ENGINEERING GREEN FLUORESCENT PROTEIN AS A DUAL FUNCTIONAL TAG

A THESIS SUBMITTED TO THE GRADUATE DIVISION OF THE UNIVERSITY OF HAWAI'I IN PARTIAL FULFILLMENT OF THE REQUIREMENT'S FOR THE DEGREE OF

MASTER OF SCIENCE

IN

MOLECULAR BIOSCIENCES & BIOENGINEERING

December 2002

By
Rosanto I. Paramban

Thesis Committee:

Wei-Wen Su, Chairperson
Monto Kumagai
Paul Patek

ACKNOWLEDGEMENTS

I am grateful to my advisor, Dr. Winston Su, for giving me the opportunity to work on this exciting project. I would like to thank him for his ideas and encouragement during this project and I really enjoyed working under his supervision. I would like to express my sincere thanks to Dr. Robert Bugos for helping me during the course of this project and for inspiring me to become a good scientist. I would like to thank my committee members Dr. Monto Kumagai and Dr. Paul Patek for their review and valuable suggestions for this project. I also appreciate Dr. Harry Yamamoto for letting me work in his laboratory for some of my experiments. I also enjoyed working with Dr. David Hieber, a very jovial person and thank him for his ideas. It was also great working with my colleagues Gabriel, Peizhu, Jian and Gwen who made the atmosphere in the lab a pleasant environment to work in. Finally, I would like to thank my family for their constant encouragement and for standing by all the decisions I have made.

ABSTRACT

A novel dual functional green fluorescent protein (GFP) tag useful for monitoring and purifying a recombinant protein fused to the tag has been developed. A poly-histidine (6×His) tag was inserted into a solvent exposed loop of the 11-stranded β-barrel GFP structure. Two variants (GFP172 and GFP157) were made depending on the site of insertion of the 6xHis tag. We produced the variants in *Escherischia coli* and purified them using immobilized metal affinity chromatography (IMAC). These purified GFP variants retained 65-68% of the fluorescence compared to having the tag inserted at the C-terminus. On comparison of recovery from a buffer by IMAC, it was found that greater than 80% of the protein could be recovered. Effect of temperature on the expression of these variants showed that GFP172 could be produced in soluble form at both 37°C and 28°C whereas negligible amounts of soluble GFP157 was obtained. GFP172 was fused to the C-terminus of Maltose Binding Protein (MBP) and the fusion protein was purified from *E. coli* lysate as well as from spiked tobacco leaf extracts. The primary advantage of this new GFP molecule is that it allows maximum flexibility for protein fusion since both N- and C-terminal ends are available for linking to a protein of interest. In conjunction with appropriate targeting/retention signals, this GFP tag can improve fusion protein accumulation and stability and at the same time perform the dual functions of protein monitoring and affinity purification. Preliminary studies were done to study the effect of a C-terminal HDEL sequence on the expression of granulocyte macrophage-colony stimulating factor (GM-CSF) fusion with GFP in tobacco suspension cultures. It was found that HDEL influences the stability of the fusion protein because in its absence, majority of the protein was not fluorescent.

TABLE OF CONTENTS

Acknowledgements.....iii
Abstract.....iv
List of Tables.....viii
List of Figures.....ix

Chapter 1. Introduction 1
1.1. The Green fluorescent protein.....1
1.1.1. Structural and spectral properties.....1
1.1.2. GFP variants.....3
1.1.3. Insertions within GFP.....4
1.2. Immobilized Metal affinity Chromatography (IMAC).....5
1.2.1. IMAC using 6xHis affinity tag.....6
1.3. Protein Production in Plants.....8
1.3.1. Plant cell culture.....9
1.4. Signal Peptides and Retention Signals.....10
1.6. Research Objectives.....11
References.....12

Chapter 2. Preliminary study on GFP with an internal histidine tag.....20
2.1. Introduction.....20
2.2. Materials and Methods.....22
2.2.1. Modification of mGFP5-ER.....22
2.2.2. GFP expression in *E coli*.....23
2.2.3. Preparation of crude *E coli* lysate.....24
2.2.4. Purification by IMAC.....24
2.2.5. Protein assay.....25
2.2.6. Analysis of proteins.....25
2.2.7. Fluorescence measurements.....25

2.3. Results and Discussion 25
2.4. Conclusions.....27
References.....28

Chapter3. Development and characterization of GFP variants as a dual functional tag...35
3.1. Introduction.....35
3.2. Materials and Methods.....37
3.2.1. Modification of mGFP5-ER.....37
3.2.2. MBP-GFP172 fusion protein.....39
3.2.3. Expression in *E coli* 39
3.2.4. Protein assay.....40
3.2.5. Analysis of proteins 40
3.2.6. Fluorescence measurements.....41
3.2.7. Percentage recovery of GFP by IMAC..... 41
3.2.8. pH titration..... 41
3.2.9. Tobacco extract and spiking studies.....42
3.2.10. Factor Xa Cleavage.....42
3.3. Results and Discussion 43
3.3.1. Effect of S65T.....43
3.3.2. Effect of temperature on GFP expression.....43
3.3.3. Fluorescence intensity.....44
3.3.4. Purity and Recovery.....45
3.3.5. Effect of pH.....45
3.3.6. MBP-GFP172 fusion protein.....47
3.3.7. Spiking studies.....48
3.3.8. Factor Xa cleavage.....52
3.4. Conclusions.....52
References.....53

Chapter 4. Effect of retention signals on fusion protein expression in tobacco cell suspensions.....74
4.1. Introduction.....74
4.2. Materials and Methods.....79
4.2.1. Fusion protein construct.....79
4.2.2. Tobacco transformation.....81
4.2.3. Screening of tobacco plants.....81
4.2.4. Callus initiation.....82
4.2.5. Suspension cell culture.....82
4.2.6. GFP extraction and fluorescence determination.....82
4.2.7. Western Blot analysis.....83
4.3. Results and Discussion.....84
References.....88

Chapter 5. Conclusions and recommendations.....100
5.1. Conclusions.....100
5.2. Recommendations.....101
5.2.1. GFP detection.....101
5.2.2. Fusion protein solubility.....102
5.2.3. Spiking studies.....103
5.2.4. Retention vs Secretion.....105
5.2.5. Chromatography.....106
References.....107

APPENDIX I 110

LIST OF TABLES

Table 1.0. Protease inhibitors.....104

LIST OF FIGURES

Figure 1.1 Excitation and Emission spectra of wild-type GFP 2
Figure 2.1 Primers used for vector construction.....31
Figure 2.2 GFP with internal 6xHis tag can be purified from E coli lysate.....32
Figure 2.3 Excitation spectra for GFP5 variants.....33
Figure 2.4 Intensity of GFP5 variants..... 34
Figure 3.1 Primers used for vector construction 58
Figure 3.2 The mutation S65T increases fluorescence intensity of GFP.....59
Figure 3.3 Improved solubility of GFP at lower temperature.....60
Figure 3.4 GFP intensity is higher when produced at 28°C.....61
Figure 3.5 Excitation and Emission spectra at 37°C.....62
Figure 3.6 Excitation and Emission spectra at 28°C.....63
Figure 3.7 GFP with internal His tags can be purified from *E coli* lysate.....64
Figure 3.8 The GFP variants can be recovered from immobilized metal affinity columns.....65
Figure 3.9 Fluorescence of GFP decreases with pH.....66
Figure 3.10 MBP-GFP172 purified from *E coli* lysate.....67
Figure 3.11 MBP-GFP172 fusion protein fluoresces.....68
Figure 3.12 MBP-GFP172 purified from tobacco leaf extract.....69
Figure 3.13 Fluorescence of fusion protein is linear with amount of protein.....70
Figure 3.14 Cleavage of fusion protein inspite of adding PMSF.....71
Figure 3.15 GFP detection on SDS-PAGE by fluorescence.....72
Figure 3.16 The GFP tag can be separated from the protein of interest by Factor Xa cleavage.....73
Figure 4.1 Screening of tobacco plants expressing fusion protein with a C-terminal HDEL.....94
Figure 4.2 Screening of tobacco plants expressing fusion protein without a C-terminal HDEL.....95
Figure 4.3 Fluorescence of intracellular fusion protein expressed in suspension cells is higher with HDEL.....96

Figure 4.4 Fluorescence of fusion protein secreted in the media.....97
Figure 4.5 Western Blot depicting intracellular expression of GMCSF-GFP5(S65T)-HDEL in suspension cells.....98
Figure 4.6 Western Blot depicting expression of GMCSF-GFP5(S65T) (without HDEL).....99

CHAPTER 1

1. INTRODUCTION

1.1 The Green fluorescent protein

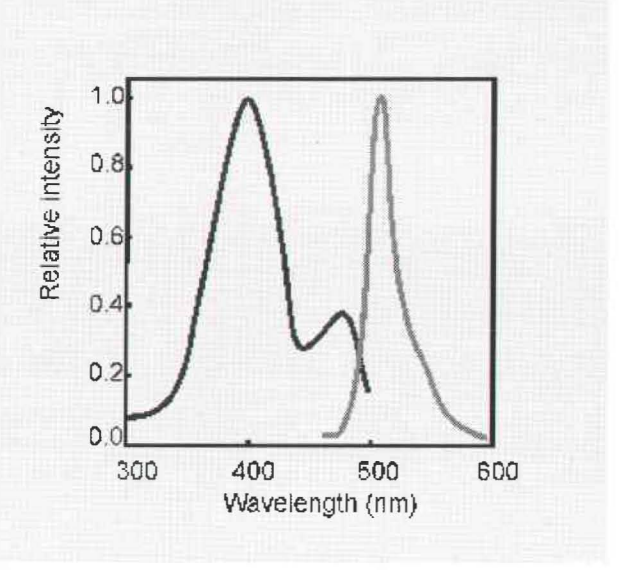
The green fluorescent protein (GFP) from the jellyfish *Aequorea victoria* is one of the most widely used reporter proteins (Tsien, 1998). In the jellyfish when exposed to physical stress or attack, the intracellular calcium ion levels increase and aequorin generates an excited state (Niwa et al., 1996). The excited state energy of aequorin is used to generate bright green GFP fluorescence through radiationless transfer, presumably to blind the attacker (Boxer, 1996). The cDNA for GFP from *Aequorea* was cloned and sequenced and encodes a 238-aa-residue polypeptide with a calculated M_r of 26, 888 Da (Prasher et al., 1992). GFP is an extremely stable protein, remaining fluorescent at temperatures up to 65ºC, pH >11, in the presence of various detergents such as 1% sodium dodecyl sulfate (SDS) (Bokman and Ward, 1981). It is even stable in 6 M guanidinium chloride and is resistant to most proteases for many hours (Ormo *et al.,* 1996). GFP is widely used as a marker for gene expression, for the study of protein localization and protein binding events. Using recombinant DNA technology, the coding sequence of GFP can be spliced with that of other proteins to create fluorescent fusion proteins. GFP tolerates N and C terminal fusions to a broad variety of proteins, many of which have shown to retain function (Cubbit et al., 1995).

1.1.1. Structural and spectral properties

In *A. victoria* light is produced as a result of energy transfer from the Ca^{2+}-activation of the photoprotein aequorin to GFP (Inouye et al., 1994). It absorbs light with an excitation maximum of 395nm (with a minor peak at 470nm) and emits light at 510nm (Morise et al., 1974; Ward et al., 1980) and this can occur in the absence of substrates, enzymes or cofactors, hence ideally suitable as a marker (Chalfie et al.,

1994). The two absorption maxima are caused by a change in the ionization state of the chromophore (Katjusa et al., 1997)

Figure 1.1 Excitation and Emission spectra of wild-type GFP



The chromophore is a p-hydroxybenzylideneimidazoline (Prasher et. al, 1992) formed from residues 65-67, which are Ser-Tyr-Gly in the native protein. Current accepted mechanism of chromophore formation first involves GFP folding into a nearly native conformation, then the imidazoline is formed by nucleophilic attack of the amide of Gly67 on the carbonyl of residue 65, followed by dehydration (Tsein, 1998). Finally molecular oxygen dehydrogenates the α-β bond of residue 66 to put its aromatic group into conjugation with imidazoline. At this stage the chromophore acquires visible absorbance and fluorescence. Atmospheric oxygen is required for fluorescence to develop (Heim et al., 1994; Inouye et al., 1994). Traditionally, fluorescence from GFP-expressing cells is produced using ultraviolet-light (UV) excitation and detected by fluorescence microscopy or by fluorescence-activated cell sorting (FACS).

1.1.2 GFP variants

Wild type GFPs are not optimal for some reporter gene applications. When excited by blue light common to fluorescence microscopy and FACS, such as 488-nm argon-ion laser, the fluorescence intensity from wild type GFPs is relatively low. In addition, a significant lag in the development of fluorescence after protein synthesis can occur and complex photoisomerization of the GFP chromophore may result in the loss of fluorescence. Furthermore, wild-type GFPs are expressed at low levels in many higher eukaryotes (Kain et al., 1995). The cDNA for GFP has been cloned and sequenced (Prasher et al., 1992) permitting the creation of engineered GFP variants to improve upon these limitations (Hiem et al., 1994).

Mutations in GFP, which shift the excitation maxima from 395nm to around 490nm, have been reported and these proteins do fluoresce more intensely when excited at 488nm (Delagrave et al., 1995; Heim et al., 1995). Three GFP mutants were isolated and expressed in *E.coli*, that fluoresced with approximately 100-fold higher intensity than bacteria expressing wild type protein (Cormack et al., 1996).

A new variant, mGFP4 was made by modifying wild type GFP by altering codons at the site of prior mis-splicing and resulted in stable and high expression in transgenic plants (Haseloff et al., 1997). It was found by the same group that it was difficult to regenerate fertile plants from the brightest *mgfp4 transformants*, because the mature protein is found throughout the cytoplasm and nucleoplasm of transformed *Arabidopsis* cells and not any localized compartment within the plant cell. Several targeting peptides were fused to GFP and these variants were expressed in transgenic *Arabidopsis* plants. It was found that the GFP (*mgfp4-ER*) targeted to the endoplasmic reticulum (ER) by a C- terminal amino acid sequence HDEL and an N-terminal signal peptide derived from an *Arabidopsis* vacuolar basic chitinase showed a substantial

improvement in terms of being able to consistently regenerate intensely fluorescent and fertile plantlets and an apparent decrease in phototoxicity. The V163A and S175G substitutions improve GFP fluorescence by enhancing proper folding of apo-GFP, particularly at higher incubation temperatures (Siemering et al., 1996). The I167T substitution alters the excitation spectrum of GFP (Heim et al., 1994) and in combination with V163A and S175G, it produces GFP5 with dual excitation peaks (395nm and 473nm) of approximately equal amplitude which can be visualised well with either long wavelength UV (eg. hand-held lamp) or blue light (eg. argon laser). *E coli* cells expressing GFP5 at 37°C fluoresce 39-fold more intensely than cells expressing GFP when excited at 395nm and 111-fold more intensely when excited at 473nm (Siemering et al., 1996). Cos-7 cells grown at 37°C expressing GFP5 showed higher fluorescence than cells expressing wt-GFP when excited by UV light (Siemering et al., 1996). In this study, we used the mGFP5-ER variant because of its versatile nature in terms of its ability to be used at different temperatures and the fact that it can be used in a broad host range.

1.1.3. Insertions within GFP

Based on the three-dimensional structure of GFP, which shows a compact molecule with solvent exposed loops (Ormo et al., 1996; Yang et al., 1996), it can be predicted that apart from the protein termini, it might be possible to insert amino acids in other sites. Out of 10 such possible sites, it was concluded that the sites Gln157-Lys158, Glu172-Asp173 and Leu194-Leu195 were most permissive for insertion of a hexapeptide, Leu-Glu-Glu-Phe-Gly-Ser (Abedi et al., 1998). Further analysis by evaluation of peptide display libraries, to see if these 3 sites could accommodate a variety of sequences resulted in the conclusion that Gln157-Lys158, Glu172-Asp173

were the most appropriate for insertion of peptides comprising 20 residues of diverse sequence (Abedi et al., 1998).

In order to design a generic biosensor, a protein binding domain TEM1 beta-lactamase (Bla) was inserted into Glu172-Asp173 and the purified fusion protein was shown to have fluorescence and Bla activity, but there was no change in the GFP spectrum upon the addition of the beta-lactamase-inhibitory protein (BLIP) (Doi & Yanagawa, 1999). On random mutagenesis and couple of rounds of selection, a BLIP sensitive sensor was obtained that showed an increase in GFP fluorescence upon the binding of ligands (Doi and Yanagawa, 1999). In a preliminary proof-of-concept study, a poly-histidine (6xHis) tag (see section 1.2.1) was inserted between amino acids 172 and 173 of GFP in this work.

1.2. Immobilized Metal Affinity Chromatography (IMAC)

Recombinant proteins can be purified by affinity based purifications by various systems, most of which rely on the interaction of an immobilized ligand with an affinity tag on the protein of interest. For example appropriate choice of immobilized antibodies have been used to selectively bind antigens in immunoaffintiy chromatography. To use antibodies as ligands requires consideration of desorption conditions and the need to recover the protein under non denaturing conditions (Nachman- Clewner et al., 1993; Sado and Katoh, 1993). Sometimes, large proteinacious affinity tags such as glutathione-S-transferase (GST) are employed to bind and recover a recombinant protein (Braun et al., 2002). Immobilized protein A derived from *Staphylococcus aureus* has been widely used for recovery of immunoglobulins for purposes of both research and therapy, and the purification of monoclonal antibodies from cell culture (Boyle et al., 1993; Kenney and Chase, 1987). All these systems are limited by the proteinaceous nature of the affinity tag

and/or the immobilized ligand. Most protein tags are only functional in their native form and hence cannot be used under denaturing conditions. Many proteins form insoluble bodies when overexpressed in *E coli* and 6M guanidinium or 8 M urea is used to enhance protein solubility. Moreover, large protein-affinity tags tend to be highly antigenic and usually interfere with the biological activity of the protein of interest. Therefore, before functional studies can be conducted or the protein used for antibody induction, the affinity tag must be removed and the protein of interest re-purified.

1.2.1 IMAC using 6xHis affinity tag

This method is based on the fact that a series of six histidine residues can interact with electropositive transition metals, including Co^{2+}, Ni^{2+}, Cu^{2+} and Zn^{2+}(Porath, 1992). An ideal affintiy system would be one that employs a small, non-antigenic protein affinty tag that does not interfere with the biological activity of the attached protein. Moreover, the affinity resin should be durable and reusable, compatible with an array of salts, detergents and alcohols and fully stable in the presence of strong denaturants. 6xHis affinity tag IMAC meets this criteria. In this technique, polydentate chelators are commonly used to fix the metal ions to the solid chromatographic substrate. The most frequently used metal chelators contain three or four electron-rich groups, leaving three or two metal coordination sites available to interact with additional ligands. Tetradentate chelators are superior to tridentate chelators like iminodiacetic acid (IDA) because they firmly bind and retain the metal ion through repeated use and under harsh conditions (Hochuli et al., 1988). Histidine binds to the metal by sharing the imidazole nitrogen electron density with the electron deficient orbitals of transition metals under conditions of physiological pH. This affinity for histidine toward imidazole follows the order $Cu^{2+} > Ni^{2+} > Zn^{2+} > Co^{2+}$

(Sundberg and Martin, 1974). Six histidines have been found to bind proteins in a reliable manner even in the presence of strong denaturants such as guanidium or urea (Hochuli et al., 1987). Very few naturally occuring proteins have multiple neighboring histidines and therefore 6xHis affinity IMAC is commonly used to achieve one-step purification. The extent of purification would depend on the metal used and the background from which the target protein is being purified. Co^{2+}-IDA was effectively used to purify beta-glucuronidase-6xHis (GUSH6) from canola protein extract (Zhang et al., 2000). Though Cu^{2+} has the highest affinity for Histidine, it may not be appropriate to use it in 6xHis affinity chromatography because the binding of the target protein to the metal would be too strong. This phenomenon was seen in the GUSH6 purification from canola protein extract where the recovery of GUS in active form was significantly lower than purifying it with the other metals (Zhang et. al, 2000). The reasons attributed were that immobilized Cu^{2+} is not entirely redox-stable and could be involved in catalytic oxidation reactions of GUSH6 deactivation (Winzerling et al., 1992). Moreover, the binding of GUSH6 might have been too strong that the Cu^{2+}-IDA bonds are broken during the imidazole elution process reulting in elution of inactive GUSH6. Co^{2+} on the other hand did not bind to native GUS (without His tag) or to any proteins from the canola extract and therefore GUSH6 purification from Co^{2+}-IDA gave the best results (Zhang et al., 2000). Therefore, just binding to the immobilized metal is not sufficient. There needs to be a preferential binding of the target protein with respect to the numerous other proteins found in the host system. In this work, TALON cobalt metal afinity resin (Clontech, Palo Alto, CA) was used for protein purification. This resin has enhanced selectivity for poly-histidine tagged proteins and there is no metal loss during purification.

1.3. Protein production in Plants

Transgenic plant technologies have become a cornerstone in modern biotechnology. In the past, most of these techniques aimed at improving disease resistance and developing herbicide resistance. With advances in plant transformation systems it is now feasible to produce biologically active proteins in plants. Plants are a suitable alternative to microbial expression systems, as the production of plant biomass is relatively inexpensive. The potential of transgenic plants as production vehicles has been demonstrated by several examples (Goddijn and Pen 1995). Inspite of the ability of mammalian cell cultures to produce authentic proteins, they often failed to meet the need for a simple efficient, high yielding and low-cost method of protein production. Sometimes, mammalian cell culture can have the risk of dealing with human pathogens. Other drawbacks of using mammalian cells are long periods of time required for high expression levels and the instability of expression (Hippenmeyer and Pegg, 1995).

E coli cell cultures and fermentation alleviate some of the problems of mammalian cultures in terms of production levels and costs but in many cases, the proteins of interest are not truly authentic because of poor post-translational modifications in *E coli* . Examples of industrial enzymes that have been produced in transgenic plants include xylanase, heat-stable β-glucanase and phytase among others (Owen and Pen, 1996). One of the earlier examples of successful heterologous expression of mammalian proteins in plants was the demonstration that leu-enkephalin, a pain-killing neuropeptide, could be produced to high levels when fused to part of the *Brassica napus* 2S seed storage protein (Vandekerckhove et al., 1989). Hepatitis B surface antigen (HbsAg) was expressed in transgenic tobacco by Mason et al. (1992). Genes for several antibodies (Hiatt et al., 1989; Owen et al., 1992),

interferon- α (Zhu et al., 1994) and human serum albumin (Sijmons et al., 1990) have been expressed in plants. Very high levels of α - trichosanthin, a root-derived component of the chinese medicinal plant *Trichosanthes kirilowii*, were obtained using a Tobacco Mossaic Virus (TMV) derived vector for overexpression (Kumagai et al., 1993). Other promising anti-cancer and/or anti-HIV plant derived compounds sucessfully engineered into transgenic plants are ricin (Gadani et al., 1995) and pokeweed antiviral protein (Hemming, 1995). Several therapeutic proteins have also been produced in plant systems. Erythropoetin has been produced in tobacco cells (Matsumoto et al., 1993, 1995). Other human proteins include epidermal growth factor (Higo et al., 1993), lactoferrin (Mitra and Zhang, 1994), Interleukin-2 and Interleukin-4 (Magnuson et al., 1998) and human granulocyte macrophage colony-stimulating factor (James et al., 2000).

1.3.1. Plant cell culture

Plant cells can be culivated in vitro by various systems such as hairy roots (Hilton and Rhodes, 1990), immobilized cells (Archambault, 1991) and free cell suspensions (Kieran et al., 1997). Among these systems, free cell suspensions are considered to be the most suitable for large-scale applications in the biotechnology industry (Su, 1995). Plant suspension cells are an in vitro system that can be used for recombinant protein production under carefully controlled conditions. Since plant cells can be grown on cheaper protein-free media, production and purification costs of target proteins from plant cells is lower (Magnuson et al., 1998). Plant cell culture for production of secondary metabolites has been used in the past. This includes production of shikonin (Tani et al., 1993) and anthocyanins (Sato et al., 1996). Even in the case of enzymes, there are potential advantages of producing the enzymes in plant cell cultures especially those enzymes that require specific post–translational

modification for activity. In comparison to mammalian cells, plant cell culture is risk free because of the absence of human pathogens in the product. A heat stable endo- β-(1-4) glucanse of *Trichoderma reesei* has been produced in suspension cell cultures of barley (Aspegren et al., 1995).

1.4. Signal Peptides and Retention Signals

Recombinant proteins expressed in plant cell suspension cultures are found in the culture supernatant or retained within the cells. This localization depends on the presence of targeting/signal peptides and the permeability of the cell wall for macromolecules (Carpita et al., 1979). Such signals have been used for secretion (Magnuson et al., 1998) or for targetting to organelles like endoplasmic reticulum, chloroplast, vacuoles and membranes (Moloney and Holbrook, 1997). Such targeting to distinct compartments of the cells preserve integrity of the recombinant protein and protects them from proteolytic degradation (Kusnadi et al., 1997). Moreover, accumulation levels could be increased.

For example, fusion of the tetra-peptide KDEL to the C-terminus of vicilin, a pea seed protein, resulted in an average 100-fold increase in accumulation of protein in tobacco leaves and an average 20-fold increase in lucerne leaves. Moreover, there was an average 12-fold increase in the half-life of vicilin expressed in transgenic lucerne leaves (Wandelt et al., 1992). A C-terminal peptide HDEF associated with the tomato Rnase LX localized solely outside the vacuole was able to accumulate proteins in the ER of *Saccharomyces cerevesiae* similar to known ER retention signals (Kalleta et al., 1998). Without the HDEF, the protein was secreted into the medium. It is also believed that this signal might be able to retain proteins in the ER of plants (Kalleta et al., 1998). The carboxy terminal HDEL sequence functions as an ER retention sequence in yeast (Pelham et al., 1988). Golgi specific modifications were

seen in the proteins tagged with HDEL in the ER suggesting that proteins bearing the signal are retrieved from the Golgi giving an insight into the accumulation of proteins within the ER (Dean and Pelham, 1990). However, disruption of the cells is necessary prior to protein purification in the case where proteins are targetted to organelles. This disruption can cause the release of various phenolic substances or proteases that can reduce protein yield. Therefore, secretion of the protein into the media would be more practically viable.

In the present study, expression of granulocyte macrophage - colony stimulating factor (GM-CSF) fused with mGFP5-ER(S65T) has been evaluated in transgenic tobacco suspension cultures. GFP could be a useful tool to monitor the retention or sorting of proteins in the secretory pathway. In this work, an N-terminal signal peptide sequence from *Arabidopsis thaliana* basic chitinase and a C-terminal ER retention signal (HDEL) was used to study accumulation and secretion of fusion proteins in tobacco suspension cell culture.

1.5. Research objectives

As stated above, monitoring, expression and purification of recombinant proteins is of great value in biotechnology. GFP has been used in the past as an efficient tool for monitoring expression and protein localization. To use GFP as a reporter, along with the use of signal peptides/retention signals and poly-histidine tags to enable its accumulation, stabilization and purification, a unique protein tag is required that could accommodate all these features. This is possible if the poly-histidine tag is inserted within the polypeptide chain of the GFP molecule. The main objectives of this study are:

1) To study the effect of insertion of 6xHis tag into the loop region of mGFP5-ER

2) To make a brighter variant of mGFP5-ER and identify the most suitable solvent exposed loop of GFP for insertion of the 6xHis tag
3) To fuse this novel GFP tag to a model protein and analyze the performance of the tag as a fusion protein in different host systems
4) To study the effect of the C-terminal retention signal HDEL on fusion protein expression in tobacco cell suspensions

REFERENCES

Abedi MR, Caponigro G and Kamb A (1998) Green fluorescent protein as a scaffold for intracellular presentation of peptides. Nucleic Acids Resch. 26, 623-630

Archambault J (1991) Large-scale culture of surface-immobilized *Catharanthus roseus* cells. Enzyme Microb Technol 13:882-92

Aspegren K, Mannonen L, Riatala A (1995) Secretion of heat-stable fungal β-glucanase from transgenic, suspension cultured barley cells. Mol Breeding 1:91-99.

Bokman SH, Ward WW (1981) Renaturation of Aequorea green fluorescent protein. Biochem Biophys Res Commun101: 1372-1380

Boyle MDP, Faulmann EL, and Metzger DW (1993) Applications of bacterial immunoglobulin-binding proteins to the purification of immunoglobulins, in Molecular Interactions in Bioseparations, T.T. Ngo, ed., Plenum press, New York, 91-112

Boxer SG (1996) Structural biology. Another green revolution. Nature 383:484-5

Braun P, Hu Y, Shen B, Halleck A, Koundinya M, Harlow E, LaBaer J (2002) Proteome-scale purification of human proteins from bacteria Proc Natl Acad Sci U S A 99:2654-9

Carpita NC, Sabularse D, Montezinos D, and Delmer DP (1979) Determination of the pore size of cell walls in living plant cells. Science 205: 1144-1147

Chalfie M, Tu Y, Euskirchen G, Ward WW, Prasher DC (1994) Green fluorescent protein as a marker for gene expression. 263:802-5

Cormack BP, Valdivia RH, Falkow S (1996) FACS-optimized mutants of the green fluorescent protein (GFP). Gene 173:33-8

Cubitt AB, Heim R, Adams SR, Boyd AE, Gross LA and Tsien RY (1995) Understanding, improving and using green fluorescent proteins Trends in Biochemical Sciences 20: 448-455

Dean N and Pelham HRB (1990) Recycling of proteins from the Golgi compartment to the ER in yeast. J Cell Biol. 111: 369-377

Delagrave S, Hawtin RE, Silva CM, Yang MM, Youvan DC (1995) Red-shifted excitation mutants of the green fluorescent protein Biotechnology (N Y) 13:151-4

Doi N and Yanagawa H (1999) Design of generic biosensors based on green fluorescent proteins with allosteric sites by directed evolution FEBS Letters 453:305-307

Gadani F, Ayers D, Hempfling W (1995) Tobacco: a tool for plant genetic engineering research and molecular farming. Part 11. Agro-Food-industry Hi-Tech 6: 3-6

Goddijn OJM and Pen J (1995) Plants as bioreactors Trends Biotechnol. 13, 379-387

Haseloff J, Siemering KR, Prasher DC and Hodge S (1997) Removal of a cryptic intron and subcellular localization of green fluorescent protein are required to mark transgenic Arabidopsis plants brightly Proc. Natl. Acad. Sci. USA Vol. 94: 2122-2127

Heim R, Prasher, DC and Tsien RY (1994) Wavelength mutations and posttranslational autooxidation of green fluorescent protein Proc. Natl. Acad. Sci. USA 91:12501-12504

Heim R, Cubbit AB and Tsien RY (1995) Improved green fluorescence. Nature 373:663-664

Hemming D (1995) Molecular farming: using transgenic plants to produce novel proteins and other chemicals. AgRiotech News Inf 7: 19N-29N

Hiatt A, Cafferkey R and Bowdish K (1989) Production of antibodies in transgenic plants. Nature 432: 76-78.

Higo K, Saito Y, Higo H (1993) Expression of a chemically synthesized gene for human epidermal growth factor under the control of cauliflower mosaic virus 35S promoter in transgenic tobacco Biosci Biotechnol Biochem 57:1477-81

Hilton MG, Rhodes MJ (1990) Growth and hyoscyamine production of 'hairy root' cultures of Datura stramonium in a modified stirred tank reactor Appl Microbiol Biotechnol 33:132-8

Hippenmeyer PJ, Pegg LE (1995) Enhancing expression of recombinant proteins in mammalian cells using the herpesvirus VP16 transactivator Curr Opin Biotechnol 6:548-52

Hochuli E, Dobeli H and Schacher A (1987) New metal chelate adsorbent selective for proteins and peptides containing neighbouring histidine residues J Chromatography 411:177-184

Hochuli E, Bannwarth W, Dobeli H, Gentz R and Stuber D (1988) Genetic approach to facilitate purification of novel recombinant proteins with a novel meta chelate adsorbent Bio/Technology 6: 1321-1325

Inouye S and Tsuji FI (1994) Aequorea Green fluorescent protein: Expression of the gene and fluorescence characteristics of the recombinant protein. FEBS letters 341: 277-280

James EA, Wang C, Wang Z, Reeves R, Shin JH, Magnuson NS, Lee JM (2000) Production and characterization of biologically active human GM-CSF secreted by genetically modified plant cells Protein Expr Purif 19:131-8

Kain SR, Adams M, Kondepudi A, Yang TT, Ward WW, Kitts P (1995) Green fluorescent protein as a reporter of gene expression and protein localization BioTechniques 19: 650-655

Kaletta K, Kunze I, Kunze G, Kock M (1998) The peptide HDEF as a new retention signal is necessary and sufficient to direct proteins to the endoplasmic reticulum FEBS Lett 434:377-81

Katjusa Brejc, Sixma TK , Kitts PA, Kain SR, Tsien RY , Ormö M, and Remington SJ (1997) Structural basis for the dual excitation and photoisomerisation of the *Aequorea victoria* green fluorescent protein Proc. Natl. Acad. Sci. USA Vol. 94: 2306-2311

Kenney AC and Chase HA (1987) Automated production scale affinity purification of monoclonal antibodies. J. Chem. Tech. Biotechnol. 39: 173-182

Kieran PM, MacLoughlin PF, Malone DM (1997) Plant cell suspension cultures: some engineering considerations J Biotechnol 59:39-52

Kumagai MH, Turpen TH, Weinzeni N, Della-Cioppa G, Turpen AM, Donson JD, Hilf ME, Grantham GL, Dawson WO, Chow TP, Piatak M, Grill LK

(1993) Rapid, high- level expression of biologically active alpha-tricosanthin in transfected plants by an RNA viral vector. Proc. Natl. Acad. Sci., USA 90: 427-430

Kusnadi AR, Nikolov ZL, Howard JA (1997) Production of recombinant proteins in transgnic plants: Practical considerations. Biotechnol. Bioeng. 56: 473-484

Magnuson NS, Linzmaier PM, Reeves R, An G, Hayglass K and Lee JM (1998) Secretion of biologically active human Interleukin-2 and Interleukin-4 from genetically modified tobacco cells in suspension culture. Protein Express. & Purif. 13, 45-52

Mason HS, Man-kit LD and Arntzen CJ (1992) Expression of hepatitis B surface antigen in transgenic plants. Proc Natl Acad Sci USA. 89: 11745-11749.

Matsumoto S, Ikura K, Ueda M and Sasaki R (1993) Expression of human erythropietin in cultured tobacco cells. Biosci. Biotech. Biochem. 57: 1249-1252.

Matsumoto S, Ikura K, Ueda M and Sasaki R (1995) Characterization of a human glycoprotein (erythropoietin) produced in cultured tobacco cells. Plant Mol. Biol. 27:1163-1172.

Mitra A, Zhang Z (1994) Expression of a human lactoferrin cDNA in tobacco cells produces antibacterial protein(s). Plant Physiol 106:977-81

Moloney MM and Holbrook LA (1997) Subcellular targeting and purification of recombinant proteins in plant production systems. Biotechnol. Genet. Eng. Rev 14, 321-336

Morise H, Shimomura O, Johnson FH, Winant J (1974) Intermolecular energy transfer in the bioluminescent system of Aequorea. Biochemistry 13:2656-62

Nachman-Clewner M, Spence C and Bailon P (1993) Receptor-affinity Chromatography (RAC) in Molecular Interactions in Bioseparations, T.T. Ngo, ed., Plenum Press, New York, 139-167

Niwa Y, Hirano T, Yoshimoto K, Shimizu M and Kobayashi H (1999) Non-invasive quantitative detection and applications of S65T-type green fluorescent protein in living plants. The Plant J.18: 455-463

Ormo M, Cubitt AB, Kallio K, Gross LA, Tsien RY and Remington SJ (1996) Crystal structure of the *Aequorea* green fluorescent protein. Science 273: 1392-1395

Owen MRL, Pen J (eds) (1996) Transgenic plants: A production system for industrial and pharmaceutical proteins. Wiley New York

Owen M, Gandecha A, Cockburn B, Whitelam G (1992) Synthesis of a functional anti-phytochrome single-chain Fv protein in transgenic tobacco. Biotechnology (NY) 10:790-4

Pelham HRB, Hardwick KG and Lewis MJ (1988) Sorting of soluble ER proteins in yeast. EMBO J. 7: 1757-1762

Porath J (1992) Immobilized metal ion affinity chromatography. Prot. Express. & Purif. 3: 263-281

Prasher DC, Eckenrode VK, Ward WW, Prendergast FG and Cormier MJ (1992) Primary structure of the *Aequorea victoria* green-fluorescent protein. GENE 111: 229-233

Sado E and Katoh S (1993) Suitable antibodies as ligands in affinity chromatography of biomolecules, in Molecular Interactions in Bioseparations, T.T. Ngo, ed., Plenum Press, New York, 205-211.

Sato K, Nakayama M and Shigeta J (1996) Culturing conditions affecting the

production of anthocyanins in suspended cell cultures of strawberry. Plant Sci. 113: 91-98.

Siemering KR, Golbik R, Sever R and Haseloff J (1996) Mutations that suppress the thermosensitivity of green fluorescent protein. Curr. Biol. 6:1653-1663

Sijimons PC, Dekker BM, Schrammeijer B, Verwoerd TC, van den Elzen PJM and Hoekema A (1990) Production of correctly processed human serum albumin in transgenic plants. Bio/Tech 8: 217-221.

Su, W.W (1995) Bioprocesssing technology for plant cell suspension cultures. Appl.Biochem. Biotechnol. 50: 189-230.

Sundberg RJ and Martin RB (1974) Interactions of histidine and other imidazole derivatives with transition metal ions in chemical and biological systems. Chem Rev 74:471-517

Tani M, Takeda K, Yazaki K, and Tabata M (1993) Effects of oligogalacturonides on biosynthesis of shikonin in *Lithospermum* cell cultures. Phytochemistry 34:1285

Tsien RY (1998) The Green Fluorescent Protein.Ann. Rev. Biochem. 67:509-544.

Vandekerchove J, Van Damme J, Lijsebetterns VM, Botterman J, De Block M, Vandewiele M, De Clercq A, Leemans J, Montagu VM and Krebbers E (1989) Enkephalins produced in transgenic plants using modified 2s seed storage proteins. Bio/tech 7, 929-932.

Wandelt CI, Khan MRI, Craig S Schroeder HE, Spencer D and Higgins TJV (1992) Vicilin with carboxy-terminal KDEL is retained in the endoplasmic reticulum and accumulation to high levels in the leaves of transgenic plants. Plant J. 2:181-192

Ward WW, Cody CW, Hart RC, and Cormier MJ (1980) Spectrophotometric identity of the energy transfer of chromophore in Renilla and Aequorea green-fluorescent proteins. Photochem. Photobiol. 31: 611-615

Winzerling JJ, Berna P, Porath J (1992) How to use immobilized metal ion affinity chromatography. Methods 4:4-13

Yang F, Moss LG and Phillips Jr., GN (1996) The molecular structure of green fluorescent protein. Nature Biotechnology 14, 1246-1251

Zhang CM, Resllewic SA, Glatz CE (2000) Suitability of immobilized metal affinity chromatography for protein purification from canola. Biotech. Bioeng. 68: 52-58

Zhu Z, Hughes KW, Huang L, Liu C, Hou Y and Li X (1994) Expression of human α-interferon cDNA in transgenic rice plants. Plant cell tissue organ culture 36:197-204.

CHAPTER 2

PRELIMINARY STUDY ON GFP WITH AN INTERNAL HISTIDINE TAG

2.1. INTRODUCTION

The green fluorescent protein (GFP) from the jellyfish *Aequorea victoria* is one of the most widely used reporter proteins (Tsien, 1998). One of the advantages of GFP is that it tolerates N- and C-terminal fusions to a variety of proteins and can be expressed in a broad range of host organisms (Cubitt et al., 1995). Over the years the number of applications has expanded tremendously. The original GFP cloned from the jellyfish had certain limitations such as low brightness and photoisomerization (Cubitt et al., 1995). This prompted development of improved variants of GFP.

One such improved variant is GFP5, which is thermostable and shows improved spectral properties (Siemering et al., 1996). GFP5 was developed by random mutagenesis of a bright and thermostable GFP4 mutant having the V163A and S175G mutations. Introducing the I167T (Heim et al., 1994) substitution along with these mutations resulted in GFP5, which has dual excitation peaks at 395nm and 473nm of nearly equal amplitude with no change in the thermostability. As a result, the range of applications in which it can be used is broadened since long wavelength UV or blue light could be used for excitation, unlike the wild-type (wt) GFP which needs to be excited with UV light. GFP5 expressed in bacterial cells that were grown at 37°C was shown to fluoresce 111-fold more brightly than wt-GFP when excited at 473nm (Siemering et al., 1996). The same group showed that GFP5 produced in COS-7 cells grown at 37°C showed higher fluorescence than wt-GFP when excited by UV light. In order to express GFP5 in *Arabidopsis*, its codon usage was modified to eliminate plant intron recognition sequences and was designated as mGFP5. This

modified GFP5 was targeted to the endoplasmic reticulum (ER) by an N-terminal signal peptide and retained on the ER network *via* a C-terminal ER-retention signal (HDEL), resulting in highly fluorescent transgenic plants (Haseloff et al., 1997). These reports demonstrated that GFP5 is a highly versatile variant that can be expressed in a variety of host organisms.

GFP has a β-can structure formed by 11-antiparallel beta strands to form a compact cylinder with some of the connecting loops exposed to solvent and the chromophore buried inside the cylinder (Ormo et al., 1996; Yang et al., 1996). From this structure, it can be predicted that apart from the protein termini, it might be possible to insert amino acids at other sites such as the solvent exposed connecting loops and retain GFP fluorescence. A GFP based biosensor was designed by inserting a binding domain between Glu172 and Asp173 and an increase in fluorescence was noted when ligand was bound (Doi and Yanagawa, 1999). To our knowledge, a hexa-histidine fragment inserted into the loop region of GFP has not yet been reported.

In this chapter, we have attempted a proof-of concept study to determine if the insertion of a 6xHis tag in one of the solvent exposed loops of mGFP5-ER (between amino acids 172 and 173) would result in a novel GFP with dual functions, that of being able to monitor any recombinant protein fused to it and also have the ability to purify it by immobilized metal afffinity chromatography. As a result of the tag being within the GFP molecule, a lot of flexibilty in designing fusion proteins is provided since appropriate signal peptides and retention signals could be attached to the N- and/or C-terminus which have advantages of better accumulation and stability of the recombinant protein. Some of the prime factors considered were to check if the GFP retained fluorescence inspite of the insertion of the tag and to determine if the 6xHis tag could bind to immobilized metal and purify the protein.

2.2. MATERIALS AND METHODS

2.2.1. Modification of mGFP5-ER

The mgfp5-ER insert was obtained by digestion of pBIN-mgfp5-ER with *Bam*HI-*Sac*I and ligated into the respective sites of Bluescript SK (Stratagene, La Jolla, CA). The sequence was modified through a number of polymerase chain reaction (PCR) amplifications using Platinum Pfx DNA Polymerase (Invitrogen, Carlsbad, CA) and oligonucleotides for site-directed mutagenesis and insertion of a poly-histidine sequence (Figure 2.1). For insertion of the poly-histidine sequence between amino acids 172 and 173 of the mgfp5-ER, a two-step PCR strategy was used. The N-terminal fragment (amino acids 1 –172) was amplified by primer A (T7 primer) and Primer H (5'- TTC GAT GTT GTG GCG GGT CTT G -3'). Similarly, a C-terminal fragment (amino acids 168-242 including the –HDEL terminal sequence) was amplified by primer E (5'-CGC CAC AAC ATC GAA **CAC CAT CAC CAT CAC CAT** GAC GGC GGC GTG CAA CTC GC -3') to incorporate a stretch of six histidine residues (bold) and a region to overlap with the N-terminal fragment (underlined region which encodes amino acids 168-172) and primer J (T3 primer). The amplified fragments were purified by agarose gel electrophoresis and used together to amplify the entire mgfp5-ER using primers A (T7 primer) and J (T3 primer). An internal *Nco*I restriction site was removed by a silent mutation (changing a A to T) using the two-step PCR strategy. The N-terminal fragment was amplified using primers A (T7 primer) and F (5'- GTG TTG GCC AAG GAA CAG GTA -3') whereas the C-terminal fragment was amplified using primers C (5'- TAC CTG TTC CTT GGC CAA CAC -3') and J (T3 primer). The purified fragments were used together to amplify the entire mgfp5-ER using primers A (T7 primer) and J (T3 primer). One final amplification of the mgfp5-ER(S65T) insert was done using

primer B (5'- GGC AGG AGG AAC CAT GGC TAG CAA AGG AGA AGA ACT TTT CAC TGG AG -3') to incorporate an N-terminal methionine (bold), an *Nco*I site (underlined) and removal of the *Arabidopsis thaliana* chitinase signal peptide and primer J (T3 primer). The amplified fragment was digested with *Nco*I-*Sac*I and ligated into the respective sites of pGEM-5Zf (Promega, Madison, WI). The product from translation of this modified gene will be referred to as LoopGFP5. All modifications to the sequence were verified by DNA sequencing.

The control, mGFP5-ER(S65T) with 6xHis in the C-terminus (GFP6xHis) was made using the primer B as sense and 5'-CGG GCA GAG CTC TTA **ATG GTG ATG GTG ATG GTG** AAG CTC ATC ATG TTT GTA TAG TTC- 3' as antisense primers to to retain SacI (underlined) and insert a 6xHis tag (bold). GFPHisER was made using primer B and 5'-CGG GCA GAG AGC TCT TAA AGC TCA TCA TGA TGG TGA TGG TGA TGG TGT TTG TAT AGT TCA TCC –3'

2.2.2. GFP expression in E coli

The modified mgfp5-ER(S65T) inserts were obtained by digestion with *Nco*I-*Sac*I and ligated into the respective sites of pET-21d vector (Novagen, Madison, WI). The resulting plasmids were transformed into *E. coli* BL21*trxB*(DE3) competent cells (Novagen, Madison, WI). Transformants containing the GFP constructs produced yellow-green colonies. The screened colonies were then used to grow up *E coli* in Luria-Bertani (LB) media. Initially, a colony was picked up and grown in 3 ml of TB media till the OD at 600 nm became around 0.6. This culture was then kept overnight at 4°C. 1 ml of this culture was then used to inoculate 50 ml of LB media with 100 µg/ml Ampicillin. When the OD at 600nm was approximately equal to 0.6 , the culture was induced with 1mM isopropyl-beta-D-thiogalactopyranoside (IPTG) and then grown for another 4 hrs.

2.2.3. Preparation of crude E coli lysate

The cells were harvested by centrifugation for 15 min at 3000xg at 4°C and resuspended in 50 mM NaH_2PO_4 (pH 7.0), 300 mM NaCl. Lysozyme was added to a final concentration of 0.75 mg/ml and the resuspended cells were incubated at room temperature for 30 min. The sample was sonicated 6 x 20 sec on ice at a minimum power setting using an ultrasonic cell disruptor equipped with a microprobe (Branson Model 250, Danbury, CT). The lysate was centrifuged for 20 min at 12,000xg at 4°C and soluble GFP in the supernatant was purified by immobilized metal affinity chromatography (IMAC) using TALON cobalt resin columns (Clontech, Palo Alto, CA).

2.2.4. Purification by IMAC

The contents of the spin columns were mixed thoroughly and the resuspended resin was transferred to a 15ml sterile tube. The tube was centrifuged at 700 x g for 2 min to pellet the resin. The 20% ethanol supernatant was removed and discarded. 5 volumes of sonication buffer (50mM NaH_2PO_4 (pH 8.0) 10 mM Tris-HCl (pH 8.0) was added to the tube and the contents mixed briefly to pre-equilibrate the resin. The tube was centrifuged as before to pellet the resin, and the supernatant was discarded. Once equilibrated, the crude *E coli* lysate was added to the resin. To bind the His-tagged protein, the suspension was gently agitated at room temperature for 45 min and then centrifuges at 700 x g for 5min. The supernatant was removed without disturbing the resin pellet. The resin was washed by adding 10 volumes of sonication buffer (pH 8.0). The suspension was gently agitated at room temperature for 10 min and centrifuged at 700 x g for 5 min. The supernatant was removed and discarded as before. The wash step was repeated and then the resin was transferred back to the column. Two column volumes of sonication buffer was added for a total of 3 times

followed by 1 x 2 column volumes of wash buffer (20mM Tris-HCl (pH 8.0), 100mM NaCl, 10 mM Imidazole. The 6xHis protein was eluted by adding elution buffer (20mM Tris-HCl (pH 8.0), 100mM NaCl, 50 mM Imidazole and the fractions were collected. The fluorescence of the fractions was measured as described below.

2.2.5. Protein Assay

Protein was determined using the Micro BCA Protein Assay Reagent (Pierce Chemical Co., Rockford, IL) with bovine serum albumin or purified recombinant GFP protein (Clontech, Palo Alto, CA) as the standard.

2.2.6. Analysis of proteins

Protein samples were denatured in SDS sample buffer containing 2-mercaptoethanol for 3 min at 100°C and resolved on a 12% SDS-polyacrylamide gel as described (Laemmli 1970). Gels were stained with Coomassie brilliant blue.

2.2.7. Fluorescence measurements

GFP fluorescence was measured using a fluorescence spectrophotometer (Hitachi Model F-2500, Tokyo, Japan). Wavelength scans were performed on all GFP variants to determine excitation and emission peaks. Unless otherwise stated, the fluorescent intensity per unit protein was determined with the GFP samples diluted in GFP buffer (50 mM NaH_2PO_4 (pH 8.0), 10 mM Tris-HCl (pH8.0), 200 mM NaCl) with excitation set at 468nm and emission set at 505 nm.

2.3. RESULTS AND DISCUSSION

Since the polyhistidine tag in the loop is exposed, the purity of the GFP obtained was quite high as can be seen from the SDS PAGE gel (Lane 2, Figure 2.1). An interesting thing to note is that having four amino acids (HDEL) after the His tag (GFPHisHDEL) does not effect the accessibility of the tag to the immobilized metal as can be seen from the purity of Lane 3 in Figure 2.1.

Since GFP tolerates additions to its N and C-terminus effect of having the HDEL after the His tag is negligible. However, HDEL is a C-terminal ER retention signal in eukaryotes like animals (Munro and Pelham, 1987) and yeast (Pelham, 1988) and is also believed to work in plants (Gomord et al., 1997). Since this protein has been expressed in *E coli* with no endoplasmic reticlum (ER), the biological function/effect of HDEL cannot be tested in this system. In yeast, two receptors ERD1 (Hardwick et al., 1990) and ERD2 (Semenza et al., 1990) have been identified that recognizes the HDEL sequence. Furthermore, the efficiency of the ER signal depends on the sequence to which it is fused in terms of presenting the signal to the receptor as suggested by Herman et al., 1990. It is not clear whether the receptor that recognizes these four amino acids will function in a normal manner with the highly positive charge introduced by the His tag adjacent to it. The same applies to other signals attached to the C-terminus of a protein. Therefore, inserting the tag in the loop would be a more viable option because the C-terminus is then free to fuse an appropriate signal.

A wavelength scan was done using HITACHI F-2500 fluorescence spectrophotometer and the excitation and emission peaks were found to be 467nm (Figure 2.2) and 505 nm respectively. Fluorescence intensities per unit protein were compared for the different variants of GFP (Figure 2.3). The fluorescence intensity of LoopGFP was about 47.8% of the intensity of GFPHDELHis. This is consistent with the excitation spectra that is obtained for the GFPs (Figure 2.2) where it was found that the 467nm excitation peak for LoopGFP was lower than the peak for GFPHDELHis by approximately the same amount. The microenvironment around the chromophore would have changed as a result of the insertion resulting in less molar absorbance and hence decrease in emission by similar amounts. Addition of four

amino acids (HDEL) after the His tag has negligible changes in the fluorescence as is seen for the 467nm peak in Figure 2.2 where the peak height is 92.7% of GFPHDELHis.

It can be seen from the above results that inserting the polyhistidine tag (6xHis) between amino acids 172 and 173, does cause a decrease in the fluorescence intensity. However, the possible advantages by the insertion could far outweigh the problem of this reduction in intensity. Inserting the tag in a brighter variant of GFP could alleviate this problem. We expect a higher fluorescence intensity by inserting the polyhistidine tag in GFP with a S65T mutation (Heim et al., 1994).

2.4. CONCLUSIONS

From this preliminary study, it can be concluded that inserting a 6xHis tag in one of the solvent exposed loops of GFP does effect the fluorescence of the new variant. In this study, we have chosen the mGFP5-ER variant for the insertion. It would be better to insert the tag in a brighter variant such as the GFP with the S65T mutation (Heim et al., 1994). This can result in a novel GFP with dual functions, that of being able to monitor expression of any recombinant protein fused to it and the also have the ability to purify it in a single step. The significance of this GFP is more prominent with appropriate choice of signal peptides and retention signals. This is because of the fact that recombinant proteins can be targeted to different compartments by N- terminal signal peptides and sometimes C- terminal ER retention signals are used for better accumulation and increasing the half life of the target protein. Therefore, by having the tag in the loop, flexibility is provided to incorporate these signals at the N- or C-terminus. Moreover, it would be worthwhile to consider other positions in the GFP molecule for the insertion because of the number of solvent exposed loops in the GFP molecule. Therefore, it is possible that there might be other

positions that would tolerate the insertions. Once the ideal site for insertion of the tag is determined, the novel GFP molecule can be then fused to a model protein and tested for its dual functional role as a fusion protein.

REFERENCES

Cubitt AB, Heim R, Adams SR, Boyd AE, Gross LA and Tsien RY (1995) Understanding, improving and using green fluorescent proteins. Trends in Biochemical Sciences 20: 448-455

Doi N and Yanagawa H (1999) Design of generic biosensors based on green fluorescent proteins with allosteric sites by directed evolution. FEBS Lett.453: 305-307

Gomord V, Denmat LA, Fitchette-Laine AC, Satiat-Jeunemaitre B, Hawes C, Faye L (1997) The C-terminal HDEL sequence is sufficient for retention of secretory proteins in the endoplasmic reticulum (ER) but promotes vacuolar targeting of proteins that escape the ER. Plant J 11:313-25

Laemmli UK (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature 227: 680-5

Hardwick KG, Lewis MJ, Semenza J, Dean N and Pelham HRB (1990) ERD1, a yeast gene required for the retention of luminal endoplasmic reticulum proteins, affects glycoprotein processing in the Golgi apparatus. EMBO J 9: 623-630

Haseloff J, Siemering KR, Prasher DC and Hodge S (1997) Removal of a cryptic intron and subcellular localization of green fluorescent protein are required to mark transgenic *Arabidopsis* plants brightly. Proc. Natl. Acad. Sci. USA 94: 2122-2127

Herman EM, Tague BW, Hoffman LM, Kjemtrump SE and Chrispeels MJ (1990) Retention of phytohaemagglutinin with carboxyterminal tetrapeptide KDEL in the nuclear envelope and the endoplasmic reticulum. Planta 182: 305 - 312

Heim R, Prasher, DC and Tsien RY (1994) Wavelength mutations and posttranslational autooxidation of green fluorescent protein. Proc. Natl. Acad. Sci. USA 91:12501-12504

Munro S, Pelham HR (1987) A C-terminal signal prevents secretion of luminal ER Proteins. Cell 48: 899-907

Ormo M, Cubitt AB, Kallio K, Gross LA, Tsien RY and Remington SJ (1996) Crystal structure of the *Aequorea* green fluorescent protein. Science 273:1392-1395

Pelham HR (1988) Evidence that luminal ER proteins are sorted from secreted proteins in a post-ER compartment. EMBO J 7: 913-8

Semenza JC, Hardwick KG, Dean N, Pelham HR (1990) ERD2, a yeast gene required for the receptor-mediated retrieval of luminal ER proteins from the secretory pathway. Cell 61:1349-57

Siemering, K.R., Golbik, R., Sever,R. and Haseloff,J. (1996) Mutations that suppress the thermosensitivity of green fluorescent protein Curr. Biol. 6:1653-1663

Semenza JC, Hardwick KG, Dean N and Pelham HRB (1990) ERD2, a yeast gene required for the receptor mediated retrieval of luminal ER proteins from the secretory pathway. Cell 61: 1349-1357

Tsien RY (1998) The Green Fluorescent Protein. Ann. Rev.Biochem.67: 509-544

Yang F, Moss LG and Phillips Jr., GN (1996) The molecular structure of green fluorescent protein. Nature Biotechnology 14, 1246-1251

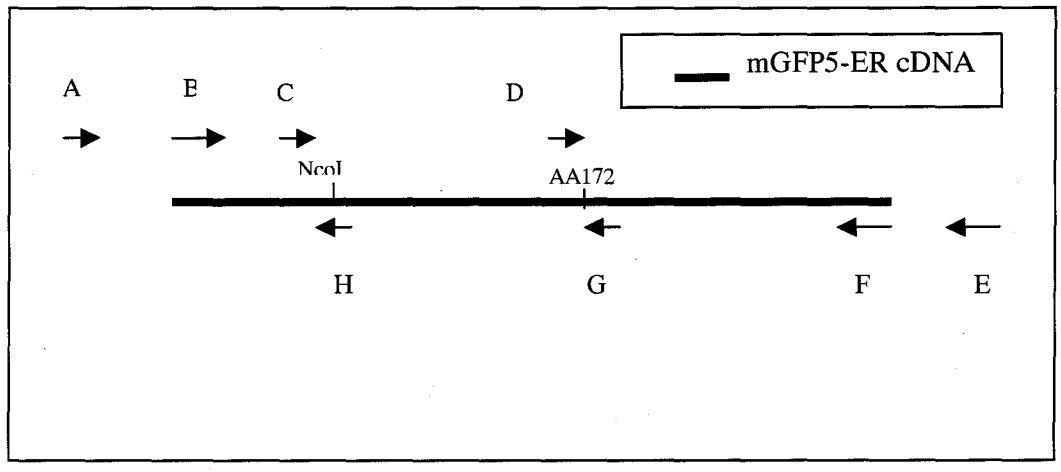


Figure 2.1: Primers used for vector construction

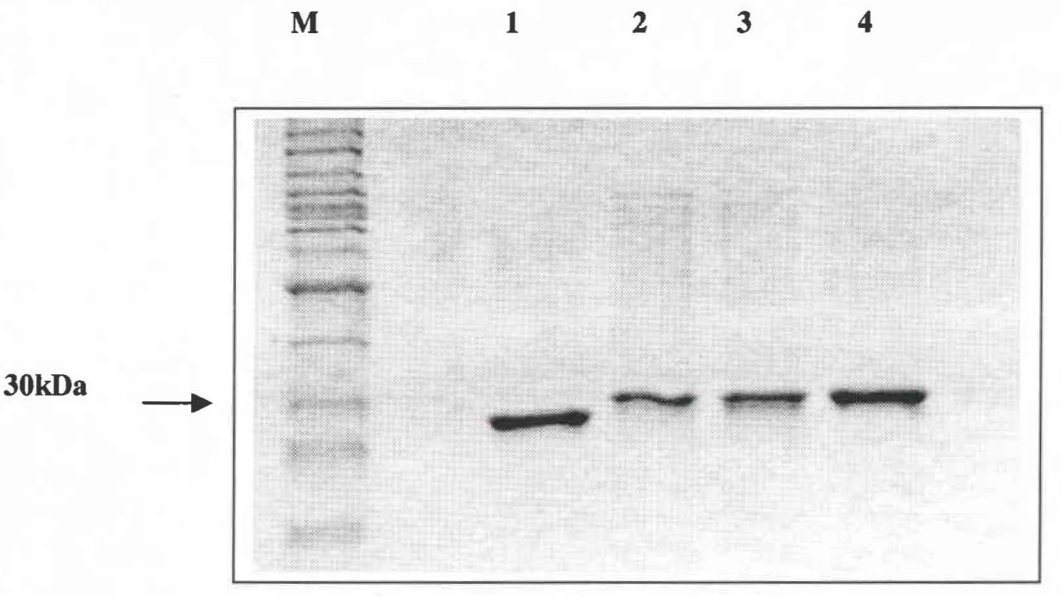


Figure 2.2 GFP with internal 6xHis tag can be purified from E coli lysate. Lanes show the purified fractions after eluting from the IMAC column. *Lane 1:* 2μg of GFP standard (Clontech, Palo Alto, CA); *Lane2:* LoopGFP; *Lane3:* GFPHisHDEL; *Lane4* GFPHDELHis.

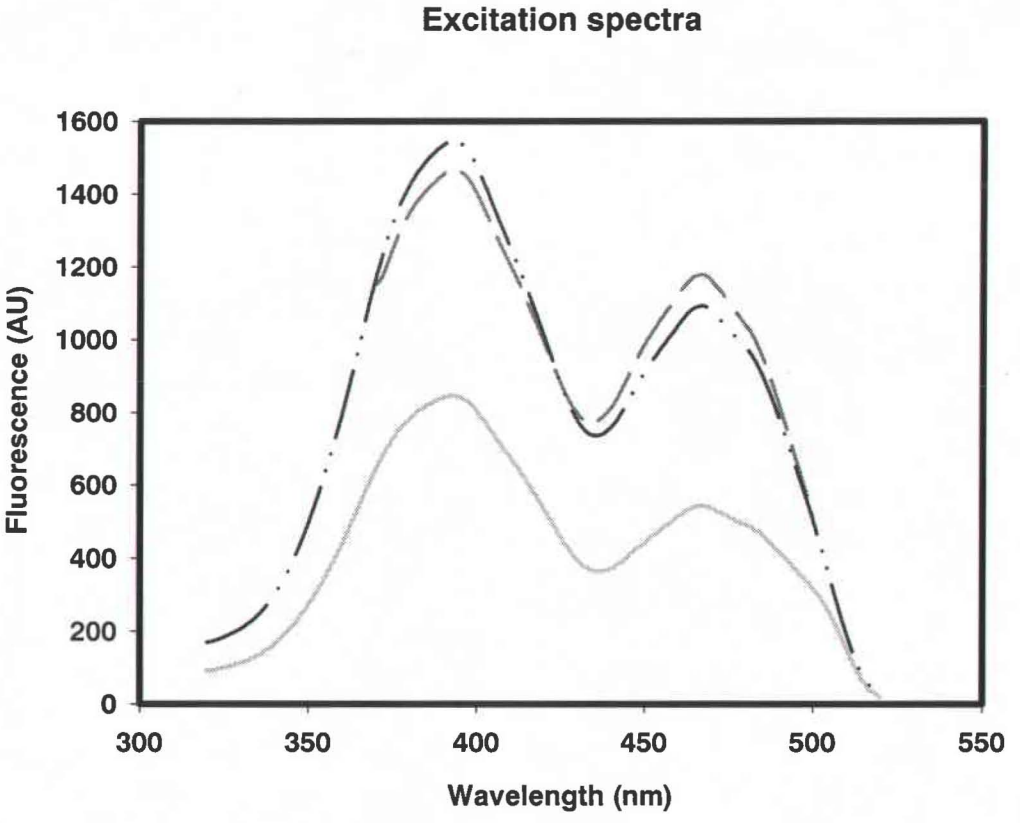


Figure 2.3: Excitation spectra for GFP5 variants. A total of 5 µg each of LoopGFP (__) ; GFPHisHDEL (__ .. __) and GFPHDELHis (_ _) was made upto 1 ml in GFP extraction buffer and a wavelength scan was done using a fluorescence spectrophotometer (HITACHI F-2500).

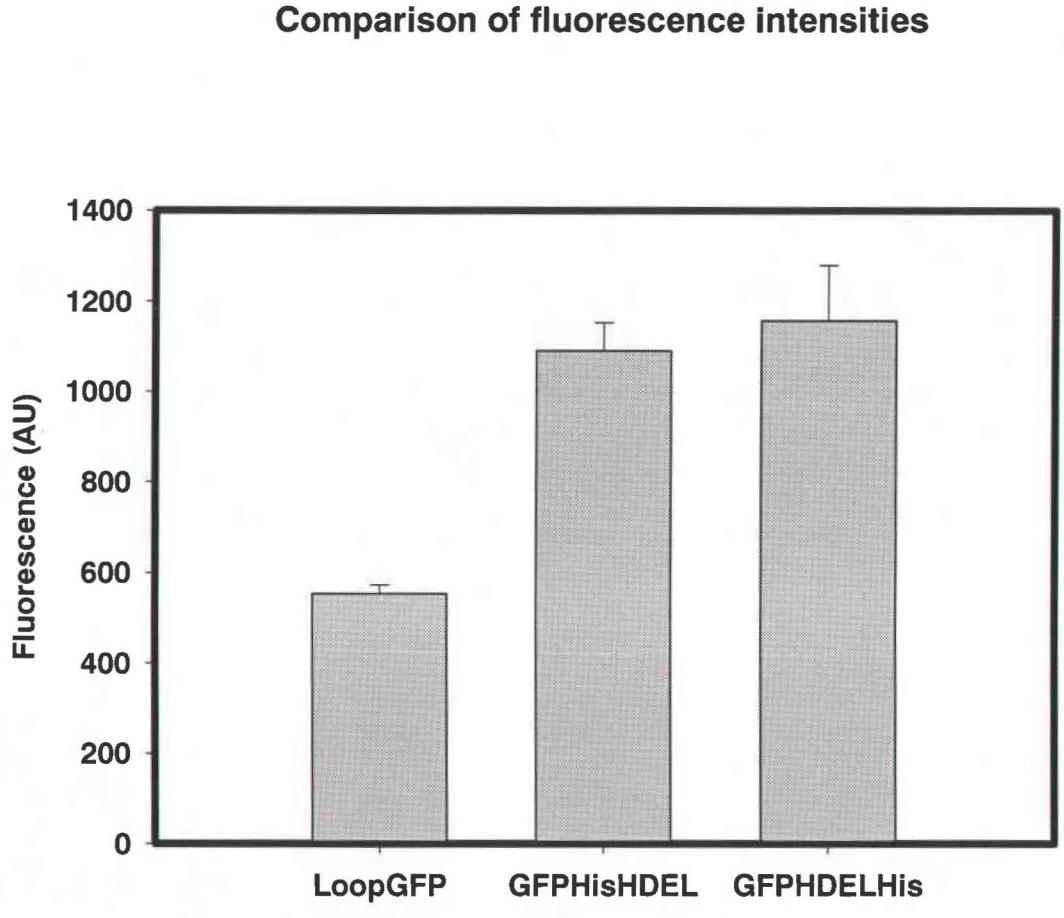


Figure 2.4: Intensity of GFP5 variants. 5 μg of GFP was made up in GFP extraction buffer to 1 ml and Fluorescence measured in a quartz cuvette with excitation set at 468nm and emission set at 505nm. Error bars represent standard deviations for measurements made in duplicate.

CHAPTER 3

DEVELOPMENT AND CHARACTERIZATION OF GFP VARIANTS AS A DUAL FUNCTIONAL TAG

3.1. INTRODUCTION

The green fluorescent protein (GFP) from the bioluminescent jellyfish *Aequorea victoria* is one of the most widely used reporter proteins (Tsien, 1998). Over the years the number of applications of GFP has expanded tremendously. The original GFP cloned from the jellyfish had certain limitations such as low brightness, photoisomerization (Cubitt et al., 1995) and improper folding at 37° C (Siemering et al., 1996). This prompted development of improved variants of GFP such as GFP5, which is thermostable and had improved spectral properties (Siemering et al., 1996). In order to further improve the functions of GFP, preliminary experiments were conducted to determine the consequences of inserting a 6xHis tag in the loop region of mGFP5. The rationale being that if there is enough fluorescence retained inspite of the insertion, then this GFP would have the dual functions of being able to monitor any protein fused to it and also purify it. It was found that the fluorescence of this LoopGFP was about 47% of the fluorescence of having the tag in the C-terminus (see Results in Chapter 2). It was found that GFP needs to be highly expressed in order to mask the autofluorescence of chlorophyll in plants and mGFP4 was suitable for this purpose (Haseloff et al., 1997). GFP5 was developed by altering codon usage to disrupt a cryptic plant intron and by incorporating the mutations V163A, I167T,and S175G to increase thermotolerance and improved spectral properties (Siemering et al., 1996). Since there was a decrease in fluorescence as a result of the insertion, it would be worthwile to insert the 6xHis tag in a brighter variant of GFP.

It was established that mutation such as S65T (replacement of serine in position 65 with threonine) made GFP fluoresce 6-fold more intensely at 490nm and the post-translational oxidation of the chromophore was 4-fold faster than wt-GFP (Heim et al., 1995). Moreover, Cos-7 cells expressing GFP5(S65T) fluoresced 1.65 times higher than cells expressing GFP5 grown at 37 °C (Siemering et al., 1996). Moreover, S65T-type GFP can be excited by blue light making it suitable for on-line monitoring applications that require long exposures to strong light. S65T-type GFP was highly expressed in transgenic *Arabidopsis* plants and non-disruptive fluorescence measurements were made in whole plants with blue light excitation (Niwa et al., 1999). In other studies involving plant cells, an engineered codon optimised sGFP(S65T) showed 100-fold brighter fluorescence than the original jellyfish GFP sequence (Chiu et al., 1996).

By analysis of the solvent exposed loops of GFP it was concluded that the sites Gln157-Lys158, Glu172-Asp173 and Leu194-Leu195 were most permissive for insertion of a hexapeptide, Leu-Glu-Glu-Phe-Gly-Ser (Abedi et al., 1998). Further analysis by evaluation of peptide display libraries to determine if these 3 sites could accommodate a variety of sequences resulted in the conclusion that Gln157-Lys158 and Glu172-Asp173 were the most appropriate for insertion of peptides comprising 20 residues of diverse sequence (Abedi et al., 1998). In another report, a GFP based biosensor was designed by inserting a binding domain between Glu172 and Asp173 and an increase in fluorescence was noted when ligand was bound (Doi and Yanagawa, 1999). An epitope for haemagglutinin was inserted at positions 172-173 and 157-158 of GFP in order to make affinity fluorescent proteins by Matsudaira et al. (2001).

In this work we describe two novel variants of GFP. The GFPs were developed based on the mGFP5 variant (Siemering et al., 1996) with an additional S65T mutation. Into this GFP backbone, a poly-histidine tag (6xHis) was inserted between Gln172 and Asp173 (GFP172) in one variant and between Gln157 and Lys158 (GFP157) in the other variant. Both GFP172 and GFP157 were found to have high fluorescence despite insertion of the 6xHis-tag and were purified by immobilized metal affinity chromatography. Since the 6xHis tag is inserted within the GFP molecule, it allows protein/peptide fusion to both N- and C-terminals of the GFP tag. On further characterization of GFP172 and GFP157, we conclude that GFP172 will serve well as a dual functional tag and have demonstrated its dual-functional ability by fusing this tag to the C-terminus of Maltose Binding Protein (MBP).

3.2. MATERIALS AND METHODS

3.2.1. Modification of mgfp5-ER

The mgfp5-ER insert was obtained by digestion of pBIN-mgfp5-ER with *Bam*HI-*Sac*I and ligated into the respective sites of Bluescript SK (Stratagene, La Jolla, CA). The sequence was modified through a number of polymerase chain reaction (PCR) amplifications using Platinum Pfx DNA Polymerase (Invitrogen, Carlsbad, CA) and oligonucleotides for site-directed mutagenesis and insertion of a poly-histidine sequence. For insertion of the poly-histidine sequence between amino acids 172 and 173 of the mgfp5-ER, a two-step PCR strategy was used. The N-terminal fragment (amino acids 1 –172) was amplified by primer A (T7 primer) and Primer H (5'- TTC GAT GTT GTG GCG GGT CTT G -3') (Figure 3.1). Similarly, a C-terminal fragment (amino acids 168-242 including the –HDEL terminal sequence) was amplified by primer E (5'-CGC CAC AAC ATC GAA **CAC CAT CAC CAT CAC CAT** GAC GGC GGC GTG CAA CTC GC -3') to incorporate a stretch of six

histidine residues (bold) and a region to overlap with the N-terminal fragment (underlined region which encodes amino acids 168-172) and primer J (T3 primer). The amplified fragments were purified by agarose gel electrophoresis and used together to amplify the entire mgfp5-ER using primers A (T7 primer) and J (T3 primer). An internal *Nco*I restriction site was removed by a silent mutation (changing a A to T) using the two-step PCR strategy. The N-terminal fragment was amplified using primers A (T7 primer) and F (5'- GTG TTG GCC AAG GAA CAG GTA -3') whereas the C-terminal fragment was amplified using primers C (5'- TAC CTG TTC CTT GGC CAA CAC -3') and J (T3 primer). The purified fragments were used together to amplify the entire mgfp5-ER using primers A (T7 primer) and J (T3 primer). Using the same two-step PCR strategy, serine in position 65 was changed to threonine (S65T). The N-terminal fragment was amplified using primers A (T7 primer) and G (5'- GAA CAC CAT AAG TGA AAG TAG TG -3') whereas the C-terminal fragment was amplified using primers D (5'- CAC TAC TTT CAC TTA TGG TGT TC -3') and J (T3 primer). The purified fragments were used together to amplify the entire mgfp5-ER using primers A (T7 primer) and J (T3 primer). One final amplification of the mgfp5-ER(S65T) insert was done using primer B (5'- GGC AGG AGG AAC **CAT GG**C TAG CAA AGG AGA AGA ACT TTT CAC TGG AG -3') to incorporate an N-terminal methionine (bold), an *Nco*I site (underlined) and removal of the *Arabidopsis thaliana* chitinase signal peptide and primer J (T3 primer). The amplified fragment was digested with *Nco*I-*Sac*I and ligated into the respective sites of pGEM-5Zf (Promega, Madison, WI). The product from translation of this modified gene will be referred to as GFP172. All modifications to the sequence were verified by DNA sequencing. A similar strategy was used to generate GFP157 by inserting a stretch of six histidine residues between Gln157 and Lys158.

The control, mGFP5-ER(S65T) with 6xHis in the C-terminus (GFPHis) was made using the primer B as sense and 5'-CGG GCA GAG CTC TTA **ATG GTG ATG GTG ATG GTG** AAG CTC ATC ATG TTT GTA TAG TTC- 3' as antisense primers to to retain SacI (underlined) and insert a 6xHis tag (bold).

3.2.2. MBP-GFP172 fusion protein

The gene coding for GFP172 was modified by PCR in order to fuse it to the C-terminus of Maltose Binding Protein. The *Nco* I site in the N-terminal of GFP172 was modified to to an *EcoR* I site (underlined) using 5'- CGG CCG AAT TCA GTA AAG GAG AAG AAC TTT TCA CTG GAG- 3' as sense and *Sac* I site was modified to a *Xba* 1 (underlined) using 5'-CGG GCA GAT CTA GAT TAA AGC TCA TCA TGT TTG TAT AG -3' as an antisense primer. The amplified PCR product was purified using GENECLEAN (BIO101 Systems, Carlsbad, CA) gene clean kit . The cleaned product was digested with *Eco*R I and *Xba* I and purified by agarose gel electrophoresis and ligated into the respective sites of pMAL–c2x vector (New England Biolabs, Beverly, MA). The product from translation of this modified gene will be referred to as MBP-GFP172.

3.2.3. Expression in E. coli

The modified mgfp5-ER(S65T) inserts were obtained by digestion with *Nco*I-*Sac*I and ligated into the respective sites of pET-21d vector (Novagen, Madison, WI). The resulting plasmids were transformed into *E. coli* BL21*trxB*(DE3) competent cells (Novagen, Madison, WI) following the manufacturer's instructions. Transformants containing the GFP constructs produced yellow-green colonies. Cultures were grown at the appropriate temperature (37°C or 28°C) in Luria-Bertani (LB) media containing 100 μg/ml ampicillin to an OD_{600} of approximately 0.6 and induced with 1 mM isopropyl-beta-D-thiogalactopyranoside (IPTG) for 4 hours. The cells were harvested

by centrifugation for 15 min at 3000xg at 4°C and resuspended in 50 mM NaH_2PO_4 (pH 7.0), 300 mM NaCl. Lysozyme was added to a final concentration of 0.75 mg/ml and the resuspended cells were incubated at room temperature for 30 min. The sample was sonicated 6 x 20 sec on ice at a minimum power setting using an ultrasonic cell disruptor equipped with a microprobe (Branson Model 250, Danbury, CT). The lysate was centrifuged for 20 min at 12,000xg at 4°C and soluble GFP in the supernatant was purified by immobilized metal affinity chromatography (IMAC) using TALON cobalt resin columns (Clontech, Palo Alto, CA) according to the manufacturer's instructions.

3.2.4. Protein assay

Protein was determined using the Micro BCA Protein Assay Reagent (Pierce Chemical Co., Rockford, IL) with bovine serum albumin or purified recombinant GFP protein (Clontech, Palo Alto, CA) as the standard.

3.2.5. Analysis of proteins

Protein samples were denatured in SDS sample buffer containing 2-mercaptoethanol for 3 min at 100°C and resolved on a 12% SDS-polyacrylamide gel and a 10% gel in the case of the fusion protein as described (Laemmli 1970). Gels were stained with Coomassie brilliant blue.

For comparison of soluble and insoluble protein following lysis of *E. coli*, a sample from the supernatant (cytosolic) was mixed with SDS loading sample buffer whereas a sample of the pellet (insoluble fraction) was dissolved in SDS loading sample buffer with volumes adjusted such that each lane in the gel represented 100 μl of original *E. coli* cell culture.

To observe fluorescence of the fusion protein, samples were made up in SDS sample buffer without mercaptoethanol and directly loaded on the gel without boiling.

3.2.6. Fluorescence measurements

GFP fluorescence was measured using a fluorescence spectrophotometer (Hitachi Model F-2500, Tokyo, Japan). Wavelength scans were performed on all GFP variants to determine excitation and emission peaks. Unless otherwise stated, the fluorescent intensity per unit protein was determined with the GFP samples diluted in GFP buffer (50 mM NaH_2PO_4 (pH 8.0), 10 mM Tris-HCl (pH8.0), 200 mM NaCl) with excitation set at 490 nm and emission set at 512 nm.

3.2.7. Percentage recovery of GFP by IMAC

The recovery of GFP by having polyhistidine in the loop was compared to having the His tag on the C-terminus. The *E coli* derived protein was quantified by BCA Protein Assay (Pierce, Rockford, IL). A total of 20 μg of protein was made up to 1 ml by the addition of Extraction buffer (50mM NaH_2PO_4, 300mM NaCl at pH 7.0) and fluorescence measured with excitation set at 490nm and emission at 512nm. The protein solution was allowed to pass through equilibrated TALON cobalt resin columns (Clontech, Palo Alto, CA) by gravity. The flow through was collected and its fluorescence was measured in the fluorescence spectrophotometer. After washing the column according to the manufacturer's instructions, the protein was eluted with 3 ml of elution buffer (50mM NaH_2PO_4, 300mM NaCl, 150 mM Imidazole at pH 7.0) and collected in 1ml fractions. Intensity measurements were made for the eluted samples as described earlier. The intensity measurements were used to estimate the amount of GFP present in the eluted fractions by standard curves of Fluorescence vs Amount of protein (μg/ml) made for the purified protein (data not shown).

3.2.8. pH Titration

Fluorescence intensities of GFP (20μg/ml) in a buffer (Kneen et al., 1998) containing 120mM KCl, 5mM NaCl, 0.5mM $CaCl_2$, 0.5mM $MgSO_4$, 10mM MES,

10mM MOPS and 10mM Citrate (Kneen et al., 1998) at pH values ranging from 8.0 to 5.0 differing by 0.5 pH units, was measured in a quartz cuvette. The pH was adjusted using 1 N HCl. Emission intensities were plotted as a function of pH and curve fitted as was done in the past (Kneen et al., 1998).

3.2.9. Tobacco extract and spiking studies

Tobacco leaves from wild-type potted plants was ground in liquid nitrogen with a mortar and pestle till a fine powder was obtained. The powder was resuspended in extraction buffer (50mM NaH_2PO_4, 300mM NaCl pH 7.0) and centrifuged at 10,000rpm for 10 minutes. The supernatant was collected and filtered using a Whatman 70mm filter paper to remove floating solid particles that remained after centrifugation. The total protein in the extract was quantified by the Biorad protein assay using Bovine Serum Albumin (BSA) as standard. 1mM Phenylmethyl sulfonyl fluoride (PMSF) was added to the extract to inhibit protease activity. The MBP-GFP172 fusion protein was spiked into this extract at a concentration representing 0.5% total soluble protein. The fusion protein was then purified using TALON (Clontech, Palo Alto, CA) cobalt resin columns and percentage recovery was calculated for the fusion protein as described earlier from a standard series of MBP-GFP172 purfied from *E coli* lysate.

3.2.10. Factor Xa Cleavage

Factor Xa (New England Biolabs, Beverly, MA) at concentrations of 200ng and 500ng was used to cut 20 μg of a 1μg/μl solution of MBP-GFP172 in elution buffer (50mM NaH_2PO_4, 300mM NaCl, 150mM Imidazole pH 7.0) at room temperature. 5 μl samples from this reaction mixture were collected at different time intervals to monitor cleavage by SDS-PAGE analysis.

3.3. RESULTS AND DISCUSSION

In this work, a 6xHis tag was inserted at positions 172-173 and 157-158 of mGFP5-ER(S65T) and the properties of these new GFPs were evaluated.

3.3.1 Effect of S65T

For the GFPs with the internal His tag, an increase in fluorescence intensities of about 1.4 times was seen, as a result of the S65T mutation, as shown in Figure 3.2. These GFPs were produced at 37°C. Cos-7 cells expressing GFP5(S65T) fluoresced 1.65 times higher than cells expressing GFP5 (Siemering et al., 1996).

3.3.2 Effect of temperature on GFP expression

Effect of temperature on GFP expression showed a significant difference in the amount of soluble GFP produced at 28°C and 37°C for the two loop GFPs. Expression of soluble GFP172 was lower at 37°C (Figure 3.3A, Lane 3) than at 28°C (Figure 3.3B, Lane 3). In the case of GFP157, most of the protein is insoluble at the two temperatures with almost no soluble protein at 37°C (Figure 3.3A, Lane 5). However, the GFP with the C-terminal His tag, expressed significant amounts of soluble protein at both temperatures. It has been shown in the past that proteins expressed at 28°C were much brighter than that expressed at 37°C in the case of wt-GFP and S65T and this was attributed to a permanent chromophore deformation at 37°C (Patterson et al.,1997). When expressed at 28°C, it was shown that the chromophores formed efficiently in many variants of GFP (Patterson et al., 1997). Contrary to this, it was shown that the post-translational oxidation to form the mature GFP chromophore in not the step responsible for the temperature sensitivity instead it's primarily due to mis-folding of GFP at elevated temperatures (Siemring et al., 1996). The group also found that His tag at the C-terminus has minimal effects on this folding property at higher temperatures. GFP5 was designed to be a thermostable

folding mutant with enhanced spectral characteristics (Siemering et al., 1996). In the case of GFP172 and GFP157, the lower temperature is favored for expression of soluble protein and this could be because of mis-folding at the elevated temperature. In the case of GFP157, the effect of the internal His tag in misfolding at the higher temperature is more prominent than for GFP172. GFP expressed in *Saccharomyces cerevisiae* were less fluorescent at high culture temperatures (Lim et al., 1995). Probably, by adding the Ser147 to Pro (S147P) mutation, this folding problem at higher temperature could be overcome. It was shown that when S147P was combined with S65T, the resulting double mutant emitted higher fluorescence than GFP with a single S65T mutation and could be used over a wide range of culturing temperatures (Kimata et al., 1997).

3.3.3. Fluorescence intensity

The fluorescence intensities of GFP172 produced at 28°C was 65% of GFPHis, shown in Figure 3.4, but was brighter than the GFP172 produced at 37°C. The intensity of GFP157 produced at 28°C was approximately the same as that of GFP172 at the same temperature. The intensities have been normalized with respect to GFPHis. It can be seen that irrespective of the position of His tag, the GFPs produced at 28°C were brighter. This is consistent with the explanation given earlier that chromophore formation was more efficient at 28°C and the fact that the fraction of completely folded GFP may be higher at the lower temperature. Moreover, the amount of light absorbed by GFP172 produced at 37°C as seen from the excitation peak in Figure 3.5 was 59% of GFPHis. Similarly, the excitation peak for GFP172 at 28°C is 63% of that absorbed by GFPHis (Figure 3.6). This reduction in the excitation peak could be because of a change in the microenvironment around the chromophore resulting in less molar absorbance and hence decrease in emission by similar amounts.

Affinity fluorescent proteins were made by Matsudaira et al (2001) by inserting an epitope for haemagglutinin at positions 172-173 and 157-158 of GFP. It was seen that the 395nm peaks were 40% and 33.3% respectively for the two positions, compared to wt-GFP. When excited at 395nm, the emission peaks at 550nm were 40% and 32.8% for 172-173 and 157-158 residues compared to wt-GFP. An interesting thing to note is that the 395nm peak, characteristic of a GFP5 variant is not seen in the excitation scan. This implies that the S65T mutation dominates and therefore, only a single 490nm peak is seen consistent with what was obtained earlier for S65T (Heim et al., 1995).

3.3.4. Purity and Recovery

Since the poly-histidine tag in the loop is accessible to the immobilized metal highly pure samples of GFP172 and GFP157 were obtained by IMAC as observed by SDS-PAGE (Figure 3.7). For the same reason, even the recovery obtained for GFP172 (87%) was comparable to that of having the tag in the C-terminus (81%) (Figure 3.8). No fluorescence was seen in the flow through and the washes indicating binding of the protein to the column and the minor loss of protein as seen in the recovery could be because of irreversible binding to the column.

3.3.5. Effect of pH

pH plays a significant role on GFP fluorescence as can be seen from the titration curve in Figure 3.9. The titration data were fitted to the following equation:

$$F = 1 / [1 + 10^{nH(pKa-pH)}] \quad (1)$$

with parameters pKa (pH at 50% maximum) and Hill coefficient nH (proportional to the slope of fluorescence versus pH at pKa) (Kneen et al 1998). Fitted pKa's were found to be 5.6, 5.45 and 5.39 for GFP172, GFP157 and mGFP5ER-6xHis. Past reports have shown that the pKa for S65T was ~ 6.0 (Elsiger et al., 1999, Kneen et al

1998). The difference in the pKa that we obtained can be attributed to the mGFP5-ER that we used as a template to insert the His tag which has the I167T, V163A and S175G mutations (Heim et al., 1994; Siemering et al., 1996). Single point mutations have known to change pKa values drastically. For example, mutagenesis was used to increase the pKa of S65T from 6.0 to 7.8 (S65T/H148D) presumably due to the proximity of the negatively charged aspartate (Elsiger et al., 1999). EGFP (F64L/S65T) is 50% quenched at approximately pH 5.5 (Patterson et al., 1998).

It can be clearly seen from the pKa values that we obtained that the position of the His tag in the GFP has minimal effects on the GFP pKa. Kneen and coworkers (1998) have shown by spectroscopic and kinetic studies that the fluorescence of GFP decreased because of a change in the molar absorbance as a result of simple protonation-deprotonation of residues for pH > 5.0 whereas below pH 5.0 the GFP unfolds. We obtained similar results where the shape of the spectra did not change with pH but there was a decrease in molar absorbance (data not shown). Crystallographic studies have shown that the phenolic hydroxyl of the chromophore is the titrating group responsible for the pH sensitive mechanism of S65T (Elsiger et al., 1999; Kneen et al., 1998) and protonation state of the imidazolinone ring nitrogen unchanged (Elsiger et al., 1999). Use of GFP as an intracellular pH indicator has been demonstrated (Llopis et al., 1998; Kneen et al., 1998). Our pH titration experiments will be useful to make appropriate corrections and calibrations in fluorescence values depending on the organelle/cytosol where the dual-functional tag fused to a target protein is produced which in turn will depend on the N- and/or C-terminal targeting and retention signals. Moreover, depending on the pH environment of where the fusion protein is expressed, appropriate GFP mutants with desirable pKa values can be used as template for the insertion of His tag at the positions we have chosen with

the assumption that the tag will not have drastic effects on the titrating group of the chromophore.

3.3.6. MBP-GFP172 fusion protein

Based on these preliminary experiments, we fused GFP172 to Maltose Binding Protein (MBP) and expressed it in *E coli* BL21(DE3) trxB cells (Novagen, Madison, WI). Though GFP172 is slightly lower in intensity than GFP157 (Figure 3.4), the fact that GFP172 is soluble at both temperatures (Figure 3.3) unlike GFP157, resulted in our choosing GFP172 as the fusion tag. The ability to remain soluble at both 28ºC and 37ºC enables the usage of the tag in a broad host range including mammalian cells, plant cell cells, yeast and *E coli.* Lots of applications of fusion proteins have been reported in the past. In essence, fusion proteins have been made, either for providing affinity purification tags or for enhancing the solubility of the fusion partner and for suppressing degradation. The solubilizing ability of commonly used fusion partners like maltose-binding protein (MBP), glutathione S-transferase (GST) and thioredoxin (TRX) were compared by fusing six diverse proteins that normally accumulate in insoluble form and it was found that MBP was the most effective fusion partner (Kapust and Waugh, 1999). MBP can also be used to purify the fusion partner by immobilized amylose resin affinity chromatography (Riggs, 1990). An MBP-GFPuv fusion protein was made to study protein localization with and without the MBP signal sequence using GFPuv as a reporter (Feilmeier et al., 2000). It was found that the fusion protein localized to the cytoplasm fluoresced but that localized to the periplasmic space did not. Solubility and folding of the fusion partner is also influenced by the position of MBP in the fusion as was demonstrated with Procathepsin D and pepsinogen, both being proteins that normally form inclusion bodies. However, MBP in the N-terminus resulted in the fusion protein being soluble

in the bacterial cytosol and was able to bind to amylose resin, but formed inclusion bodies if fused to the C-terminus (Sachdev and Chirgwin, 1998).

We fused the GFP172 tag to the C-terminus of Maltose Binding Protein and expressed it in *E coli* BL21(DE3) trxB cells (Novagen, Madison, WI) in soluble form to test the ability of the GFP tag to serve its dual functions as a fusion protein. Figure 3.10 (Lane 2) shows that the GFP172 with the internal 6xHis tag can be used to purify the fusion protein to near homogenous purity in a single step from a crude *E coli* lysate. This implies that the 42.5 kDa Maltose Binding Protein attached to GFP172, does not interfere with the accessibility of the internal 6xHis tag to the immobilized metal. As mentioned earlier, this accessibility is because of the presence of the 6xHis tag in the solvent accessible loop of GFP. The purified fusion protein was soluble and fluoresced which was determined by running a SDS-PAGE gel (Figure 3.11) where the samples were loaded without boiling so that the fusion protein maintains its native state. It has been established that GFP fluoresces even in the presence of 1% SDS (Bokman and Ward, 1981) and GFP-S65T does not get quenched in the presence of acrylamide (Kneen et al., 1998). Since the GFP fluoresces even as a fusion, it can be concluded that the MBP is also active because full length protein is produced as can be seen from the size of 72.5 Kda on the gel (Figure 3.10) and the fact MBP is fused to the N-terminus of GFP172. As mentioned earlier, position of MBP in the fusion influences the solubilty of the fusion partner. Moreover, MBP-GFP (Kapust and Waugh, 1999) and MBP-GFPuv (Feilmeier B.J et al., 2000) fusion proteins were reported to be soluble.

3.3.7. Spiking studies

In order to test, if the GFP tag would be able to purify the fusion protein from plant extracts, we spiked MBP-GFP172 into tobacco leaf extract and analyzed the

ability of the tag to recover the fusion protein. We spiked at a concentration of 0.5% total soluble protein which is representative of the level of expression achieved for GFP in plants from past work. Expression of mGFP5-ER driven by Cauliflower Mosaic Virus (CaMV) 35S promoter ranged upto 7.05 μg GFP per mg of extractable protein (0.7%) (Remans et al., 1999). GFP expression levels in tobacco ranging from 0.0-0.5% of total extractable protein has been reported and a minimal amount of approximately 0.1% GFP was required for unambiguous macroscopic detection (Leffel et al., 1997). Tobacco plants grown under artificial light driven by CaMV 35S promoter was analyzed for GFP expression patterns and it was found that GFP synthesis levels varied between 0.12% to 0.15% in leaves (Harper and Stewart, 2000).

In our spiking and recovery experiments we found that the average recovery was about 75% and purity as judged by SDS-PAGE was greater than 75%, determined by scanning the Coomassie stained gel (Figure 3.12, Lane 2) and densitometry analysis using SigmaScan Pro 5.0 image-analysis software (SPSS Inc., Chicago, IL). The recovery was calculated as described in Materials and Methods from a standard series of the fusion protein purified from *E coli* lysate (Figure 3.13). Such a methodology has been used in the past wherein a GFP-standard series was prepared for purified GFP in untransformed plant extract to determine GFP concentration in transgenic plant extracts (Remans T et al., 1999). Lactate dehydrogenase (LDH) was purified from tobacco extract using a 6xHis tail at the N-terminus by IMAC with a Zn^{2+} chelated gel (Mejare M et al., 1998). Enzmye recovery of 55% was obtained for $LDHHis_6$ compared to 7% for native LDH having no His tail. It was also reported that native LDH did not bind to Co^{2+} similar to results obtained in another work where native β-glucuronidase (GUS) did not bind to Co^{2+} chelated to imnodiacetate (IDA) (Zhang et al., 2000). However, β-glucuronidase-his_6

(GUSH6) was purified from canola protein extract with almost homogenous purity in a single chromatographic step using Co^{2+} with iminodiacetate (IDA) as the chelating ligand. It was found that Co^{2+}-IDA had the least amount of non-specific binding of canola proteins than metals like Cu^{2+}, Ni^{2+} and Zn^{2+} and therefore higher purity of the target protein (Zhang et al., 2000). This highlights the importance of the 6xHis tag in the performance of IMAC compared to the native protein. Just binding to the immobilized metal is not sufficient. There needs to be a preferential binding of the target protein with respect to the numerous other proteins found in the host system. For the same reasons, using Cu(II) to purify GFPuv as was reported by Li et al., (2001) without a His tail may not be feasible in most cases for reasons of non-specific binding of host proteins because Cu(II) has the ability to recognize surface histidines. Moreover, the Cu(II) system requires high amounts of NaCl to work effectively (Li et al., 2001) and that in turn would result in hydrophobic interactions causing more non-specific binding. The Cu(II) system would also be ineffective if there was a His tail on the target protein because the binding of the target protein would be too strong requiring harsh conditions for elution which may disrupt the target protein and therefore lower recoveries as was seen in GUSH6 recovery on Cu^{2+}-IDA (Zhang et al., 2000).

From the gel in Figure 3.12, we can see that the purity of the fusion protein (~71kda) is quite high. However, a lower band having the size of approximately 32 Kda is seen in the lane and this is seen inspite of adding a protease inhibitor as mentioned in the Materials and Methods. Since no other significant band apart from that of the fusion protein and the lower band is seen, leads us to believe that the lower band is either a product of protease activity on the fusion protein or is a protein from the tobacco extract. The PMSF may not have been effective enough to inactivate

protease activity to completion. A western blot using anti-GFP antibody on this fraction (Figure 3.14) does show this lower band and therefore rules out the second possibility of the band being a protein from the extract. Since the size of this lower band is approximately the size of GFP, it was necessary to determine if this band is responsible for any fluorescence. Since our recovery data is obtained from a standard series of the fusion protein, fluorescence if any contributed by the lower band will distort the values obtained for recovery. We ran samples on SDS-PAGE without boiling the samples so that the protein retains its fluorescence. Lane 3 in Figure 3.15 represents protein recovered from an extract with no PMSF added and Lanes 4 & 5 show protein recovered from extract with 1mM PMSF. No fluorescent band is seen around the size of the GFP172 standard in Lanes 3,4 or 5 which implies that, there is no contribution to the fluorescence of the eluted fraction by the lower band. Based on densitometry analysis using SigmaScan Pro 5.0 image-analysis software (SPSS Inc., Chicago, IL) and comparing with intensity of fusion protein standards on the Coomassie stained gel (Figure 3.12), the lower bands in Lanes 4 and 5 is estimated to have atleast 520ng and 780ng respectively. However, none or negligible amounts of this protein fluoresces by comparing to the bands in Lanes 7, 8 and 9 which have different concentrations of purified GFP172 with the lowest being 100ng in Lane 7. Since no band in Lane 3, 4 or 5 is seen around the size of GFP172, we can conclude that majority of the protein in the lower band is non-fluorescent. The amount of fluorescent protein if any would be far less than 100ng and since Lanes 3 and 4 have 2μg of total protein and Lane 5 has 3 μg total protein, the error would be far less than 5%. Therefore, the fluorescence of the recovered fractions is a good estimate of the amount of fusion protein recovered. The cleavage of the fusion protein would depend on the recombinant protein of interest fused to the tag, the linker and the host system

used for expression. Therefore, the use of protease inhibitors needs to be optimized on a case-by-case basis.

3.3.8. Factor Xa cleavage

In a test run to determine Factor Xa (New England Biolabs, Beverly, MA) cleavage, the enzyme was added in a ratio of 1:100 of Factor Xa to fusion protein. As can be seen from the gel in Figure 3.16, even after 40 hrs, the cleavage of the fusion protein into two distinct bands is not complete. This is probably because the cleavage site is not completely exposed to cleave the fusion protein to completion at the recommended test concentration of Factor Xa. On raising the Factor Xa concentration by 2.5 times, complete cleavage was achieved. Factor Xa concentration cannot be increased to a great extent because secondary cleavage products would be produced. Therefore, an optimum concentration of Factor Xa needs needs to be arrived at, for each fusion protein with Ile-Glu/Asp-Gly-Arg Factor Xa recognition site in between the two proteins. Factor Xa cleaves after Arg and sometimes at other basic residues, depending on the conformation of the protein substrate (Nagai et. al, 1984; Eaton et al 1986). Factor Xa can be removed from the reaction mixture by passing through benzamidine sepharose resin, a product of Amersham pharmacia biotech.

3.4. CONCLUSIONS

A novel dual functional GFP tag was developed which can be effective in different host systems. This multifunctional GFP tag can be used to monitor expression of any target protein fused to it and can also monitor purification and recovery by IMAC, by measuring fluorescence of the fractions. Once the fusion protein is purified, the dual functional tag could be removed from the protein of interest by appropriate choice of linkers.

It has been demonstrated that by simply measuring GFP fluorescence in *E coli* cultures, amounts of active Organophosphorous hydrolase (OPH) in the culture could be estimated because GFP was fused to the N-terminus of OPH (Wu et al., 2000). The GFP fluorescence was also used to track OPH purification by a 6xHis tag on the N-terminus of GFP. In a more recent finding, GFP fused to the N-terminus of μ-opioid receptor (HuMOR) was used to quantify HuMOR expression levels and the fusion did not effect the expression levels and ligand binding properties of the functional receptor (Sarramegna et. al, 2002). However, by having an internal 6xHis tag in the GFP as demonstrated in this work, more flexibility is provided in terms of attaching signal peptides and retention signals which has the advantage of better accumulation and stability of the recombinant protein. Commercial vectors can be developed to incorporate these features and recombinant proteins could be expressed and monitored in different host organisms and also purified by IMAC.

REFERENCES

Abedi MR, Caponigro G and Kamb A (1998) Green fluorescent protein as a scaffold for intracellular presentation of peptides. Nucleic Acids Resch. 26: 623-630

Bokman SH, Ward WW (1981) Renaturation of Aequorea green fluorescent protein. Biochem. Biophys. Res. Commun.101: 1372-1380

Chalfie M, Tu Y, Euskirchen G, Ward WW, Prasher DC (1994) Green fluorescent protein as a marker for gene expression. Science 263: 802-805

Chiu W, Niwa Y, Zeng W, Hirano T, Kobayashi H, Sheen J (1996) Engineered GFP as a vital reporter in plants. Curr Biol. 6:325-30

Cubitt AB, Heim R, Adams SR, Boyd AE, Gross LA and Tsien RY (1995) Understanding, improving and using green fluorescent proteins. Trends in Biochemical Sciences 20: 448-455

Doi N and Yanagawa H (1999) Design of generic biosensors based on green fluorescent proteins with allosteric sites by directed evolution. FEBS Lett. 453: 305-307

Eaton D, Rodriguez H, Vehar GA (1986) Proteolytic processing of human factor VIII. Correlation of specific cleavages by thrombin, factor Xa, and activated protein C with activation and inactivation of factor VIII coagulant activity. Biochem. 25:505-12

Elsliger MA, Wachter RM, Hanson GT, Karen Kallio and Remington SJ (1999) Structural and spectral response of Green Fluorescent protein variants to changes in pH. Biochem.38:5296-5301

Feilmeier BJ, Iseminger G, Schroeder D, Webber H, Phillips GJ (2000) Green fluorescent protein functions as a reporter for protein localization in *Escherichia coli.* J Bacteriol. 4068-4076

Harper BK and Stewart CN Jr. (2000) Patterns of Green fluorescent protein expression in transgenic plants. Plant Molec. Biol. Rep.18: 141a-141i

Haseloff J, Siemering KR, Prasher DC and Hodge S (1997) Removal of a cryptic intron and subcellular localization of green fluorescent protein are required to mark transgenic Arabidopsis plants brightly. Proc. Natl. Acad. Sci. USA 94: 2122-2127

Heim R, Prasher, DC and Tsien RY (1994) Wavelength mutations and posttranslational autooxidation of green fluorescent protein. Proc. Natl. Acad. Sci. USA 91:12501-12504

Heim R, Cubbit AB and Tsien RY (1995) Improved green fluorescence. Nature 373: 663-664

Kapust RB and Waugh DS (1999) *Escherichia coli* maltose-binding protein is uncommonly effective at promoting the solibility of polypeptides to which it is fused. Protein Sci.1668-1674

Kimata Y, Iwaki M, Lim C R, Kohno K (1997) A novel mutation that enhances the fluorescence of green fluorescent protein at high temperatures. Biochem Biophys Res Commun 232: 69-73.

Kneen M, Farinas J, Li Y, and Verkman AS (1998) Green fluorescent protein as a noninvasive intracellular pH indicator. Biophys. J. 74: 1591-1599

Laemmli UK (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature 227:680-5

Leffel SM, Mabon SA, Stewart CN Jr (1997) Applications of green fluorescent protein in plants. Biotechniques 23:912-8

Li Y, Agrawal A, Sakon J, Beitle RR (2001) Characterization of metal affinity of green fluorescent protein and its purification through salt promoted, immobolized metal affinity chromatography. J of Chromatgr. A 909: 183-190

Lim CR, Kimata Y, Oka M, Nomaguchi K, Kohno K (1995) Thermosensitivity of green fluorescent protein utilized to reveal novel nuclear-like compartments ina mutant nucleophorin NSP1. J Biochemistry (Tokyo) 118:13-17

Llopsis J, McCaffery JM, Miyawaki A, Farquhar MG and Tsien RY.(1998) Measurement of cytosolic , mitochondrial, and Golgi pH in single living cells with green fluorescent proteins Proc. Natl. Acad. Sci. USA 95: 6803-6808

Matsudaira P.T, Ehrlich DJ, Zhong Q and Freyson Y (2001) Affinity fluorescent proteins and uses thereof. International Patent, Publication number WO 01/09177 A2

Mejàre M, Lilius G, Bulow L (1998) Evaluation of genetically attached histidine affinity tails for purification of lactate dehydrogenase from transgenic tobacco. Plant Science 134: 103 –114

Nagai K, Thogersen HC (1984) Generation of beta-globin by sequence-specific proteolysis of a hybrid protein produced in *Escherichia coli.* Nature 309:810-2

Niwa Y, Hirano T, Yoshimoto K, Shimizu M and Kobayashi H (1999) Non-invasive quantitative detection and applications of S65T-type green fluorescent protein in living plants. Plant J. 18: 455-463

Patterson GH, Knobel SM, Sharif WD, Kain SR and Piston DW (1997) Use of the Green Fluorescent Protein and its mutants in quantitative fluorescence microscopy. Biophy. J 73: 2782-2790

Remans T, Schenk PM, Manners JM, Grof CPL, Elliott AR (1999) A protocol for the fluorimetric quantification of mGFP5-ER and sGFP(S65T) in transgenic plants. Plant Molecular Biology Reporter 17: 385-395

Riggs P, in "Current Protocols in Molecular Biology" (Ausubel FM et al., 1990) p.16.4.1/16.6.1. Greene Associates/Wiley Interscience, New York

Sachdev D and Chirgwin JM (1998) Order of fusions between bacterial and mammalian proteins can determine solubility in *Escherichia coli* Biochem. Biophys. Res. Commun. 244, 933

Sarramegna V, Talmont F, de Roch S, Milon A, Demange P (2002) Green fluorescent protein as a reporter of human μ-opioi receptor overexpression and localization in the methylotrophic yeast *Pichia pastoris.* J of Biotech. 99: 23-39

Siemering KR, Golbik R, Sever R and Haseloff J (1996) Mutations that suppress the thermosensitivity of green fluorescent protein. Curr. Biol. 6:1653-1663

Tsien RY (1998) The Green Fluorescent Protein. Ann. Rev. Biochem.67: 509-544

Wu CF, Cha HJ, Rao G, Valdes JJ, Bentley WE (2000)A green fluorescent protein fusion strategy for monitoring the expression, cellular location, and separation of biologically active organophosphorus hydrolase. Appl Microbiol Biotechnol 54:78-83

Zhang CM, Resllewic SA, Glatz CE (2000) Suitability of immobilized metal affinity chromatography for protein purification from canola. Biotech.Bioeng. 68: 52-58

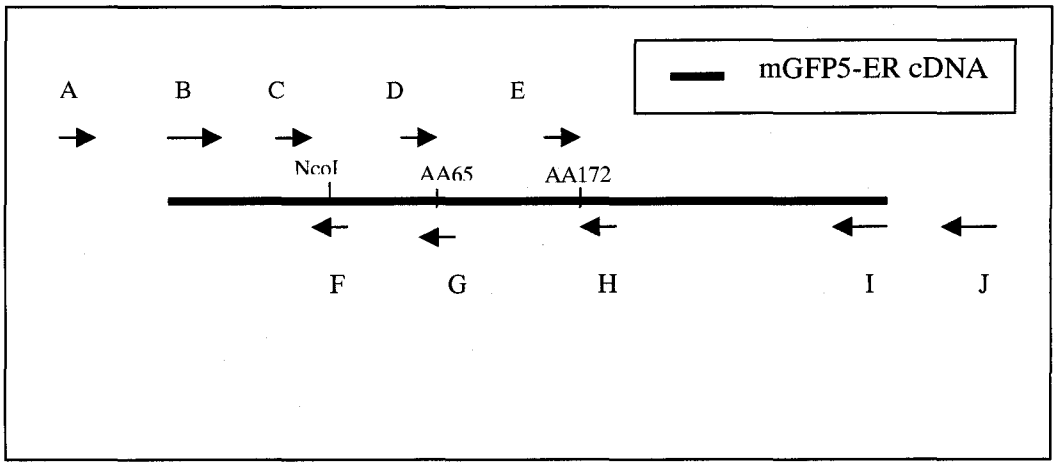


Figure 3.1: Primers used for vector construction

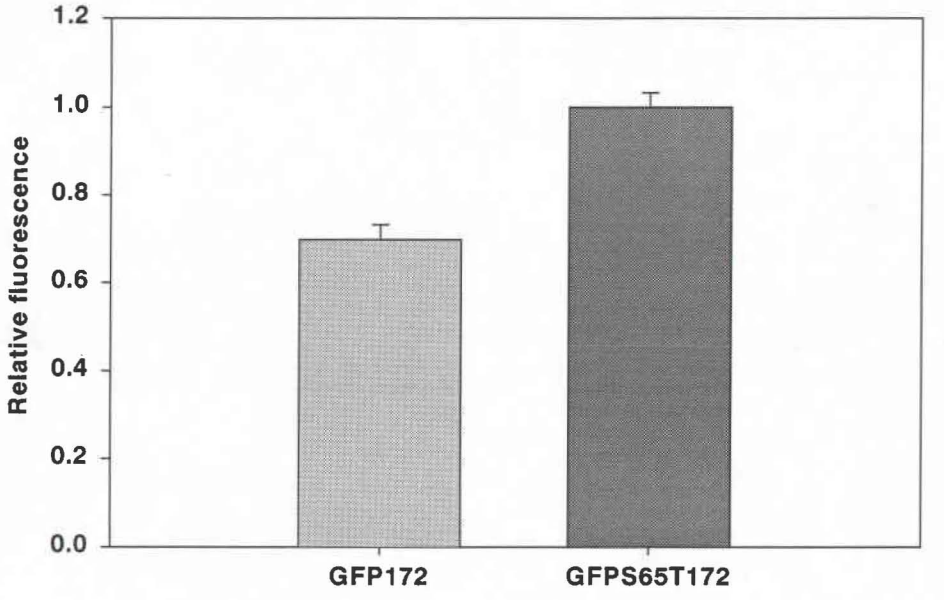


Figure 3.2: The mutation S65T increases fluorescence intensity of GFP. Fluorescence intensity readings were measured for 5μg/ml of GFP in GFP extraction buffer. Intensities were measured with excitation set at 467 nm and emission set at 505 nm for GFP172 and 490 nm excitation and 505 nm emission for GFPS65T172. Error bars denote standard deviations for measurements made in duplicate.

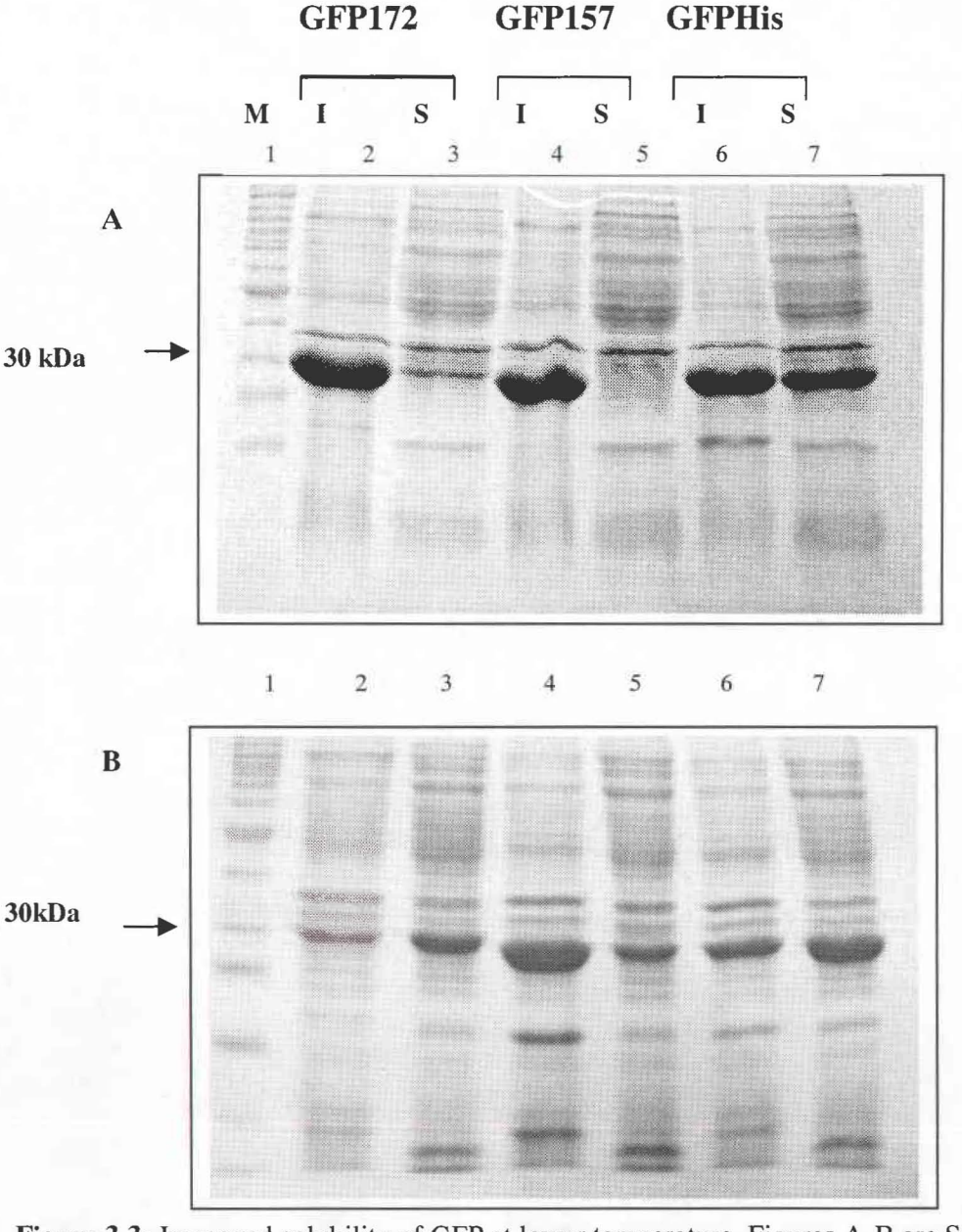


Figure 3.3: Improved solubility of GFP at lower temperature. Figures A-B are SDS-PAGE gels comparing the solubilities of the GFP variants (with the S65T mutation) during expression in bacterial cells at 37°C, shown in Figure 3.3A, and 28ºC, shown in Figure 3.3B. Samples equivalent to 100 µl of cell culture were loaded on a 12% polyacrylamide gel. I = Insoluble fraction; S = Soluble fraction.

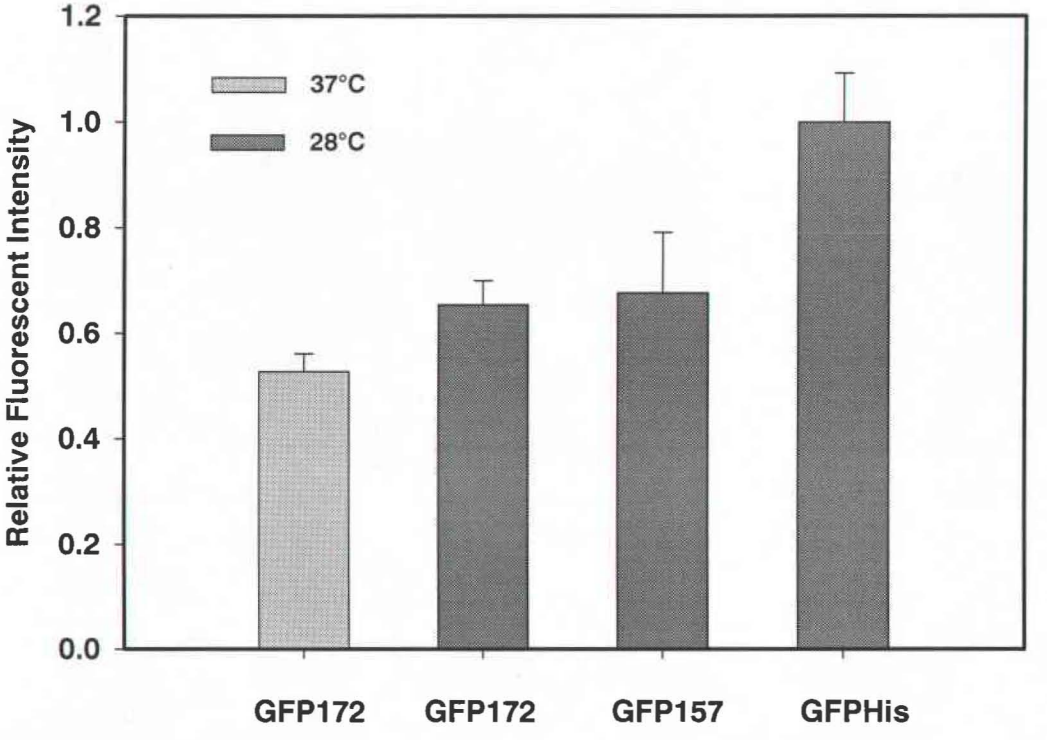


Figure 3.4 : GFP intensity is higher when produced at 28°C. Fluorescence intensity readings were taken for 5μg/ml of GFP in GFP extraction buffer. Readings were taken by setting the excitation at 490 nm and emission at 512 nm. Each intensity reading is an average of 4 replicates with each replicate reading an average intensity measured at time intervals of 10 seconds, 30 seconds, and 60 seconds after placing the cuvette in the F-2500 fluorescent spectrophotometer. Error bars denote standard deviations for measurements that were repeated for a total of 4 times.

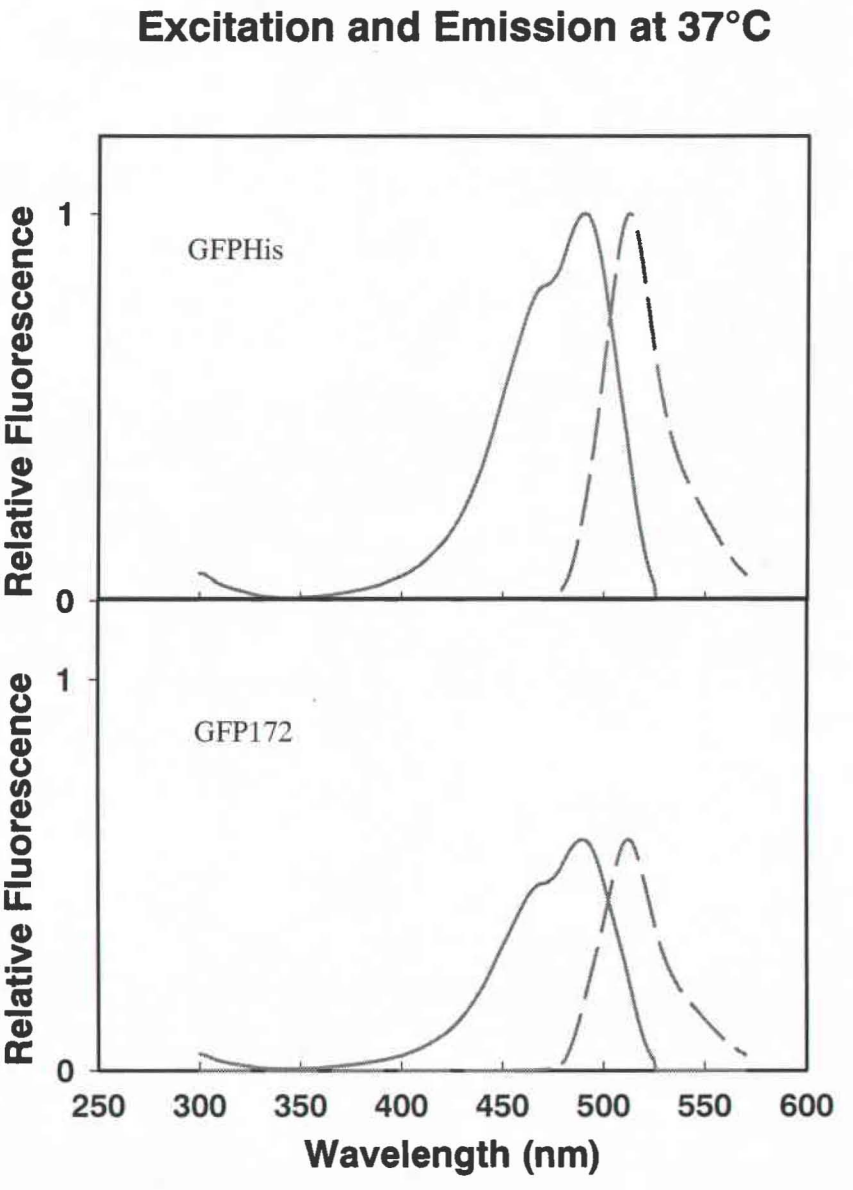


Figure 3.5: Excitation and Emission spectra at 37°C. Protein concentrations were 5 μg/ml in GFP extraction buffer (pH 8.0) and samples were allowed to stabilize for 1 minute after mixing before the scan was performed. All data points of the spectra have been normalized based on the peak height of the GFPHis variant.

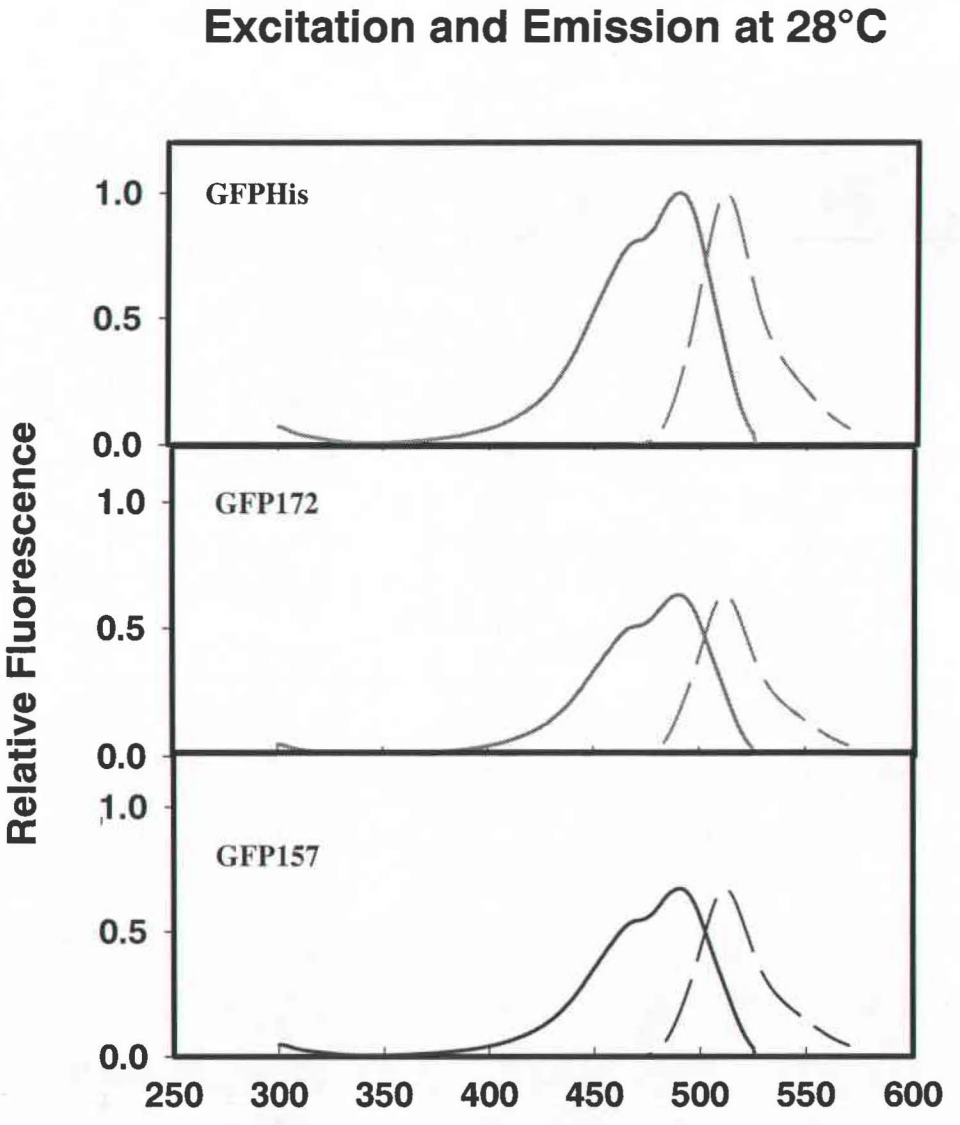


Figure 3.6: Excitation and Emission spectra at 28°C. Protein concentrations were 5 μg/ml in GFP extraction buffer (pH 8.0) and samples were allowed to stabilize for 1 minute after mixing before the scan was performed. All data points of the spectra have been normalized based on the peak height of the GFPHis variant

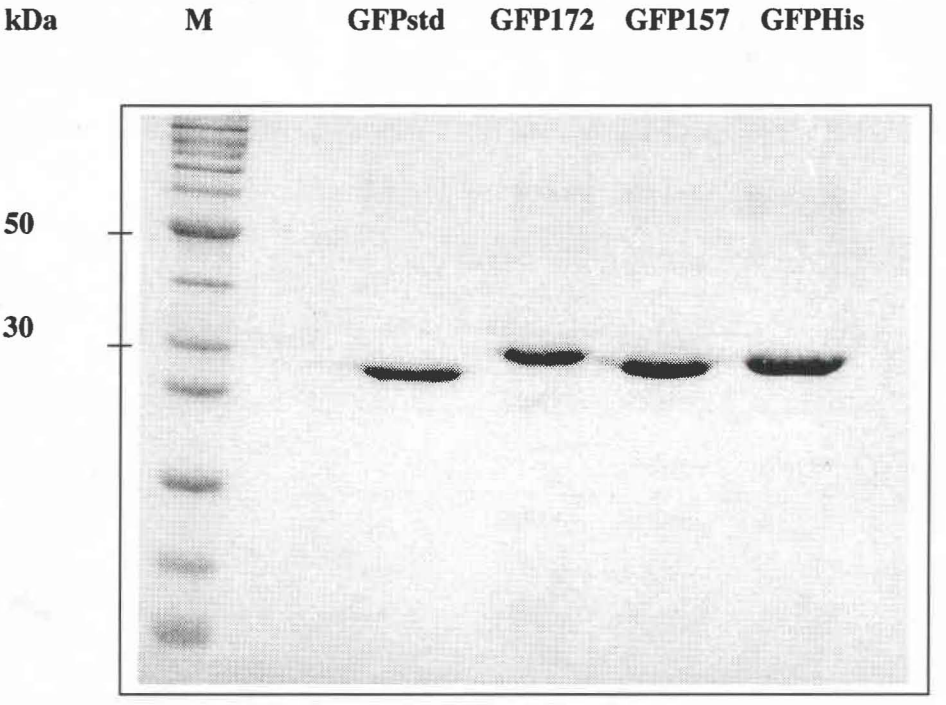


Figure 3.7: GFP with internal His tags can be purified from *E coli* lysate. The figure depicts an SDS-PAGE gel showing protein samples expressed in *E. coli* and purified by immobilized metal affinity chromatography (IMAC). Proteins were quantified by BCA assay using recombinant GFP (Clontech, Palo Alto, CA) as standard. A total of 2 μg protein was loaded on a 12% polyacrylamide gel and stained with Coomassie Brilliant Blue. M and Std denotes the protein ladder (Invitrogen, Carlsbad, CA) and a GFP standard without 6xHis tag (Clontech, Palo alto, CA), respectively.

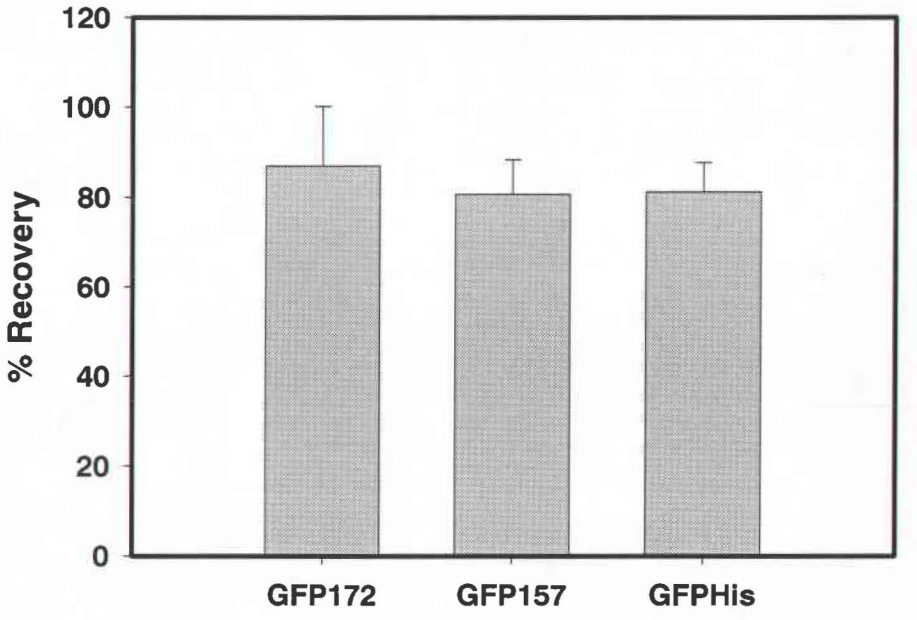


Figure 3.8: The GFP variants can be recovered from immobilized metal affinity columns. A total of 20 μg of GFP172, GFP157, or GFPHis was spiked into extraction buffer and protein was recovered using IMAC. Protein was eluted in 1 ml fractions to a total of 3 ml. Fluorescence of each eluted fraction was measured by setting excitation at 490 nm and emission at 512 nm. GFP fluorescence was highly linear in the 0.5- 20 μg/ml range in extraction buffer and there was hardly any change in linearity when fluorescence was measured in elution buffer. The fluorescence of eluted sample gives a measure of GFP recovered by 6xHis tag metal affinity chromatography. Error bars denote standard deviations for measurements made in duplicate.

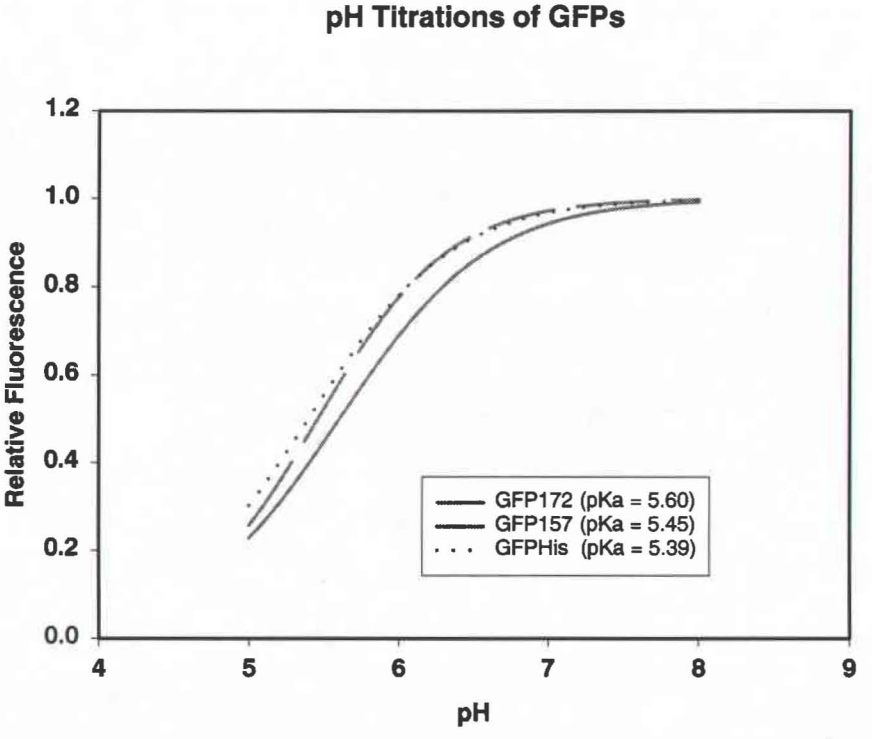


Figure 3.9: Fluorescence of GFP decreases with pH. Intensity readings were made for 5μg/ml of GFP variants, as noted, (produced at 28°C) at 10 seconds, 30 seconds, and 60 seconds and averaged. The averaged values were normalized based on the fluorescence of GFPHis at pH 8.0. The data was curve fitted as described before.

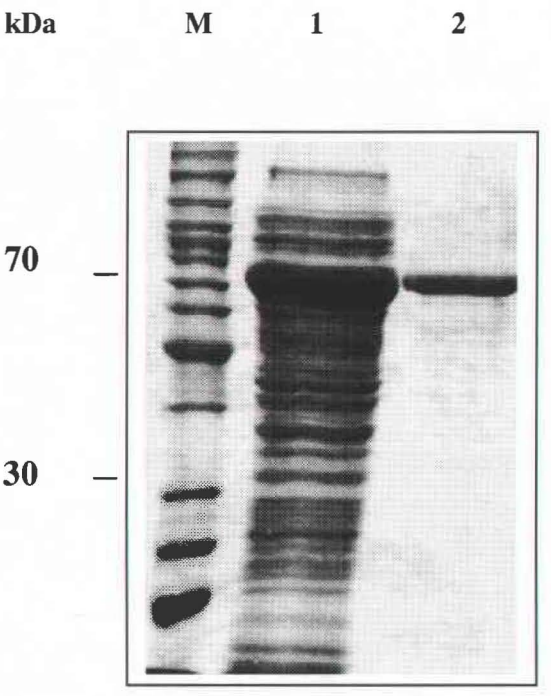


Figure 3.10: MBP-GFP172 purified from *E coli* lysate. M: Benchmark Protein Ladder (Invitrogen , Carlsbad, CA); Lane 1: *E. coli* lysate equivalent to 100µl cell culture; Lane 2*:* fusion protein purified by IMAC (3.5µg).

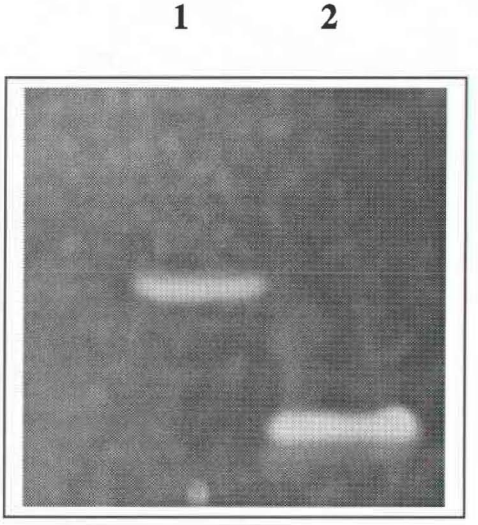


Figure 3.11: MBP-GFP172 fusion protein fluoresces. Samples were loaded without boiling on 10% SDS-PAGE gels. Lane 1: 10 μg of MBP-GFP172; Lane 2: 4 μg of GFP172.

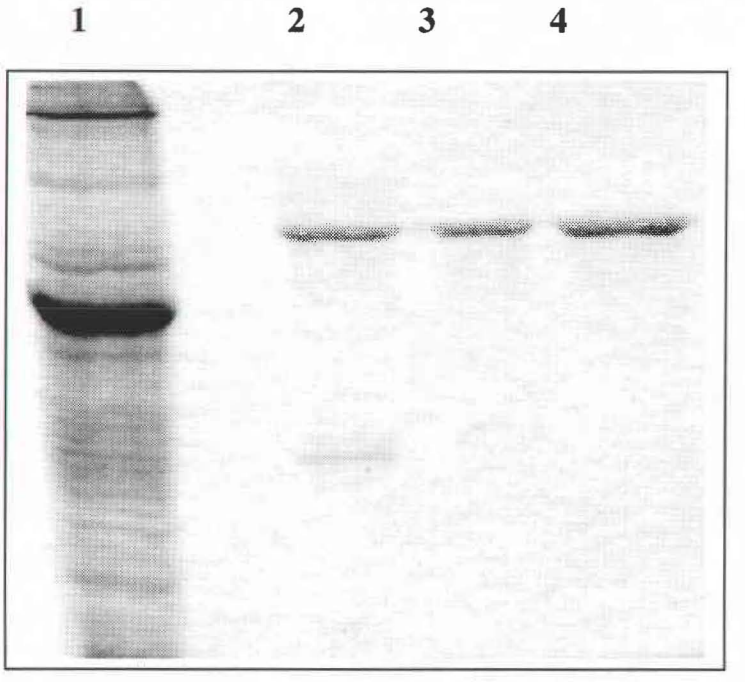


Figure 3.12: MBP-GFP172 purified from tobacco leaf extract. The figure shows a Coomassie stained gel showing the purification of the MBP-GFP172 fusion protein from tobacco extract using IMAC. Lane 1: 100μg of soluble proteins from tobacco extract; Lane 2: 3μg of eluted fraction having the highest fluorescence; Lane 3: 2μg of fusion protein purified from *E. coli*; Lane 4:3.0 μg of fusion protein purified from *E. coli.*

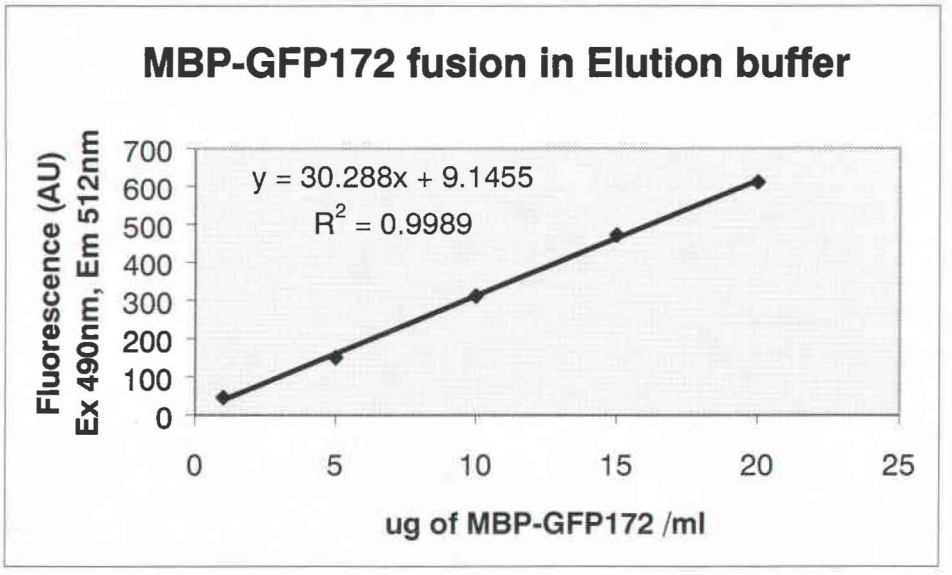


Figure 3.13: Fluorescence of fusion protein is linear with amount of protein. A standard curve of MBP-GFP172 in elution buffer (50mM NaH2PO4, 300mM NaCl, 150 mM imidazole pH 7.0) was made. Fusion protein purified from *E. coli* was made up in elution buffer at different concentrations and fluorescence measured with excitation set at 490 nm and emission set at 512 nm.

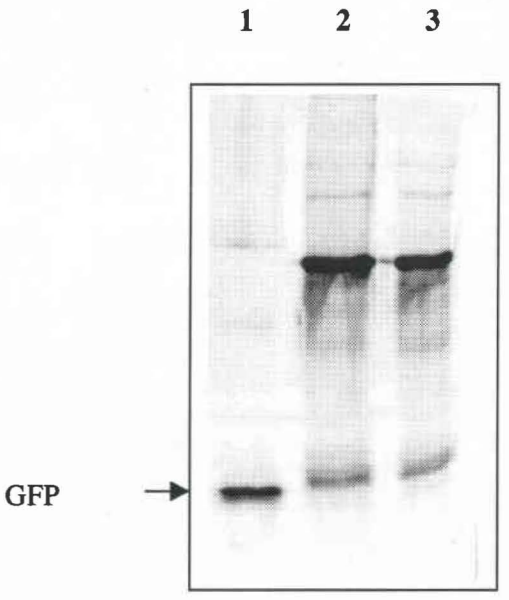


Figure 3.14: Cleavage of fusion protein inspite of adding PMSF. Fusion protein was spiked into tobacco leaf extract and purified by IMAC. A western blot was done using anti-GFP rabbit polyclonal antibody (1:2000) as the primary antibody and alkaline phosphatase conjugated goat anti-rabbit IgG was used as the secondary antibody on the most concentrated fraction that was eluted after IMAC. Lane 1: 50ng of GFP standard (Clontech, Palo Alto, CA); Lane 2: 500ng total protein without PMSF in the extraction buffer; Lane 3: 500ng total protein with 1mM PMSF in the extraction buffer

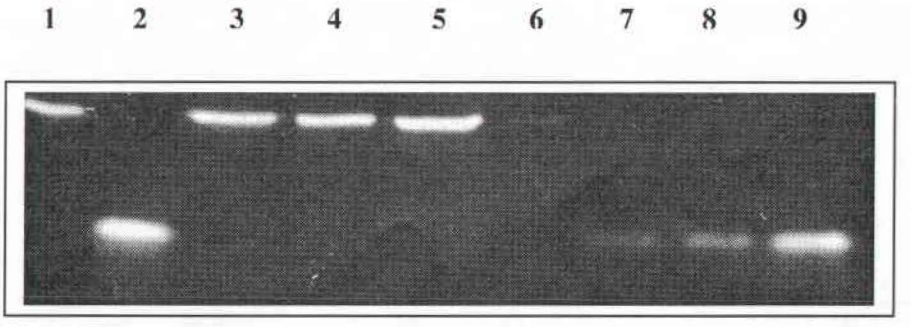


Figure 3.15: GFP detection on SDS-PAGE by fluorescence. Lane 1: 2μg fusion protein purified from *E. coli*; Lane 2: 2μg GFP172; Lane 3: 2μg of MBP-GFP172 purified from tobacco extract (with no phenylmethyl sulfonyl fluoride (PMSF); Lane 4: 2μg of MBP-GFP172 purified from tobacco extract (with 1mM PMSF); Lane 5: 3μg of MBP-GFP172 purified from tobacco extract (with 1mM PMSF); Lane 6: 100ng of MBP-GFP172; Lane 7: 100ng of GFP172; Lane 8: 500ng of GFP172; Lane 9: 1μg of GFP172.

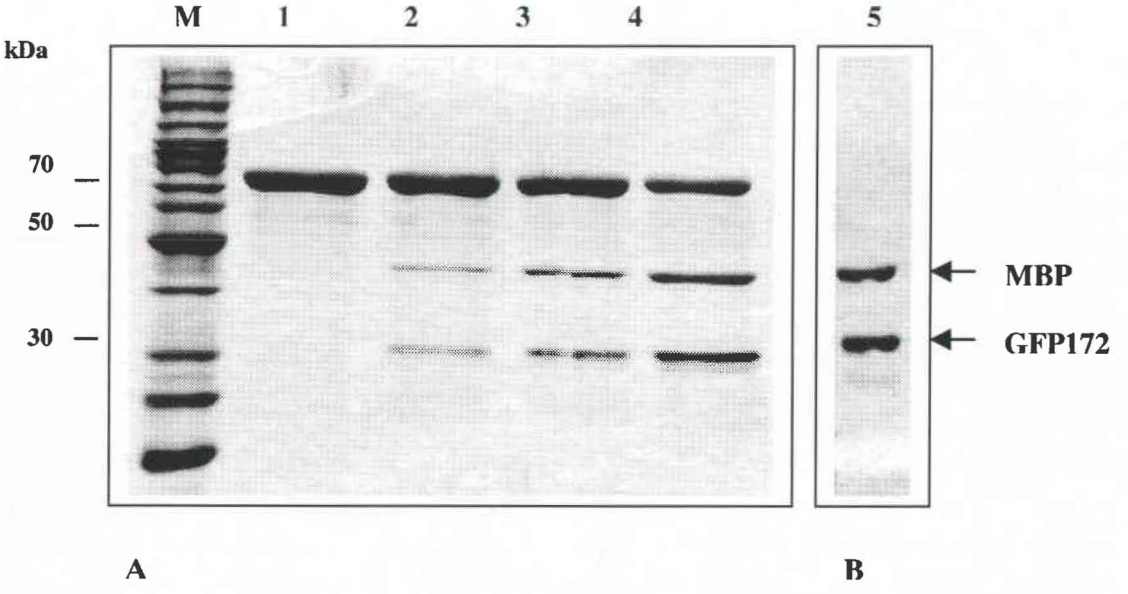


Figure 3.16: The GFP tag can be separated from the protein of interest by Factor Xa cleavage. Figure depicts Factor Xa cleavage of MBP-GFP172 fusion protein at different concentrations of Factor Xa. In each, 5µl (having a concentration of 1µg/µl) samples were taken at different time intervals. Figure 15A shows results using 200ng Factor Xa per 20μg fusion protein; M=Benchmark protein ladder; Lane 1: uncut fusion (0 hours); Lane 2: 6 hours; Lane 3: 24 hours; Lane 4: 40 hours. Figure 15B shows results using 500ng Factor Xa per 20μg fusion protein; Lane 5: complete cleavage at 6 hrs.

CHAPTER 4

EFFECT OF RETENTION SIGNALS ON FUSION PROTEIN EXPRESSION IN TOBACCO CELL SUSPENSIONS

4.1 INTRODUCTION

Transgenic plant technologies have become a cornerstone in modern biotechnology. In the past, most of these techniques aimed at improving disease resistance and developing herbicide resisitance. With advances in plant transformation systems it is now feasible to produce biologically active proteins in plants, sometimes referred to as "molecular farming". Plants are a suitable alternative to microbial expression systems, as the production of plant biomass is relatively inexpensive. The potential of transgenic plants as production vehicles has been demonstrated by several examples (Goddijn and Pen 1995). Inspite of the ability of mammalian cell cultures to produce authentic proteins, they often failed to meet the need for a simple efficient, high yielding and low-cost method of protein production. Other drawbacks of using mammalian cells are the ample amounts of time required to reach high expression levels and the instability of expression (Hippenmeyer and Pegg 1995). *E. coli* cell cultures and fermentations alleviate some of the problems of mammalian cultures in terms of production levels and costs but in many cases, the proteins of interest are not truly authentic because of poor post-translational modifications in *E coli* .

Transgenic plants, with its post translational abilities alongwith the advantage of low cost of producing plant biomass, infinite capacity and existing infrastructure in harvesting, storing and processing of crop plants provide a viable source of producing enzymes and therapeutic proteins. Examples of industrial enzymes that have been

produced in transgenic plants include xylanase, heat-stable β-glucanase and phytase among others (Owen and Pen 1996). One of the earlier examples of successful heterologous expression of mammalian proteins in plants was the demonstration that leu-enkephalin, a pain-killing neuropeptide, could be produced to high levels when fused to part of the *Brassica napus* 2S seed storage protein (Vandekerckhove et al., 1989). Genes for several antibodies (Hiatt et al., 1989; Owen et al., 1992), interferon- α (Zhu et al., 1994) and human serum albumin (Sijmons et al., 1990) have been expressed in plants. Very high levels of α - trichosanthin, a root-derived component of the chinese medicinal plant *Trichosanthes kirilowii* , were obtained using a Tobacco Mossaic Virus (TMV) derived vector for overexpression (Kumagai et al., 1993). Other promising anti-cancer and/or anti-HIV plant derived compounds sucessfully engineered into transgenic plants are ricin (Gadani et al., 1995) and pokeweed antiviral protein (Hemming, 1995). Ricin is a potent ribosome-inactivating lectin originally derived from the castor bean, *Ricinus communis*, whose toxic properties have been known for many years. Transgenic potato tubers having hepatitis B surface antigen (HbsAg) has been tested as an oral vaccine in pre-clinical animal trials (Richter et al., 2000).

In any protein production system, the raw material is of primary consideration. Usually, therapeutic proteins have stringent purity requirements and therefore the less complex the raw material the simpler the the downstream processing. Using plant tissues as a raw material for protein recovery have problems like protein interactions with other protoplasmic constituents like lipids and nucleic acids. Moreover, the protein may be unstable and sensitive to proteases. Disruption of the plant cells is made difficult because of the cellulose fibers in the cell wall. In most cases, large amounts of phenolic

compounds are present in the vacuole and once plant tissue is broken to get the target protein out of the protoplasm, the compartmentalization ceases to exist and the proteins get exposed to the phenolic compounds. These compounds form complexes with the protein. Therefore, most purification techniques for isolating proteins from plants should specifically separate the phenols from proteins. One way to overcome problems in production in plants is to get the protein produced in the extra-cellular fluids since these fluids are less complex than the intracellular products. Moreover, instead of using whole plants, plant cell culture could be a viable alternative. Since plant cells can be grown on cheaper protein-free media, production and purification costs of target proteins from plant cells are lower (Magnuson et al., 1998). Plant cell culture for production of secondary metabolites has been used in the past. Even in the case of enzymes, there are potential advantages of producing the enzymes in plant cell cultures especially those enzymes that require specific post–translational modification for activity. In comparison to mammalian cells, plant cell culture is risk free because of the absence of human pathogens in the product. A heat stable endo- β-(1-4) glucanse of *Trichoderma reesei* has been produced in suspension cell cultures of barley (Aspegren et al., 1995). Chloramphenicol acetyltransferase (CAT) was produced in genetically modified tobacco cells (Hogue et al., 1990). Several therapeutic proteins have also been produced in plant systems. Erythropoetin has been produced in tobacco cells (Matsumoto et al., 1993, 1995). Other human proteins include epidermal growth factor (Higo et al., 1993) and lactoferrin (Mitra and Zhang, 1994). The human cytokines, IL-2 and IL-4 were produced in suspension cells of tobacco and it was found to be secreted into the media which was facilitated by mammalian leader sequences (Magnuson et al., 1998). Biologically active human

granulocyte-macrophage colony stimulating factor (hGM-CSF) was produced for the first time in plant cell culture (James et al., 2000).

GM-CSF is a blood protein and is of immense therapeutic value since it has been found to stimulate the proliferatation, maturation and function of variety of haematopoetic cells (Sieff , 1987; Nicola 1989; Gasson , 19891). It is produced by many cell types including T cells, B cells, macrophages, endothelial cells and fibroblasts in response to cytokine or immune and inflammatory stimuli. GM-CSF is a growth factor for erythroid, megakaryocyte and eosinophil progenitors. GM-CSF can also stimulate the proliferation of a number of tumor cell lines, including osteogenic sarcoma, carcinoma and adenocarcinoma cell line. GM-CSF is species specific and and human GM-CSF has no biological effects on mouse cells. GM-CSF exerts its biological effects through binding to specific cell surface receptors. GM-CSF has many beneficial uses in chemotherapy, bone marrow transplantation, cancer and AIDS treatments (Metcalf, 1991). The mature GM-CSF is a glycoprotein consisting of 127 amino acid residues. GM-CSF is produced in very low levels in human tissues. Therefore, alternative methods need to be evaluated. Successful expression of GM-CSF has been achieved in transgenic tobacco seeds (Sardana et al., 1998) and in tobacco cell culture (James et al., 2000).

Signal peptides and retention signals have been used in the past to improve accumulation and stability of proteins. Different signal sequences were tested on a modified beta-glucuronidase (GUS) protein to direct GUS into the secretory pathway of transgenic tobacco. It was found that the protein was not secreted to the extra-cellular space but was associated with the ER and the plasma membranes whereas without the signal sequence, the occurrence of GUS was cytosolic (Yan et al., 1997). Molecular

signals that can target proteins to the endoplasmic reticulum (ER) or to other compartments have been identified. Proteins have been retained in the ER of yeast (Pelham, 1988) and animals (Munro and Pelham, 1987) by HDEL and KDEL C-terminal end signals and it is believed that they could be used in plants. Fusion of the tetra-peptide KDEL to the C-terminus of vicilin, a pea seed protein, resulted in an average 100-fold increase in accumulation of protein in tobacco leaves and an average 20-fold increase in lucerne leaves. Moreover, there was an average 12-fold increase in the half-life of transgenic lucerne leaves (Wandelt et al., 1992). In another report, single-chain antibody ScFv (32 kDa) was produced and secreted in soluble form in tobacco cell suspension culture (Xu et al., 2002). The presence of the plant leader sequence, the tobacco etch virus (TEV) sequence, and the KDEL peptide in the expression construct resulted in the highest expression of total scFv (Xu et al., 2002). When HDEL was fused to the C-terminus of vacuolar or extracellular forms of sporamin, it was found by subcellular fractionation studies that the resultant proteins accumulated in the ER (Gomord et al., 1997). It was also proposed that HDEL may have quality control functions by targeting chaperones or chaperone-bound protiens that escape the ER to lysosomal compartments of plants for degradation (Gomord et al., 1997).

Use of green fluorescent protein as a secretory reporter in transgenic plant cell cultures has been established (Liu et al., 2001). In order to test the utility of GFP and the effect of the C-terminal retention signal HDEL, we used GM-CSF as our model protein. In this preliminary study, expression constructs were made with the *Arabidopsis* basic chitinase signal peptide to target the GMCSF-GFP fusion protein and to determine the effect of having HDEL on the C-terminus, on fusion protein accumulation/secretion.

4.2 MATERIALS AND METHODS

4.2.1 Fusion protein construct

The mgfp5-ER insert was obtained by digestion of pBIN-mgfp5-ER with *Bam*HI-*Sac*I and ligated into the respective sites of Bluescript SK (Stratagene, La Jolla, CA). The sequence was modified through a number of polymerase chain reaction (PCR) amplifications using Platinum Pfx DNA Polymerase (Invitrogen, Carlsbad, CA) and oligonucleotides for site-directed mutagenesis as described in Chapter 2 to remove an internal *Nco*I restriction site and to change the serine in position 65 to a threonine (S65T). The sequence was further modified by amplification with the oligonucleotide 5'-GGA GAT ATA ACC ATG **GCG** ACT AAT CTT-3' (sense primer) to incorporate an *Nco*I restriction site at the N-terminal methionine of the *Arabidopsis thaliana* chitinase signal peptide (underlined) and to change the second amino acid from a lysine to an alanine (K2A, bold), and the T3 primer (antisense primer). In addition, another amplification was performed using the same sense primer and the oligonucleotide 5'-GCTGGAGCTCTTAAAGCTCATC**TTA**TTTGTATAG-3' (antisense primer) to incorporate a termination codon (bold) to remove the C-terminal (-HDEL) ER retention signal. The amplified products were digested with *Nco*I-*Sac*I and ligated into the respective sites of pGEM-5Zf (Promega, Madison, WI). All modifications to the sequence were verified by DNA sequencing.

The human granulocyte-macrophage colony-stimulating factor (GM-CSF) was amplified by PCR from plasmid pGA748 (James et al., 2000). The sequence was modified using the oligonucleotides 5'-GATC**GAATTC**GCACCCGCCCGCTCGCCC-3' (sense primer) and 5'-CTAC**GAATTC**CTCCTGGACTGGCTCCCA-3' (antisense

primer) to incorporate *Eco*RI restriction sites (bold) on both ends of the cDNA that are in-frame with the GM-CSF coding sequence. The sense primer was designed to remove the N-terminal human signal peptide of GM-CSF and the antisense primer removed the termination codon. The amplified fragment was digested with *Eco*RI and ligated into *Eco*RI-digested Bluescript SK (Stratagene, La Jolla, CA). The modifications to the sequence were verified by DNA sequencing.

The GMCSF-GFP fusion proteins were generated by ligating the *Eco*RI-digested GM-CSF into the *Eco*RI-digested GFP constructs described above. Ligation into the *Eco*RI restriction site of the GFP constructs would position the GM-CSF between the *Arabidopsis thaliana* chitinase signal peptide and the modified GFP, and would integrate the GM-CSF in the same reading frame to produce a fusion protein. Fusion proteins with and without a C-terminal (-HDEL) ER retention signal were developed. The correct orientation of the GM-CSF in the fusion construct was verified by PCR.

The GMCSF-GFP fusion constructs were digested with *Nco*I-*Sac*I to remove the entire fusion protein construct and were ligated into the respective sites of a modified plant promoter cassette derived from pBI-525 (Dattla et al., 1993). The pBI-525 vector was previously modified to incorporate a *Sac*I restriction site just before the NOS terminator and the cassette was digested with *Hin*dIII-*Eco*RI and ligated into the respective sites of Bluescript SK (Stratagene, La Jolla, CA). The entire fusion protein-plant promoter cassette was excised with *Hin*dIII-*Bam*HI (*Bam*HI is in the multiple cloning site of Bluescript SK at the 3' end of the cassette) and ligated into the respective sites of the plant transformation vector pBINPLUS (van Engelen et al., 1995).

4.2.2 Tobacco Transformation

The plant expression construct both with and without the C-terminal HDEL was transferred into *Agrobacterium tumefaciens* LBA4404 by electroporation. Tobacco (*Nicotiana tabacum* cv Xanthi) transformation was carried out as described previously (Horsch et al., 1985) with modifications as suggested (Fisher and Guiltinan, 1995). Transformants were selected on Murashige and Skoog medium (Murashige and Skoog, 1962) containing Gamborg's vitamins and 3% sucrose, and supplemented with 300 mg/l kanamycin and 500 mg/l carbenicillin. Individual shoots were excised and rooted on the same medium supplemented with 200mg/l kanamycin and 250 mg/l carbenicillin.

4.2.3 *Screening of tobacco plants*

A single leaf from the middle of the plant that was grown for approximately a month in magenta jars was cut out and ground with liquid nitrogen using a mortar and pestle for a total of 3 times so as to obtain a fine powder. The powder was collected in eppendorf tubes and resuspended in 1.1 ml of GFP extraction buffer (50 mM NaH_2PO_4 (pH 8.0), 10 mM Tris-HCl (pH 8.0), 200 mM NaCl). The samples were thoroughly mixed by vortexing and then fluorescence of 1 ml of sample was measured at an excitation wavelength of 490nm and an emission wavelength of 512nm using a fluorescence spectrophotometer (Hitachi Model F-2500, Tokyo, Japan).

The protein content in the samples was determined using a protein assay reagent (Bio-Rad, Hercules, CA) with bovine serum albumin as the standard (Bradford, 1976). Out of 29 plants, the top 3 plants were selected based on fluorescence per milligram of protein for callus initiation.

4.2.4 *Callus Initiation*

Sections of stem tissue from young primary transformants was cut into small pieces and placed on solid Murashige and Skoog medium containing Gamborg's vitamins and 2% sucrose, and supplemented with 1.0 mg/l 2,4-D, 0.1 mg/l kinetin, 300 mg/l kanamycin and 500 mg/l carbenicillin for callus initiation and growth.

4.2.5. *Suspension Cell Culture*

Suspension cell cultures were established by introducing transgenic calli into Murashige and Skoog medium containing Gamborg's vitamins with 2% sucrose, and supplemented with 1.0 mg/l of 2,4-D. Cultures were initially grown for 2-3 weeks in the liquid medium and thereafter routinely transferred to new medium with a 15-20% (v/v) inoculum every 7-10 days. The suspension cell cultures were maintained at 25°C in the dark with shaking at 100 rpm on a gyratory shaker.

4.2.6. *GFP Extraction and Fluorescence Determination*

A three ml sample of cell suspension was centrifuged at 10,000 rpm for 5 mins and the supernatant removed using a 0.2 μm syringe needle connected to the lab vacuum. The cells were resuspended in 3 ml of buffer (50 mM NaH_2PO_4 (pH 8.0), 10 mM Tris-HCl (pH 8.0), 200 mM NaCl) and lysed using an ultrasonic cell disrupter equipped with a microprobe for 5 min at its lowest output setting (Branson Model 250, Danbury, CT). The supernatant was collected after centrifugation at 10,000 rpm 5 min, and used for determining intracellular GFP. For extracellular GFP determination, a sample of cell suspension was centrifuged at 10,000 rpm for 5 min and filtered through a 0.2 μm filter.

GFP fluorescence intensity was measured using a fluorescence spectrophotometer (Hitachi Model F-2500, Tokyo, Japan) at an excitation wavelength of 490 nm and an

emission wavelength of 512 nm. Culture samples were diluted to levels at which the fluorescence is linearly correlated with cell density to alleviate the interference of the inner filter effect (Srinivas and Mutharasan, 1987). Subsequently, a linear regression was made and the culture fluorescence of the undiluted sample was extrapolated. Protein was determined using a protein assay reagent (Bio-Rad, Hercules, CA) with bovine serum albumin as the standard (Bradford, 1976).

4.2.7. *Western Blot Analysis*

Samples (20 μl) were mixed with a third of the sample volume with 4X reducing sample buffer and resolved on a 12% SDS-polyacrylamide gel as described by Laemmli (1970) and electrophoretically transferred to a PVDF membrane (Millipore, Bedford, MA). Purified recombinant GFP (Clontech, Palo Alto, CA) was used to show the difference in migration of the fusion protein w.r.t GFP by itself. Western analysis was performed as previously described (Bugos et al., 1999). The blot was incubated for one hour with an anti-GFP rabbit polyclonal antibody (Molecular Probes, Inc., Eugene, OR) diluted 1:2000 in TBST (10 mM Tris-HCl, pH 8.0, 150 mM NaCl, 0.05% Tween 20) containing 5% (w/v) nonfat dry milk. After washing the blot with TBST, it was then incubated for 45 min with an alkaline phosphatase conjugated goat anti-rabbit IgG (Southern Biotechnology Associates, Birmingham, AL) diluted 1:2000 in TBST with milk. GFP protein was detected by exposing the membrane to NBT/BCIP.

4.3. RESULTS AND DISCUSSION

Expression constructs with and without a C-terminal HDEL were used to transform tobacco (*Nicotiana tabaccum* cv Xanthi). The protein expression was driven by the cauliflower mosaic virus (CaMV) 35S promoter. A total of 29 independent kanamycin resistant transformants were selected and leaf tissue was extracted to analyze protein expression. As can be seen from Figure 4.1 different plants express the fusion protein at different levels. This can be attributed to reasons like the position effect ie., position where the transgene gets inserted on the plant genome (Dean et al., 1988) or methylation of the transgene (Prols and Meyer, 1992). Similarly 22 plants expressing constructs with no HDEL were screened. The best lines were selected based on fluorescence per unit protein. GFP fluorescence can provide good estimates of transgene expression as was reported by (Remans et al., 1999). The best transgenic lines (those with a star in Figure 4.1) were selected for initiating callus. Similarly, the best lines without a C-terminal HDEL were selected (transgenic lines with the star in Figure 4.2) for initiating callus. The callus was maintained on media containing kanamycin and carbenicillin. In Figure 4.2 the level of florescence was lower than most lines in Figure 4.1

Suspension cells were developed from the kanamycin resistant callus and maintained for evaluation of fusion protein expression. Samples were taken from an 8-day culture which is approximately the time period for maximum GFP production in suspension cells (Liu et al., 2001; Guan et al., unpublished results). Fluorescence of the suspension cells was measured in a quantitative manner in the fluorescent spectrophotometer. Measurements were made both for intracellularly expressed protein

and for the media to see if there was any secretion. Fluorescence intensity values were obtained by extrapolating the readings obtained by diluting the samples (Liu et al., 2001) to alleviate the interference of the inner filter effect (Srinivas and Mutharasan, 1987). From Figure 4.3, it can be seen that HDEL seems to have an effect on fluorescence intensity since all the lines with the HDEL are brighter than those without HDEL. This is consistent with results obtained in the past where C-terminal signals improved accumulation. For example, fusion of the tetra-peptide KDEL to the C-terminus of vicilin, a pea seed protein, resulted in an average 100-fold increase in accumulation of protein in tobacco leaves and an average 20-fold increase in lucerne leaves. However, there is hardly any fluorescence in the media (Figure 4.4) for lines with the HDEL, consistent with the reports that no secretion occurred in the media because of the presence of the HDEL which is an ER retention signal (Gomord et al., 1997; Munro and Pelham, 1987). Without the HDEL, there seemed to be a marginal increase in fluorescence of the media with the exception of one transgenic line which had approximately a 5-fold increase (Figure 4.4).

To make sure that fluorescence was exhibited by the fusion protein and not any cleaved GFP bands, western blots were done (Figures 4.5 and 4.6) with an anti-GFP polyclonal antibody (Molecular Probes, Inc., Eugene, OR). No bands were seen for samples from the media (data not shown) for the lines with the HDEL. As can be seen from the anti-GFP blot, majority of the protein is intact as a fusion protein (theoretical weight 44 kDa) and no cleavage products are seen. There probably is a minor amount of glycosylation but cannot be accurately determined from this blot because of excess amounts of protein in the lane (Figure 4.5). GM-CSF by itself has a projected weight of

14 kDa (Lee et al., 1985). It could be higher depending on the glycosylation of GM-CSF which has five O-linked sites and two N-linked sites (Voet and Voet, 1995). In the blot, no distinct bands heavier than the fusion protein is seen. In the case of GM-CSF, glycosylation does not affect its conformational stability (Wingfield et al., 1988).

Western blot analysis samples without the HDEL (Figure 4.6) shows that there is significant amount fusion protein that is being expressed intracellularly with the best line (Lane 1, Figure 4.6) representing about 2.6% total soluble protein, which was obtained by densitometry. However, there is almost a 10-fold increase in fluorescence intensity between lines having HDEL and those without HDEL (Figure 4.3). This seems to indicate that a portion of the fusion protein is not fluorescent. It has been shown that absence of HDEL causes the protein to go into the default protein secretion pathway (Denecke et al., 1990) . Secretion of the protein to the cell wall is facilitated by this default pathway. Transient expression of GFP in onion epidermal cells localized to different organelles showed distinct patterns of fluorescence (Scott et al., 1999). It was found that the GFP localized to the cell wall was not fluorescent at the normal cell wall pH but starts fluorescing when bathed in a pH 7.0 buffer. Since our cell cell culture samples were sonicated, the protein is most likely not associated with the cell wall. Moreover, the samples were made up in GFP extraction buffer (50mM NaH_2PO_4, 10mM Tris-HCl, 200 mM NaCl pH 8.0) and therefore pH may not be the reason for the lack of fluorescence. GFP fluorescence is pH dependent (Kneen et al., 1998) and fluorescence decreases with pH. However, the effect of pH is also reversible above pH 5.0, wherein the fluorescence is restored when the pH is raised to the original value (Kneen et al., 1998).

Another possibility is that the protein may bind to other proteins that have a retention signal which may provide a mechanism for complete assembly in the ER before being secreted (Pelham, 1988). Gethering et al. 1986 have shown that a number of unfolded proteins bind to Binding Protein (BiP). This indirectly would mean that the GMCSF-GFP fusion protein is not completely folded which is consistent with the results obtained wherein enough protein is detected in the western blot but with much lower fluorescence. This implies that the absence of the HDEL causes the fusion protein either to be unstable or is not properly folded, which needs to be determined. This is consistent with the results of Wandelt et al. 1992. They reported an increase in half-life from 4.5 hrs to over 48 hrs by adding the KDEL on to the C-terminus of Vicilin, a pea seed protein. Therefore, this role played by HDEL seems to depend on the nature of the protein. Typically, absence of HDEL/KDEL results in secretion of the protein (Yan et al., 1997; Gomord et al., 1997), contrary to the results that we obtained. This would mean that the protein would be expressed intracellularly and therefore appropriate purification strategies need to be considered especially if the process needs to be scaled up. Ideally, if the protein is secreted into the surrounding media, then purification is much easier because cell material can be removed by vacuum filtration followed by clarification of the filtrate before initial purification steps (Fischer et al., 1999). However, for proteins like GM-CSF, which has very high therapeutic value, stability of the protein is of utmost priority. If the stability, can be improved by HDEL, it would be more viable to chose an appropriate strategy to recover the intracellular protein, rather than have unstable protein secreted into the media.

REFERENCES

Aspegren K, Mannonen L, Riatala A (1995) Secretion of heat-stable fungal β-glucanase from transgenic, suspension cultured barley cells. Mol Breeding 1: 91-99.

Bradford MM (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal Biochem. 72:248-54

Chiu W, Niwa Y, Zeng W, Hirano T, Kobayashi H, Sheen J (1996) Engineered GFP as a vital reporter in plants. Curr Biol 6:325-30

Dattla RSS, Bekkaoui F, Hammerlindl JK, Pilate G, Dunston DI, Crosby WL (1993) Improved high-level constitutive foreign gene expression in plants using an AMV RNA 4 untranslated leader sequence. Plant Science 94:139-149.

Dean C, Jones J, Favreau M, Dunsmuir P, Bedbrook J (1988) Influence of flanking sequences on variability in expression levels of an introduced gene in transgenic tobacco plants. Nucleic Acids Res 16: 9267-83

Denecke J, Botterman J, Deblaere R (1990) Protein secretion in plant cells can occur via a default pathway. Plant Cell 2: 51-59

Fisher DK and Guiltinan MJ (1995) Rapid, efficient production of homozygous transgenic tobacco plants with *Agrobacterium tumefaciens*: A seed-to-seed protocol. Plant Mol. Biol. Rep. 13: 278-289.

Gasson JC (1991) Molecular physiology of granulocyte-macrophage colony-stimulating Factor. Blood 77:1131-45

Goddijn OJM and Pen J (1995) Plants as bioreactors Trends Biotechnol. 13: 379-387

Gadani F, Ayers D, Hempfling W (1995) Tobacco: a tool for plant genetic engineering research and molecular farming. Part 11. Agro-Food-industry Hi-Tech 6: 3-6

Gomord V, Denmat L, Fitchete-Laine A, Satiat-Jeunemaitre B, Hawes C and Faye L (1997) The C-terminal HDEL sequence is sufficient for retention of secretory proteins in the endoplasmic reticulum (ER) but promotes vacuolar targeting of proteins that escape the ER. Plant J 11: 313-325

Hemming D (1995) Molecular farming: using transgenic plants to produce novel proteins and other chemicals. AgRiotech News Inf 7: 19N-29N

Hiatt A, Cafferkey R and Bowdish K (1989) Production of antibodies in transgenic plants. Nature 432: 76-78.

Higo K, Saito Y and Higo H (1993) Expression of a chemically synthesized gene for human epidermal growth factor under the control of cauliflower mosaic virus 35s promoter in transgenic tobacco. Biosci. Biotech. Biochem. 57:1477-1481.

Hippenmeyer PJ, Pegg LE (1995) Enhancing expression of recombinant proteins in mammalian cells using the herpesvirus VP16 transactivator. Curr Opin Biotechnol 6: 548-552

Hogue RS, Lee JM, An G (1990) Production of a foreign protein product with genetically modified plant cells. Enzyme Microb Technol 12:533-8

Horsch RB, Rogers SG, Fraley RT (1985) Transgenic plants Cold Spring Harb Symp Quant Biol 50:433-7

James EA, Wang C, Wang Z, Reeves R, Shin JH, Magnuson NS, Lee JM (2000) Production and characterization of biologically active human GM-CSF secreted by genetically modified plant cells. Protein Expr Purif 19:131-8

Kneen M, Farinas J, Li Y, and Verkman AS (1998) Green fluorescent protein as a noninvasive intracellular pH indicator. Biophysical J. 74: 1591-1599.

Kumagai MH, Turpen TH, Weinzeni N, Della-Cioppa G, Turpen AM, Donson JD, Hilf ME, Grantham GL, Dawson WO, Chow TP, Piatak M, Grill LK (1993) Rapid, high- level expression of biologically active alpha-tricosanthin in transfected plants by an RNA viral vector. Proc. Natl. Acad. Sci., USA 90: 427-430

Laemmli UK (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature 227: 680-5

Lee F, Yokota T, Otsuka T, Gemmel L, Larson N, Luh J, Arai K, and Rennick D (1985) Isolation of cDNA for a human granulocyte-macrophage colony stimulating factor by functional expression in mammalian cells. Proc. Natl. Acad. Sci. USA 82, 4360-4364

Liu S, Bugos RC, Dharmasiri N, Su WW (2001) Green fluorescent protein as a secretory reporter and a tool for process optimization in transgenic plant cell cultures: J Biotechnol 87:1-16

Magnuson NS, Linzmaier PM, Reeves R, An G, Hayglass K and Lee JM (1998) Secretion of biologically active human Interleukin-2 and Interleukin-4 from genetically modified tobacco cells in suspension culture. Protein Express. & Purif. 13: 45-52

Matsumoto S, Ikura K, Ueda M and Sasaki R (1993) Expression of human erythropietin in cultured tobacco cells. Biosci. Biotech. Biochem. 57:1249-1252.

Matsumoto S, Ikura K, Ueda M and Sasaki R (1995) Characterization of a human glycoprotein (erythropoietin) produced in cultured tobacco cells. Plant Mol. Biol. 27:1163-1172.

Metcalf D (1991) Control of granulocytes and macrophages: molecular, cellular, and clinical aspects. Science 254: 529-533

Mitra A, Zhang Z (1994) Expression of a human lactoferrin cDNA in tobacco cells produces antibacterial protein(s) Plant Physiol 106:977-81

Munro S, Pelham HR (1987) A C-terminal signal prevents secretion of luminal ER Proteins. Cell 48:899-907

Murashige T and Skoog F (1962) A revised medium for rapid growth and bioassays with tobacco tissue cultures. Physiol. Plant. 15, 473-497.

Nicola NA (1989) Hemopoietic cell growth factors and their receptors Annu Rev Biochem;58:45-77

Owen MRL, Pen J (eds) (1996) Transgenic plants: A production system for industrial and pharmaceutical proteins. Wiley New York

Owen M, Gandecha A, Cockburn B, Whitelam G (1992) Synthesis of a functional anti-phytochrome single-chain Fv protein in transgenic tobacco. Biotechnology (NY) 10:790-4

Pelham HR (1988) Evidence that luminal ER proteins are sorted from secreted proteins in a post-ER compartment. EMBO J 7: 913-8

Prols F, Meyer P (1992) The methylation patterns of chromosomal integration regions influence gene activity of transferred DNA in Petunia hybrida. Plant J 2: 465-75

Remans T, Schenk PM, Manners JM, Grof CPL, Elliott AR (1999) A protocol for the fluorimetric quantification of mGFP5-ER and sGFP(S65T) in Transgenic plants. Plant Molecular Biology Reporter 17: 385-395

Richter LJ, Thanavala Y, Arntzen CJ, Mason HS (2000) Production of hepatitis B surface antigen in transgenic plants for oral immunization Nat. Biotechnol. 18: 1167-71.

Sardana RK, Ganz PR, Dudani A, Tackaberry ES, Cheng X, Altosaar I (1998) In: Cunningham C, Porter AJR (eds) Recombinant Proteins from plants. Humana, Totowa pp 77-87

Scott A, Wyatt S, Tsou PL, Robertson D and Allen NS (1999) Model system for plant cell biology: GFP imaging in living onion epidermal cells. Biotechniques 26: 1125-1132

Sieff CA (1987) Hematopoietic growth factors J Clin Invest 79:1549-57

Sijmons PC, Dekker BM, Schrammeijer B, Verwoerd TC, van den Elzen PJ, Hoekema A (1990) Production of correctly processed human serum albumin in transgenic plants. Biotechnology (NY) 8:217-21

Srinivas SP and Muthurasan R (1987) Inner filter effects and their interferences in the interpretation of culture fluorescence. Biotechnol. Bioeng. 30, 769-774

Vandekerchove J, Van Damme J, Lijsebetterns VM, Botterman J, De Block M, Vandewiele M, De Clercq A, Leemans J, Montagu VM and Krebbers E (1989) Enkephalins produced in transgenic plants using modified 2s seed storage proteins. Bio/tech 7, 929-932.

Voet D and Voet JG (1995) Biochemistry second ed., pp. 271-273, Wiley , New York.

van Engelen FA, Molthoff JW, Conner AJ, Nap JP, Pereira A, Stiekema WJ (1995) pBINPLUS: an improved plant transformation vector based on pBIN19. Transgenic Res 4:288-90

Wandelt CI, Khan MRI, Craig S Schroeder HE, Spencer D and Higgins TJV (1992) Vicilin with carboxy-terminal KDEL is retained in the endoplasmic reticulum and accumulation to high levels in the leaves of transgenic plants. Plant J. 2:181-192

Wingfield P, Graber P, Moonen P, Craig S, Pain RH (1988) The conformation and stability of of recombinant-derived granulocyte-macrophage colony stimulating factors. Eur. J. Biochem 173: 65-72

Xu H, Montoya FU, Wang Z, Lee JM, Reeves R, Linthicum DS, Magnuson NS (2002) Combined use of regulatory elements within the cDNA to increase the production of a soluble mouse single-chain antibody, scFv, from tobacco cell suspension cultures. Protein Expr Purif 24:384-94

Yan X, Gonzales A and Wagner GJ (1997) Gene fusions of signal sequences with a modified β- Glucuronidase protein in the secretory pathway/plasma membrane. Plant Physiol. 115: 915-924

Zhu Z, Hughes KW, Huang L, Liu C, Hou Y and Li X (1994) Expression of human α-interferon cDNA in transgenic rice plants. Plant cell tissue organ culture 36: 197-204.

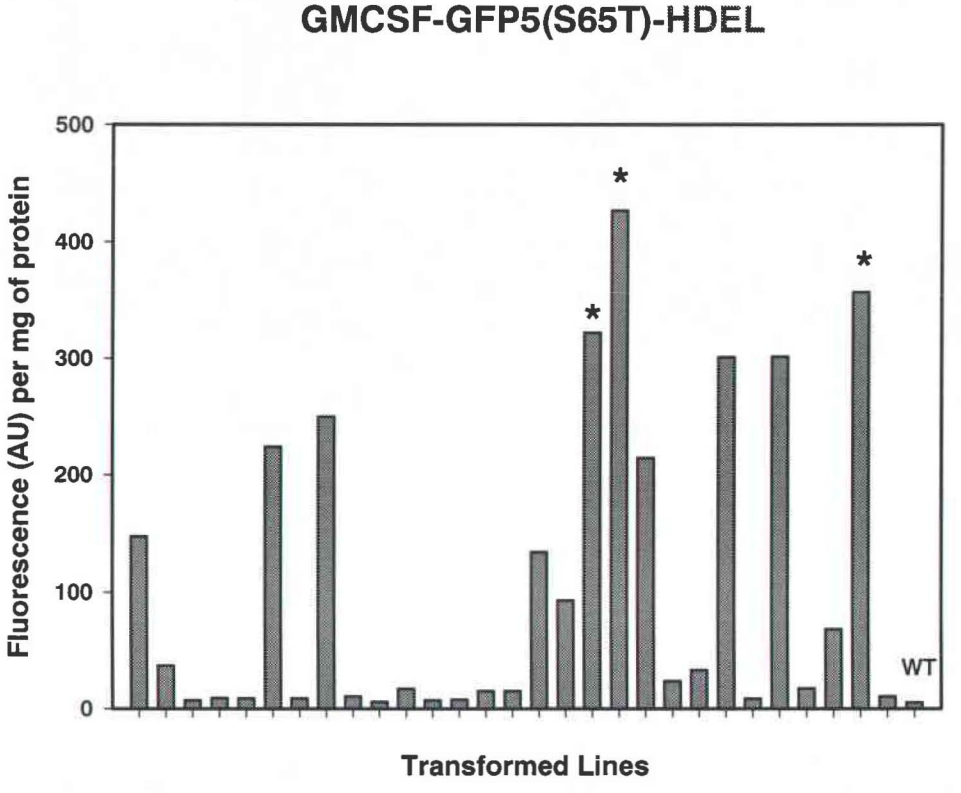


Figure 4.1: Screening of tobacco plants expressing fusion protein with a C-terminal HDEL. 29 plants were screened based on fluorescence of soluble protein extracted from leaf tissue by grinding in liquid nitrogen. The ground tissue was resuspended in 1.1 ml of GFP extraction buffer (50 mM NaH_2PO_4 (pH 8.0), 10 mM Tris-HCl (pH 8.0), 200 mM NaCl). Fluorescence of 1 ml of sample was measured at an excitation wavelength of 490nm and an emission wavelength of 512 nm using a fluorescence spectrophotometer (Hitachi Model F-2500, Tokyo, Japan)

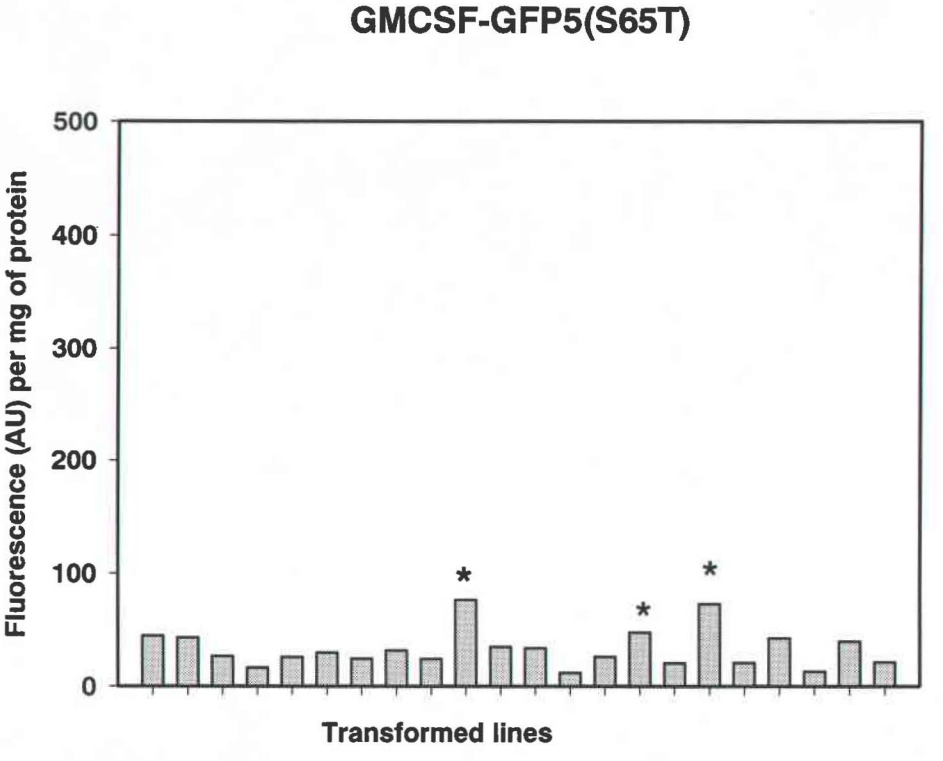


Figure 4.2: Screening of tobacco plants expressing fusion protein without a C-terminal HDEL. 23 plants were screened based on fluorescence of soluble protein extracted from leaf tissue by grinding in liquid nitrogen. The ground tissue was resuspended in 1.1 ml of GFP extraction buffer (50 mM NaH_2PO_4 (pH 8.0), 10 mM Tris-HCl (pH 8.0), 200 mM NaCl). Fluorescence of 1 ml of sample was measured at an excitation wavelength of 490nm and an emission wavelength of 512 nm using a fluorescence spectrophotometer (Hitachi Model F-2500, Tokyo, Japan)

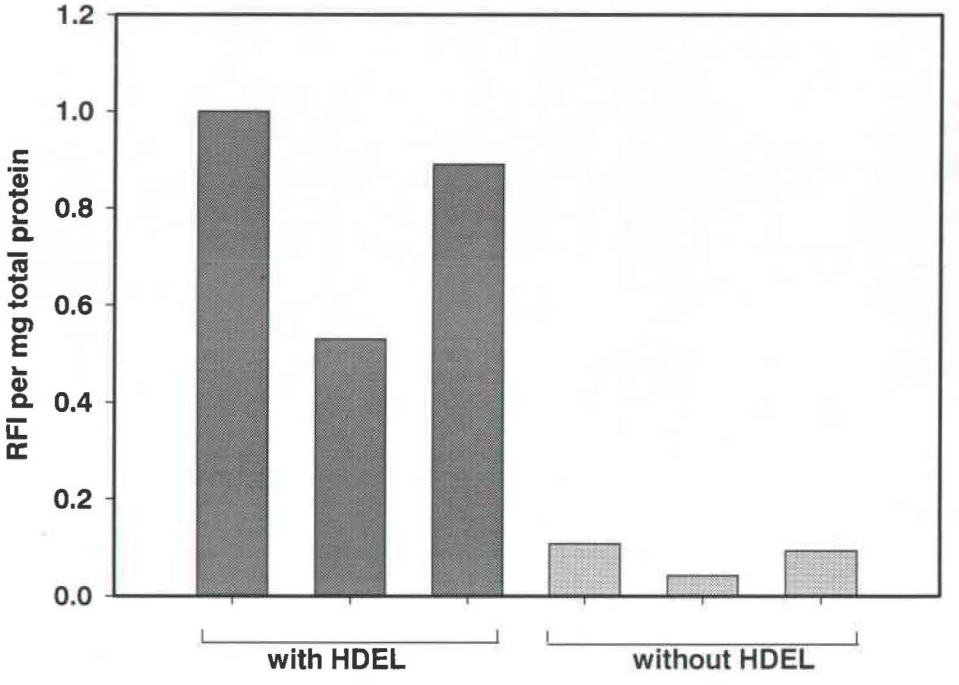


Figure 4.3 Fluorescence of intracellular fusion protein expressed in suspension cells is higher with HDEL. 3 ml of an 8-day cell suspension culture was taken and centrifuged at 10,000 rpm for 5min. 1 ml of supernatant was collected and fluorescence measured at 490nm excitation and 512nm emission to give an estimate of fusion protein in the media (M). The pelleted cells were resuspended in GFP extraction buffer to make up the volume to 3ml and sonicated at the lowest output for 5min. The sonicated sample was centrifuged as before and fluorescence of the supernatant gave an estimate of Intracellular (IC) expression. Data was normalized based on the highest intensity with the HDEL on the fusion protein.

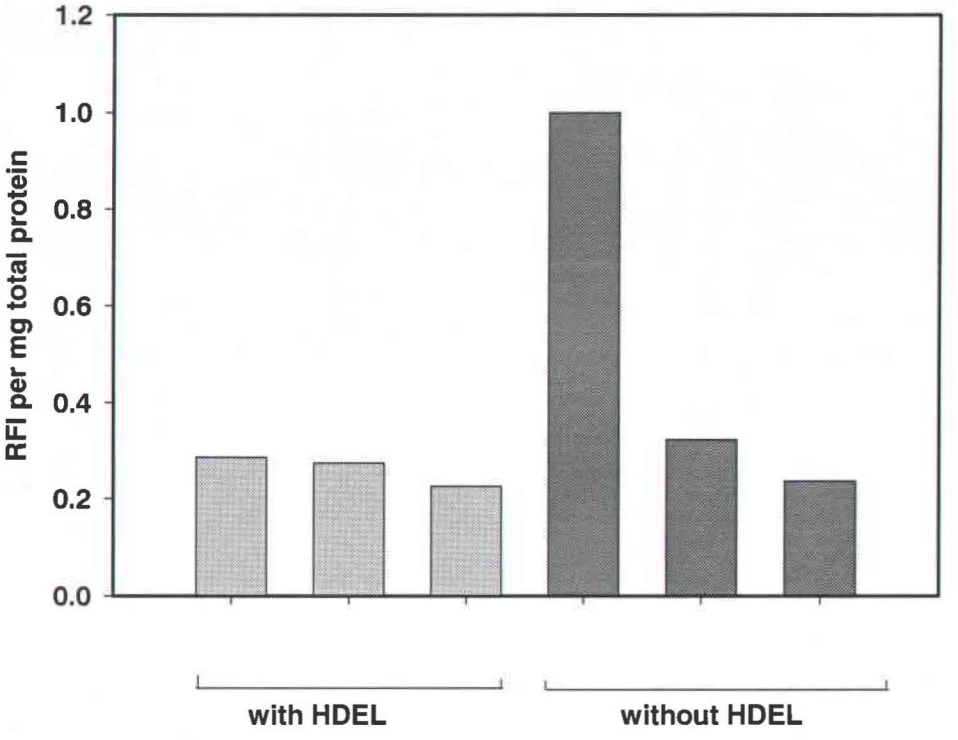


Figure 4.4 Fluorescence of fusion protein secreted in the media. 3 ml of an 8-day cell suspension culture was taken and centrifuged at 10,000 rpm for 5min. 1 ml of supernatant was collected and fluorescence measured at 490nm excitation and 512nm emission to give an estimate of fusion protein in the media (M). Data was normalized based on the highest intensity which was seen for a line without the HDEL.

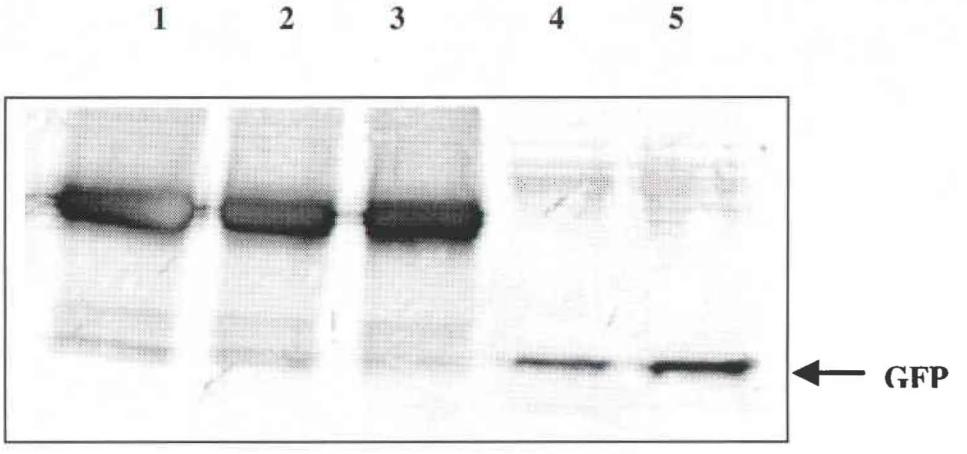


Figure 4.5. Western Blot depicting intracellular expression of GMCSF-GFP5(S65T)-HDEL in suspension cells. Anti-GFP rabbit polyclonal antibody (1:2000) was used as the primary antibody and alkaline phosphatase conjugated goat anti-rabbit IgG was used as the secondary antibody. Lanes 1, 2 & 3 represent 10μg total protein expressed of the best transgenic lines having the HDEL. Densitometry analysis was done using SigmaScan Pro 5.0 image-analysis software (SPSS Inc., Chicago, IL) and lanes 1 and 3 were estimated to have about 4.5% total soluble protein. Lanes 4 & 5 have 10 and 50ng of GFP standard (Clontech, Palo alto, CA)

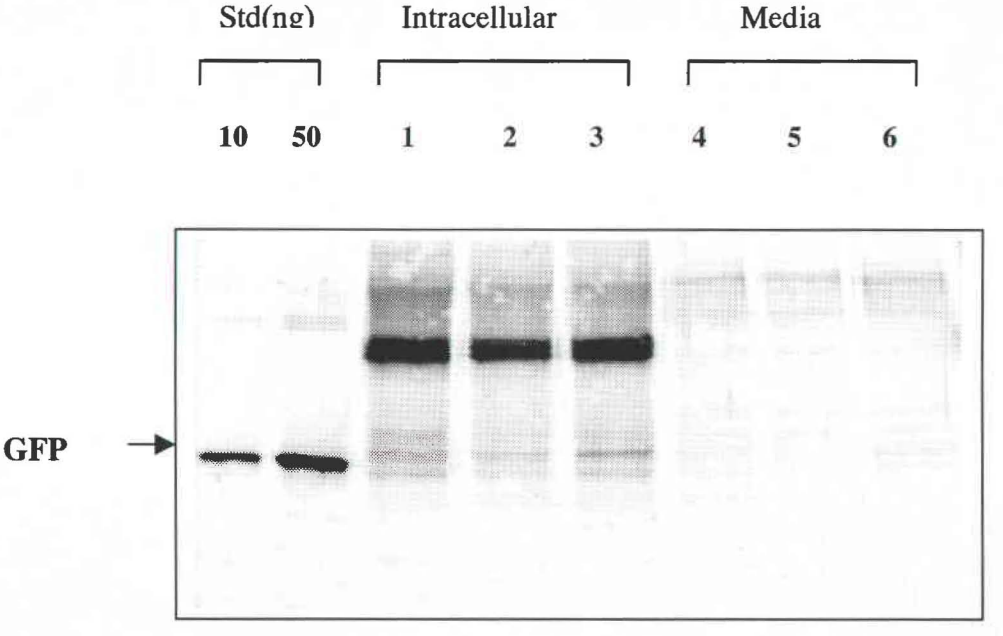


Figure 4.6. Western Blot depicting expression of GMCSF-GFP5(S65T) (without HDEL). Anti-GFP rabbit polyclonal antibody (1:2000) was used as the primary antibody and alkaline phosphatase conjugated goat anti-rabbit IgG was used as the secondary antibody. 20µl samples were loaded in each lane. 10 and 50ng of GFP standard (Clontech, Palo alto, CA) was used. Lanes 1, 2 & 3 show intracellular expression with about 2.6% total soluble protein in Lane 1 obtained by densitometry. Lanes 4, 5 and 6 shows protein secreted into the media.

CHAPTER 5

CONCLUSIONS AND RECOMMENDATIONS

5.1 CONCLUSIONS

Preliminary studies to see the effect of inserting a 6xHis tag into one of the solvent exposed loops (position 172-173) of mGFP5-ER resulted in the conclusion that there is a decrease in fluorescence (about 50%) in the GFP as a result of the insertion (see Chapter 2). However, the 6xHis tag in the loop was exposed and accessible to immobilized metal as was judged by the purity of the protein obtained. It was decided that the problem of loss of fluorescence could be alleviated to a great extent by inserting the 6xHis tag in a brighter variant of GFP which could be obtained by mutating the serine at position 65 to a threonine (S65T) (Heim et al., 1995).

The mutation did indeed increase the fluorescence of GFP. It was also decided to test the insertion of the 6xHis tag in another loop (position 157-158) of this brighter GFP5(S65T) variant (see Chapter 3). Based on the effect of temperature on the expression of soluble GFP variants, GFP172 was used as a fusion tag because soluble protein was obtained both at 28° and 37°C thereby enabling its use in a broad host range. GFP172 was fused to the C-terminus of Maltose Binding Protein (MBP), and it was found that the 6xHis tag within this fusion protein could purify it to homogenous purity in a single step. This fusion protein was spiked into tobacco leaf extract and it was found that the 75% of the protein could be recovered with some loss of fusion protein due to protease activity. This cleavage took place inspite of adding phenylmethanesulphonyl fluoride (PMSF) in the extraction buffer that was used to resuspend the ground leaf tissue. Western blot results (with anti-GFP antibody) on the eluted fractions from the

tobacco extract, showed that the cleaved product was approximately the size of GFP. However, on further analysis it was found that this product didn't fluoresce and therefore the recovery results obtained by using GFP-fusion protein standard series (fluorescence vs amount of protein) was a good estimate of the recovery of protein by IMAC.

In this study it was found that the C-terminal retention signal HDEL played a role in protein stability (Chapter 4). When GMCSF-GFP fusion protein was expressed in tobacco suspension cell culture, absence of HDEL resulted in far less amounts of fluorescent protein inspite of seeing sufficient protein on the western blot. This implies that the protein that was produced was mis-folded and hence the decrease in fluorescence compared to the fusion proteins expressed with the C-terminal HDEL.

This novel GFP molecule has potential advantages in recombinant protein production in different host systems like plants, yeast and animals where N-terminal signal peptides and C-terminal retention signals are used in order to accumulate protein and to increase its stability. By inserting the 6xHis tag within the GFP molecule, great amount of flexibility is provided in terms of attaching these signals or any other peptides since the N- and C-terminals of the fusion protein are free. Therefore, this novel GFP molecule can now monitor expression and also purify any protein attached to it with the added advantage of increased accumulation and better stability.

5.2 RECOMMENDATIONS

5.2.1. GFP detection

Though we see a decrease in the fluorescence, having the internal His tag has advantages that far outweigh the problem of decrease in fluorescence. The decrease could be because of mis-folding of the protein to a certain extent. The micro-environment

around the chromophore of GFP might have changed as a result of the insertion which is reflected by the excitation spectra of GFP172 (Figures 3.5 and 3.6). Probably, by adding the Ser147 to Pro (S147P), this folding problem could be overcome. It was shown that when S147P was combined with S65T, the resulting double mutant emitted higher fluorescence than GFP with a single S65T mutation and could be used over a wide range of culturing temperatures (Kimata et al., 1997). Ultimately what matters is whether the GFP fluorescence can be detected with respect to the background. It was found by Harper et al. (1999) that mGFP5-ER (the gene used in this study) could monitor expression of *Bacillus thuringiensus cry1Ac* gene when co-introduced in tobacco and oilseed rape. An excellent review on possible ways to tackle autofluorescence has been done (Billinton and Knight, 2001). Alternatively, advanced instrumentation and techniques for auto-fluorescence discrimination like optimized filter sets, dual wavelength differential fluorescence correction, fluorescence polarization, confocal laser scanning microscopy and advanced image analysis software could be used (Billinton and Knight, 2001). Of course, the strategy needs to be identified and optimized for each application. A quantitative fluorescent imaging system with blue laser excitation was developed (Niwa et al., 1999) to detect a codon optimized sGFP(S65T) that provided up to 100-fold brighter signals in eukaryotes (Chiu et al., 1996).

5.2.2. Fusion protein solubility

A temperature effect on GFP solubility was found. The other factor to be considered is the position of the GFP in the fusion protein. GFP tolerates N- and C-terminal fusions (Cubbitt et al., 1995). However, it has been found that in expressing fusion proteins, the solubility of the upstream protein has an influence on the solubility of

the entire fusion. MBP, the model protein chosen in this study, is known to be a soluble protein and it has been shown that if MBP is used at the N-terminus of the fusion partner, then the resultant fusion protein is usually soluble. This was demonstrated with MBP on the N-terminus of Procathepsin D and pepsinogen, both being proteins that normally form inclusion bodies by itself (Sachdev and Chirgwin, 1998). It was demonstrated in another report that the GFP fluorescence was influenced by the folding pattern of the upstream protein expressed alone and therefore could be used as a folding reporter (Waldo et al., 1999). This possibly brings out another advantage of the novel GFP reported in this work where it could not only monitor and purify any recombinant protein attached to its N-terminus but also give information about the folding properties of the upstream protein.

5.2.3. Spiking studies

In order to test the purification abilities of the GFP tag from tobacco leaf extract, spiking studies were done (Chapter 3). A problem encountered here was that of protease activity, in spite of the presence of PMSF in the extraction buffer. Different types of proteases like serine proteases usually present in all cells and cysteine proteases that dominate in plant cells, exist. PMSF is commonly used protease inhibitor (Hiat et al., 1989; Sijmons et al., 1990; Mason et al., 1992). The table below (Table 1.0) adapted from the website of Uptima (www.interchim.com) can be used as reference in selecting protease inhibitors depending on the system that is being used and the kind of proteases that are expected to be present. Protease inhibitor cocktail tablets (Roche Applied Science, Indianapolis, IN) that can target more than one protease at a time are also available.

Table 1.0 **Protease inhibitors**

Proteases	**Inhibitors**
Serine proteases (Chymotrypsin, Plasmin, Subtilisin, Thrombin, Trypsin)	General inhibitor: AEBSF Chymostatin, Soybean trypsin inhibitor PMSF Soybean trypsin inhibitor
Cysteine proteases (Cathepsin, Papain)	General inhibitor: Antipain & E64 CA-074 Chymostatin, PMSF
Metalloproteases (Ca^{+} metalloproteases, Metalloendoproteinases)	General inhibitor: EDTA EGTA Phosphoramidon
Aspartic acid proteases (Cathepsin D, Pepsin, Renin)	General inhibitor: Pepstatin
Other proteases Apopain Elastase HIV protease	 Bestatin Phosphoramidon Elastase, Soybean trypsin inhibitor

If the problem of proteolytic activity still persists, then the recovery based on fluorescence is not straightforward. This is because the cleaved GFP fragment might be fluorescent and could distort the results obtained for recovery from the standard curve of the fluorescence of the fusion protein. GFP has been used as an *in vivo* reporter of organophosphorous hydrolase (OPH) expression and cellular location in *E coli (*Wu et al., 2000). It was reported that the GFP got cleaved from the fusion protein and that the cleaved product was fluorescent. A correlation between GFP fluorescence and the sum of the intensity of the two bands on the western blot (bands of fusion and GFP by itself) was

developed. This correlation gave a better estimate of OPH quantity by fluorescence measurements. Therefore, if cleaved products that fluoresce are seen, then correlations could be made to get better estimates of recovery of the target protein. Alternatively, a second purification step (gel filtration) could be used to separate the cleaved product from the fusion protein, provided there is enough difference in the size between the two. Another option would be to use other linkers between the two proteins being fused.

In this work spiking studies were done, but spiking just gives a general idea about purification from the tobacco extract, but information about what happens to the protein as it goes through the secretory pathway cannot be established. Therefore, expression constructs need to be made to express GFP-fusion proteins and then actually determine if the protein can be recovered and if the recovered protein is stable.

5.2.4. Retention vs Secretion

From the results obtained in Chapter 4, it was found that the C-terminal HDEL retention signal plays a role in protein stability apart from its role in accumulation. This is consistent with other results. A 12-fold increase in the half-life of vicilin, a pea seed protein was seen in when expressed in transgenic lucerne leaves with a C-terminal KDEL signal (Wandelt et al., 1992). This would mean that the protein would be expressed intracellularly and therefore appropriate purification strategies need to be considered especially if the process needs to be scaled up. Ideally, if the protein is secreted into the surrounding media, then purification is much easier because cell material can be removed by vacuum filtration followed by clarification of the filtrate before initial purification steps (Fischer et al., 1999). However, for proteins like GM-CSF, which has very high therapeutic value, stability of the protein is of utmost priority. If the stability, can be

improved by HDEL, it would be more viable to chose an appropriate strategy to recover the intracellular protein, rather than have unstable protein secreted into the media. Cell disruption devices like bead mills could be used but problems due to heat generation, disruption of sub-cellular organelles accompanied by liberation of noxious chemicals (alkaloids, phenolics) arise (Fischer et al., 1999). When cells are disrupted, additives like polyvinlypolypyrolidone (PVPP), dithiothreitol (DTT), ascorbic acid, EDTA and others are added to reduce proteolytic activity and oxidation. In the case of IMAC, DTT and EDTA cannot be used because they are not compatible. A viable option would be to use enzymatic digestion to disrupt the cell wall like technical-grade pectinase (Fischer et al., 1999).

5.2.5. Chromatography

All the IMAC work in this study was done either in batch mode or by gravity. A continuous flow system like that in FPLC needs to be tested to see if the binding and elution of the fusion protein does take place from the column as the kinetics of binding may be different when the carrier solvent (both binding and elution buffers) is pumped through the column.

REFERENCES

Billinton N and Knight AW (2001) Seeing the wood through the trees: A review of Techniques distinguishing green fluorescent protein from endogenous autofluoresence. Analy. Biochem. 291:175-197

Cubitt AB, Heim R, Adams SR, Boyd AE, Gross LA and Tsien RY (1995) Understanding, improving and using green fluorescent proteins. Trends in Biochemical Sciences 20: 448-455

Fischer R, Emans N, Schuster F, Hellwig S and Drossard J (1999) Towards molecular farming in the future: using plant-cell-suspension cultures as bioreactors. Biotechnol. Appl. Biochem. 30: 109-112

Fischer R, Liao YC and Drossard J (1999) Affinity-purification of a TMV-specific recombinant full-size antibody from a transgenic tobacco suspension culture. J. Immunol. Methods 226: 1- 10

Harper BK, Mabon SA, Leffel SM, Halfhill MD, Richards HA, Moyer KA and Stewart NC (1999) Green fluorescent protein as a marker for expression of a second gene in transgenic plants. Nature Biotech. 17:1125-1129.

Heim R, Cubbit AB and Tsien RY (1995) Improved green fluorescence. Nature 373: 663-664

Hiatt A, Cafferkey R and Bowdish K (1989) Production of antibodies in transgenic plants. Nature 432: 76-78.

Kimata Y, Iwaki M, Lim C R, Kohno K (1997) A novel mutation that enhances the fluorescence of green fluorescent protein at high temperatures. Biochem Biophys Res Commun 232: 69-73.

Mason HS, Man-kit LD and Arntzen CJ (1992) Expression of hepatitis B surface antigen in transgenic plants. Proc Natl Acad Sci USA. 89: 11745-11749.

Niwa Y, Hirano T, Yoshimoto K, Shimizu M and Kobayashi H (1999) Non-invasive quantitative detection and applications of S65T-type green fluorescent protein in living plants. Plant J. 18: 455-463

Sachdev D and Chirgwin JM (1998) Order of fusions between bacterial and mammalian proteins can determine solubility in *Escherichia coli* Biochem. Biophys. Res. Commun. 244: 933

Sijimons PC, Dekker BM, Schrammeijer B, Verwoerd TC, van den Elzen PJM and Hoekema A (1990) Production of correctly processed human serum albumin in transgenic plants. Bio/Tech 8, 217-221.

Waldo GS, Standish BM, Berendzen J and Terwilliger TC (1999) Rapid protein-folding assay using green fluorescent protein. Nature Biotech. 17:691-695

Wandelt CI, Khan MRI, Craig S Schroeder HE, Spencer D and Higgins TJV (1992) Vicilin with carboxy-terminal KDEL is retained in the endoplasmic reticulum and accumulation to high levels in the leaves of transgenic plants. Plant J. 2:181-192

Wu CF, Cha HJ, Rao G, Valdes JJ, Bentley WE (2000)A green fluorescent protein fusion strategy for monitoring the expression, cellular location, and separation of biologically active organophosphorus hydrolase. Appl Microbiol Biotechnol. 54:78-83

APPENDIX 1

SEQUENCE OF GFP172

```
       CCATGGCTAGCAAAGGAGAAGAACTTTTCACTGGAGTTGTCCCAATTCTTGTTGAATTAG
1        M  A  S  K  G  E  E  L  F  T  G  V  V  P  I  L  V  E  L  D

       ATGGTGATGTTAATGGGCACAAATTTTCTGTCAGTGGAGAGGGTGAAGGTGATGCAACAT
21       G  D  V  N  G  H  K  F  S  V  S  G  E  G  E  G  D  A  T  Y

       ACGGAAAACTTACCCTTAAATTTATTTGCACTACTGGAAAACTACCTGTTCCTTGGCCAA
41       G  K  L  T  L  K  F  I  C  T  T  G  K  L  P  V  P  W  P  T

       CACTTGTCACTACTTTCACTTATGGTGTTCAATGCTTTTCAAGATACCCAGATCATATGA
61       L  V  T  T  F  T  Y  G  V  Q  C  F  S  R  Y  P  D  H  M  K

       AGCGGCACGACTTCTTCAAGAGCGCCATGCCTGAGGGATACGTGCAGGAGAGGACCATCT
81       R  H  D  F  F  K  S  A  M  P  E  G  Y  V  Q  E  R  T  I  F

       TCTTCAAGGACGACGGGAACTACAAGACACGTGCTGAAGTCAAGTTTGAGGGAGACACCC
101      F  K  D  D  G  N  Y  K  T  R  A  E  V  K  F  E  G  D  T  L

       TCGTCAACAGGATCGAGCTTAAGGGAATCGATTTCAAGGAGGACGGAAACATCCTCGGCC
121      V  N  R  I  E  L  K  G  I  D  F  K  E  D  G  N  I  L  G  H

       ACAAGTTGGAATACAACTACAACTCCCACAACGTATACATCATGGCCGACAAGCAAAAGA
141      K  L  E  Y  N  Y  N  S  H  N  V  Y  I  M  A  D  K  Q  K  N

       ACGGCATCAAAGCCAACTTCAAGACCCGCCACAACATCGAACACCATCACCATCACCATG
161      G  I  K  A  N  F  K  T  R  H  N  I  E  H  H  H  H  H  H  D

       ACGGCGGCGTGCAACTCGCTGATCATTATCAACAAAATACTCCAATTGGCGATGGCCCTG
181      G  G  V  Q  L  A  D  H  Y  Q  Q  N  T  P  I  G  D  G  P  V

       TCCTTTTACCAGACAACCATTACCTGTCCACACAATCTGCCCTTTCGAAAGATCCCAACG
201      L  L  P  D  N  H  Y  L  S  T  Q  S  A  L  S  K  D  P  N  E

       AAAAGAGAGACCACATGGTCCTTCTTGAGTTTGTAACAGCTGCTGGGATTACACATGGCA
221      K  R  D  H  M  V  L  L  E  F  V  T  A  A  G  I  T  H  G  M

       TGGATGAACTATACAAACATGATGAGCTTTAAGAGCTC
241      D  E  L  Y  K  H  D  E  L  *
```

Notes: Sequence of GFP172 (with His tag). The 6 Histidines are shown in bold. *Nco*1 and *Sac*1 sites at the N- and C-terminus are underlined. The S65T mutation in the chromophore (TYG) is shown. Since an *Nco*1 site was introduced at the N-terminus, an internal *Nco*1 site was removed by a single base change without changing the codon (boxed). Based on deletion analysis, the minimal domain required for GFP fluorescence is amino-acids 7-229 (Li et al., 1997). Therefore, any changes at the N- or C-terminus will not effect fluorescence. All mutations shown above have been named w.r.t to the original mGFP5 (Hasellof et al., 1997).

REFERENCES

Haseloff J, Siemering KR, Prasher DC and Hodge S (1997) Removal of a cryptic intron and subcellular localization of green fluorescent protein are required to mark transgenic Arabidopsis plants brightly Proc. Natl. Acad. Sci. USA Vol. 94: 2122-2127.

Li X, Zhang G, Ngo N, Zhao X, Kain SR, Huang C (1997) Deletions of the *Aequorea* green fluorescent protein define the minimal domain required for fluorescence. J of Biol. Chem. 272: 28545-28549