; Orci et al. 2000a), and the “cisternal progression” model (Martinez-Menarguez et al. 2001; Mironov et al. 2001), is still emerging, and does not speak against non-exclusive operation of both models. It

should be noted that Golgi biogenesis, i.e. whether it is a self-sustained stable organelle, or a dynamic, steady-state structure maintained by ER export, has also been the subject of recent debates (Seemann et al. 2000; Ward et al. 2001; Rossanese & Glick 2001; Seemann et al. 2002).

The *Trans* Golgi Network (TGN) is a sorting compartment where secretory proteins are separated from those destined for lysosome/vacuole (Orci et al. 1987; Gu et al. 2001). By default, proteins with no additional sorting signals are secreted (section 1.4.5).

The passage to lysosome/vacuole is not strictly viewed as part of the secretory pathway. However, understanding this passage, and its targeting signals is crucial and integral to the understanding of the secretory pathway. Therefore, a very brief review of the lysosomal/vacuolar targeting is presented below.

Lysosomal/vacuolar proteins have their vacuolar sorting signals (VSSs), found at the N-terminus (right after the signal peptide), or at the C-terminus. These signals are often proteolytically removed from mature proteins during or after transport to the vacuole. They are referred to as NTPP or CTPP, for N- or C-terminal propeptide, respectively. In general, NTPP and CTPP are necessary and sufficient for vacuolar targeting (Raikhel & Chrispeels 2000). Other VSSs are less obvious signals, often found as “surface loops” in the middle of the protein sequence, and are not normally removed from mature proteins. Their roles in vacuolar targeting tend to be protein-specific (Tague et al. 1990; Saalbach et al. 1991).

In many mammalian lysozymes, the presence of propeptides allows binding of GlcNAc phosphotransferase, and subsequent phosphorylation of a mannose residue on the N-linked glycan. The resulting Man6P is recognized by the lysosomal targeting Man6P receptor in the Golgi. Mammalian β -glucuronidase and many other acid hydrolases are sorted into lysosomes via this mechanism (von Figura & Hasilik 1986; Kornfeld 1987; Roberts et al. 1998). Other lysosomal proteins, like cathepsin D, are transported via

Man6P-independent mechanisms, which involve “surface loops” signals, or some other complex protein-specific interactions (Glickman & Kornfeld 1993; Zhu & Conner 1994).

Transport to yeast vacuoles has been extensively studied, resulting in a very large number of VPS (vacuolar protein sorting) proteins identified (Stack et al. 1995; Cowles et al. 1997). Unlike mammals, yeast and plants do not have the N-glycan dependent Man6P receptor. Instead, they have receptors that bind directly to amino acid sequence determinants within the ligands. Two vacuolar transport pathways have been well characterized in yeast. The 'CPY pathway' mediates transport of the soluble hydrolase CPY (carboxypeptidase Y) by means of its transmembrane sorting receptor, VPS10 (Marcusson et al. 1994; Cooper & Stevens 1996; Jorgensen et al. 1999). The 'ALP pathway' mediates transport of some membrane proteins, like ALP (alkaline phosphatase) and Vamp3 (a vacuolar t-SNARE), which contain an N-terminal vacuolar sorting signal (Cowles et al. 1997; Piper et al. 1997; Conibear & Stevens 1998).

Plant cells have two functionally and structurally distinct types of vacuoles: protein storage vacuoles and lytic vacuoles. Thus, individual soluble and membrane proteins must be recognized and sorted into one or the other type of vacuole by distinct, specific mechanisms (Neuhaus & Rogers 1998).

Transport to plant lytic vacuoles is, in many aspects, similar to that found in yeast vacuoles and mammalian lysosome. From the TGN, proteins are first delivered to the prevacuolar compartment (PVC), or endosome, by clathrin-coated vesicles, before they are delivered to lysosome/vacuole (Raikhel & Chrispeels 2000). For review about clathrin coated vesicles, see Schmid (1997). Note that the clathrin coat disassembles quickly after the vesicles form, unlike coatomer coats (COPI/COPII), which are removed after the vesicles dock on their target membrane (Alberts et al. 1994).

In plants, the N-terminal vacuolar targeting 'NPIR' motif and its receptor, the type I membrane protein BP-80, have been well-characterized (Kirsch et al. 1994; Kirsch et al. 1996; Paris et al. 1997; Matsuoka & Nakamura 1999; Cao et al. 2000). Other C-

terminal signals have also been shown to be necessary and sufficient for lytic vacuolar targeting (Bednarek & Raikhel 1991; Neuhaus et al. 1991).

Transport to storage vacuoles is specific only to plants, and does not involve BP-80 (Hinz et al. 1999). The transport is mediated by electron-dense protein-filled vesicles, which do not have clathrin coat, and bud off from TGN or direct from ER, bypassing the Golgi (Raikhel & Chrispeels 2000).

1.4.5 Transport from *Trans* Golgi Network to the cell surface

Secretory proteins are transported to the cell surface via vesicles, which bud off the TGN and fuse to the plasma membrane, releasing their content to the extracellular space. These vesicles clearly have a protein coat, whose composition has not been determined in any system (Traub & Kornfeld 1997).

For most cells and proteins, vesicle fusion is by default and continuous. In certain cells (e.g. pancreatic acinar cells, hormone-secreting endocrine cells, or nerve cells), secretion of a specific set of proteins is not continuous. These proteins (e.g. digestive enzymes, hormones, or neurotransmitters) are concentrated and sorted into secretory vesicles that are stored inside the cell awaiting a stimulus for exocytosis (Alberts et al. 1994; Lodish et al. 2000).

Various SNARE proteins have been identified for plasma membrane fusion (Timmers et al. 1996; Bajjalieh 2001). However, the precise mechanism for constitutive and regulated membrane fusion is still unclear (Raikhel & Chrispeels 2000).

Proteolytic processing (for example: proinsulin or proalbumin) in transport of secretory vesicles is common to yield mature, active proteins (Alberts et al. 1994; Lodish et al. 2000). In plants, the processing and maturation of proteins within the vesicles is largely unknown, partly due to the difficulties associated with isolating and indentifying these vesicles (Hohl et al. 1996). Many plant substilases (a super family of serine proteases

that predominantly cleave dibasic residues) have been identified (Siezen & Leunissen 1997), but their exact roles in post-Golgi proteolytic processing remain to be elucidated.

1.5 Previous work on the development of better or secreted GUS

As mentioned earlier, *E. coli* GUS has many characteristics of a nearly ideal reporter, capable of meeting most requirements in plant reporter analyses. Therefore, no system based on other GUS enzymes from various organisms has been developed. Published work done to improve *E. coli* GUS has also been modest. Among these are various mutageneses to improve the enzyme's tolerance to fixatives (Matsumura et al. 1999), or thermal stability (Flores & Ellington 2002). To eliminate background bacterial GUS expression during the plant transformation process using *Agrobacterium*, an intron was introduced to the original *gusA^{Eco}* (Vancanneyt et al. 1990). Use of an intron-containing GUS has also been known to increase the level of GUS expression in monocots, through increases in steady-state mRNA levels (Tanaka et al. 1990). Fusions of GUS and GFP have also been developed (Quaedvlieg et al. 1998), although they have been not very popular, mainly due to the compromised activities of both proteins in the fusion.

The "Achilles' heel" of the current GUS system is its destructive assay. This limitation can be overcome by either delivering substrate into the cell using the permease gene *gusB*, or engineering GUS to be secreted to the extracellular space (Jefferson 1989). So far, the former approach has not been extensively pursued, and is not discussed here.

A secreted GUS has the potential to offer combined advantages of the current GUS and GFP reporter system. More interestingly, it would also lead the way to many novel applications when combined with pro-active substrates.

Attempts have been made to obtain a secreted GUS by engineering GUS^{Eco}. Using plant signal peptides, GUS^{Eco} has been efficiently targeted to the endoplasmic reticulum, or ER (the first destination in the default secretory pathway). In its native form the enzyme is inactive in the ER due to glycosylation at a cryptic N-glycosylation site. When this site is mutated, the enzyme is active in the ER, but is not secreted (Iturriaga et al. 1989; Denecke et al. 1990; Farrell & Beachy 1990; Diaz et al. 1992; Pang et al. 1992; Firek et al. 1994; Yan et al. 1997). Our current hypothesis is that the enzyme, which has nine cysteine residues, could be retained in the ER via thiol-mediated retention mechanism. This mechanism has been shown to be responsible for the retention of unassembled immunoglobulin (Ig) chains, and unassembled subunits of acetylcholinesterase (Sitia et al. 1990; Kerem et al. 1993; Fra et al. 1993; Reddy et al. 1996; Isidoro et al. 1996; Reddy & Corley 1998; Ellgaard et al. 1999). GUS^{Eco} mutants, with various numbers and combinations of cysteines eliminated, were developed. But due to the large number of mutants, and their often-compromised activities, testing of these materials has been a low priority (R. Jefferson et al., unpublished).

An alternative approach to engineering GUS^{Eco} is to isolate new GUSes from microbial sources. We choose microbes because we can quickly screen large number of candidates, and because of the likelihood of finding enzymes that are very robust (to cope with the diverse and changing environments in which they operate). Besides, use of GUSes of microbial origin in transgenic plants would possibly be less subject to public concerns than those of vertebrate origin.

With this approach, a novel GUS with improved enzymatic characteristics was isolated from a soil bacterium, identified as *Staphylococcus* sp. by 16S rDNA sequence analysis (J. Mayer, unpublished). The new enzyme (referred herein as GUS^{Ssp}), shares about 47% identity with that of *E. coli*, but has only one cysteine in contrast to nine cysteines from its *E. coli* counterpart. It is therefore an excellent candidate for a better

and secretable GUS. However, the native gene is very AT-rich, and consequently has low levels of expression in heterologous systems. A synthetic version was designed for optimized expression in *E. coli* and plants.

1.6 Aims of the project

My research is part of the continuing work aimed at developing novel GUS enzymes that can be used as improved reporters, or as *effectors* when employed with pro-bioactive GUS substrates. This PhD project is focused on the novel GUS^{Ssp} described above. In particular, it aims to achieve the following objectives:

1. To construct a synthetic, codon-optimized *gusA*^{Ssp} gene

The poor expression of the native *gusA*^{Ssp} in heterologous systems prevents it to be used effectively as a reporter gene. Therefore, a codon-optimized version of the gene needs to be constructed. Successful construction of the gene is obviously a prerequisite for further studies described in this thesis and beyond.

2. To purify GUS^{Eco} and GUS^{Ssp} for their biochemical characterization and antibody production

Both enzymes are to be purified to high homogeneity, so that comparison of their biochemical characteristics can be made. Purified enzymes are also used to produce polyclonal antibodies. These antibodies are necessary for biochemical studies to identify possible post-translational modifications of the enzymes when they are used in targeting and transport studies. In particular, these antibodies are essential to biochemical methods for secretion studies of GUS^{Eco} and GUS^{Ssp}.

3. To study GUS^{Eco} and GUS^{Ssp} expression and secretion in yeast and plants

Yeast is used mainly because it is an excellent model for secretion studies in eukaryotes. Besides, examination of the potential use of GUS^{Ssp} as an improved GUS reporter for yeast is also important. Expression, secretion and potential applications of GUS^{Ssp} in plants are the main foci of the project. The enzyme, if secreted, will be the desirable form of GUS sought after during the last 15 years or so, and will also provide important insights to the unexplained secretory-incompetence of GUS^{Eco}.

4. To investigate some applications using GUS^{Ssp}

Conditional to the attainment of previous goals, some key experiments are performed to confirm and demonstrate various applications using GUS^{Ssp}, such as the potential use of GUS^{Ssp} as an improved and non-destructive GUS reporter, and other novel applications employing pro-bioactive GUS substrates (e.g. positive selection or selective manipulation of specific cells and tissues).

