Purification of Recombinant Proteins in Plants Using Small-Molecule Dependent Inteins Fused to ELP or HFBI

Kira Liu, The University of Western Ontario

Supervisor: Dr. Rima Menassa, The University of Western Ontario

A thesis submitted in partial fulfillment of the requirements for the Master of Science degree in Biology

© Kira Liu 2015

Follow this and additional works at: https://ir.lib.uwo.ca/etd

Part of the Cell Biology Commons

Recommended Citation

https://ir.lib.uwo.ca/etd/3048

This Dissertation/Thesis is brought to you for free and open access by Scholarship@Western. It has been accepted for inclusion in Electronic Thesis and Dissertation Repository by an authorized administrator of Scholarship@Western. For more information, please contact wlsadmin@uwo.ca.
PURIFICATION OF RECOMBINANT PROTEINS IN PLANTS USING SMALL-MOLECULE DEPENDENT INTEINS FUSED TO ELP OR HFBI

(Thesis format: Monograph)

By

Kira Liu

Graduate Program in Biology

A thesis submitted in partial fulfillment of the requirements for the degree of Master of Science

School of Graduate and Postdoctoral Studies
The University of Western Ontario
London, Ontario, Canada

© Kira Liu 2015
Abstract

Elastin-like polypeptides (ELP) and hydrophobins (HFBI) are fusion tags which enhance the accumulation levels of recombinant proteins in plants and aid in the non-chromatographic purification of these proteins. An intein is inserted between the target protein and fusion tag to replace the protease cleavage site. In this study, transient expression of intein-GFP fused to ELP or HFBI in *Nicotiana benthamiana* yielded the full size fusion protein. The ELP-fused proteins were successfully purified using inverse transition cycling (ITC) with 1.0 M NH₄(SO₄)₂ at room temperature. However, the purification using membrane ITC was unsuccessful due to clogged membrane pores. Purification of HFBI-fused proteins through aqueous two-phase system (ATPS) using Triton X-114 was also unsuccessful. The proteins did not separate into the correct phase for recovery. In crude extract, cleavage of the intein at the C-terminus could be induced with a pH change, while DTT was not sufficient for cleavage at the N-terminus.

**Keywords:** elastin-like polypeptides, hydrophobin I, inverse transition cycling, aqueous two-phase system, intein splicing, recombinant protein production
Acknowledgements

Before we dive into the rest of the document, I would like to take a moment to thank everyone who has helped make this possible. From professional to personal contributions, these people have been here for me through this experience and I would not be here without them.

First and foremost, I would like to thank my supervisor, Dr. Rima Menassa at the Agriculture and Agri-Food Canada in London, ON, for inviting me into her lab nearly six years ago. I was only a summer student then, but she later extended the invitation to include my Masters. Her expertise in the field makes her a great supervisor and her amiable nature makes her great company. Through this journey, she has been nothing but encouraging and understanding, and her support has been invaluable. Thank you for believing in me.

I would also like to thank my co-supervisor, Dr. Susanne Kohalmi at the University of Western Ontario, whom I had the happy chance of meeting prior to the beginning of my Masters. Always there with a reassuring smile and wise words, your guidance has been incomparable.

To my advisors, Dr. Richard Gardiner and Dr. Aiming Wang, your expertise in your respective fields helped open my eyes and mind to think outside the box. To the experts Dr. Udo Conrad, Dr. Jussi Joensuu, Dr. Hoang Trong Phan and Dr. Lauri Reuter, I want to thank them for their input in how to troubleshoot and carry forward with this project.

I cannot forget the amazing people of the Menassa lab. Without the help from our lab technicians Angelo Kaldis and Hong Zhu, the lab would not run as smoothly as it does, and we, the students, would most certainly be lost without your teachings. A special thank
you goes out to Alex Molnar for making these incredible images for my thesis. And to everyone else in and around the lab, especially Dr. Reza Saberianfar, Sean Miletic, Dr. Igor Kolotilin, Mandana Miri, Dr. Alexandre Tromas and Zayn Khamis, thank you for making the work environment so enjoyable.

Finally, I would like to give my gratitude to my friends who have supported me through all this: Athena Nguyen, Celina Tsang, Dennis Choi, Dr. Frederick Lam, Luyen Nguyen, and Dr. Thomas Wong. And of course, I could not have done this without the unconditional support, encouragement and love from my mom, dad and brother.

Thank you.
Table of Contents

Abstract ........................................................................................................................................ i

Acknowledgements ................................................................................................................... ii

Table of Contents ....................................................................................................................... iv

List of Figures .............................................................................................................................. vii

List of Abbreviations ................................................................................................................ viii

CHAPTER 1: INTRODUCTION ................................................................................................... 1

1.1 Plants as a Platform for Recombinant Protein Production ............................................. 1

1.2 Transient Expression of Proteins in *Nicotiana benthamiana* .................................. 2

1.3 Fusion Tags for Increased Accumulation and Simplified Purification ...................... 3

1.4 Inteins ................................................................................................................................. 9

1.5 Intein Cleavage .................................................................................................................. 11

1.6 Hypotheses ......................................................................................................................... 16

1.7 Objectives and Approaches ............................................................................................ 17

CHAPTER 2: MATERIALS AND METHODS ............................................................................. 18

2.1 Common Media, Solutions and Reagents .................................................................... 18

2.1.1 Media and Media Components .................................................................................... 18

2.1.2 Solutions and Reagents .............................................................................................. 19

2.2 Strains and Plasmids ....................................................................................................... 23

2.2.1 Strains and Growth Conditions .................................................................................. 23

2.2.2 Plasmids ...................................................................................................................... 23

2.3 Design of Constructs and Cloning Procedures ............................................................. 24

2.3.1 Design of Intein-HFBI Constructs .............................................................................. 24
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.3.2 Cloning of HFBI Constructs</td>
<td>24</td>
</tr>
<tr>
<td>2.4 <em>Escherichia coli</em> and <em>Agrobacterium tumefaciens</em> Transformations</td>
<td>25</td>
</tr>
<tr>
<td>2.5 Transient Expression in <em>N. benthamiana</em> Leaves</td>
<td>25</td>
</tr>
<tr>
<td>2.6 Protein Extraction</td>
<td>26</td>
</tr>
<tr>
<td>2.7 Protein Quantification</td>
<td>27</td>
</tr>
<tr>
<td>2.7.1 SDS-PAGE and Staining</td>
<td>27</td>
</tr>
<tr>
<td>2.7.2 Immunoblotting</td>
<td>28</td>
</tr>
<tr>
<td>2.7.3 Fluorometry</td>
<td>28</td>
</tr>
<tr>
<td>2.8 Purification</td>
<td>29</td>
</tr>
<tr>
<td>2.8.1 Aqueous Two-Phase System</td>
<td>29</td>
</tr>
<tr>
<td>2.8.2 Inverse Transition Cycling</td>
<td>29</td>
</tr>
<tr>
<td>2.8.3 Membrane Inverse Transition Cycling</td>
<td>30</td>
</tr>
<tr>
<td>2.9 Separation and Recovery of GFP</td>
<td>30</td>
</tr>
<tr>
<td>2.9.1 Intein Cleavage</td>
<td>30</td>
</tr>
<tr>
<td>2.9.2 Cleavage of ITC-Purified ELP Fusions and Recovery of GFP</td>
<td>31</td>
</tr>
<tr>
<td>CHAPTER 3: RESULTS</td>
<td>32</td>
</tr>
<tr>
<td>3.1 Transient Expression of ELP Constructs</td>
<td>32</td>
</tr>
<tr>
<td>3.2 Design and Transient Expression of HFBI Constructs</td>
<td>35</td>
</tr>
<tr>
<td>3.3 Aqueous Two-Phase System with HFBI Constructs</td>
<td>37</td>
</tr>
<tr>
<td>3.4 Optimal Conditions for Inverse Transition Cycling</td>
<td>41</td>
</tr>
<tr>
<td>3.5 Purification of ELP Fusions Using Membrane ITC</td>
<td>43</td>
</tr>
<tr>
<td>3.6 Intein Cleavage in Crude Extract</td>
<td>44</td>
</tr>
<tr>
<td>3.7 Recovery of GFP from Cleaved Intein-ELP Fusions</td>
<td>47</td>
</tr>
</tbody>
</table>
CHAPTER 4: DISCUSSION

4.1 Transient Expression of ELP and HFBI Fusions in *N. benthamiana* ..........49

4.2 Purification of ELP-intein- and HFBI-intein-GFP Fusion Proteins...........50

4.3 Controlled Intein Cleavage in N30 and C30......................................53

4.4 Summary and Future Work....................................................................54

References..................................................................................................57

Curriculum Vitae .......................................................................................64
List of Figures

Figure 1. Schematic of inverse transition cycling and membrane inverse transition cycling.................................................................5

Figure 2. Conceptual schematic of aqueous two-phase system purification of a GFP-intein-HFBI fusion protein ..........................................................8

Figure 3. Diagram of steps involved in intein cleavage..........................................................13

Figure 4. Modifications of an intein to promote ease in separation of target protein and fusion tag.................................................................15

Figure 5. Constructs for transient expression of intein-GFP in *N. benthamiana*............33

Figure 6. Expression of ELP constructs in *N. benthamiana*........................................34

Figure 7. Transient expression of HFBI constructs in *N. benthamiana*.........................36

Figure 8. Aqueous two-phase system for HFBI constructs and pJJJ161.........................38

Figure 9. Aqueous two-phase system under visible and UV light.................................40

Figure 10. Testing of ITC conditions for the purification of C30 using (NH₄)₂SO₄........42

Figure 11. Membrane ITC with N30 and C30 at room temperature with 1.0 M (NH₄)₂SO₄.................................................................45

Figure 12. Intein cleavage treatments for crude extracts of N30 and C30 ....................46

Figure 13. Recovery of cleaved GFP through ITC, post intein cleavage ......................48
List of Abbreviations

* Standard SI units are not listed

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>4-HT</td>
<td>4-hydroxytamoxifen</td>
</tr>
<tr>
<td>APS</td>
<td>Ammonium persulfate</td>
</tr>
<tr>
<td>ATPS</td>
<td>Aqueous two-phase system</td>
</tr>
<tr>
<td>BSA</td>
<td>Bovine serum albumin</td>
</tr>
<tr>
<td>BY-2</td>
<td>Bright yellow 2</td>
</tr>
<tr>
<td>C##</td>
<td>C-terminus cleaving intein fused to GFP and ELP repeats of 30, 60 or 90</td>
</tr>
<tr>
<td>CHFBI</td>
<td>C-terminus cleaving intein fused to GFP and HFBI</td>
</tr>
<tr>
<td>DNA</td>
<td>Deoxyribonucleic acid</td>
</tr>
<tr>
<td>dpi</td>
<td>Days post infiltration</td>
</tr>
<tr>
<td>DTT</td>
<td>Dithiothreitol</td>
</tr>
<tr>
<td>EDTA</td>
<td>Ethylenediaminetetraacetic acid</td>
</tr>
<tr>
<td>ELP</td>
<td>Elastin-like polypeptide</td>
</tr>
<tr>
<td>ER</td>
<td>Endoplasmic reticulum</td>
</tr>
<tr>
<td>GFP</td>
<td>Green fluorescent protein</td>
</tr>
<tr>
<td>H5</td>
<td>Hemagglutinin subtype 5</td>
</tr>
<tr>
<td>HFBI</td>
<td>Hydrophobin I from <em>Trichoderma reesei</em></td>
</tr>
<tr>
<td>HRP</td>
<td>Horseradish peroxidase</td>
</tr>
<tr>
<td>ITC</td>
<td>Inverse transition cycling</td>
</tr>
<tr>
<td>KDEL</td>
<td>Lysine-aspartate-glutamate-leucine ER retrieval signal</td>
</tr>
<tr>
<td>KOH</td>
<td>Potassium hydroxide</td>
</tr>
<tr>
<td>LB</td>
<td>Luria-Bertani medium</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
</tr>
<tr>
<td>LBD</td>
<td>Ligand binding domain</td>
</tr>
<tr>
<td>MCS</td>
<td>Multiple cloning site</td>
</tr>
<tr>
<td>MES</td>
<td>Morpholinoethanesulfonic acid</td>
</tr>
<tr>
<td>mITC</td>
<td>Membrane inverse transition cycling</td>
</tr>
<tr>
<td>mRNA</td>
<td>Messenger RNA</td>
</tr>
<tr>
<td>N##</td>
<td>N-terminus cleaving intein fused to GFP and ELP repeats of 30, 60 or 90</td>
</tr>
<tr>
<td>NHFBI</td>
<td>N-terminus cleaving intein fused to GFP and HFBI</td>
</tr>
<tr>
<td>NOS</td>
<td>Nopaline synthase</td>
</tr>
<tr>
<td>PBS</td>
<td>Phosphate buffered saline</td>
</tr>
<tr>
<td>PBS-T</td>
<td>Phosphate buffered saline with 0.1% v/v Tween-20</td>
</tr>
<tr>
<td>PEB</td>
<td>Protein extraction buffer</td>
</tr>
<tr>
<td>PMSF</td>
<td>Phenylmethylsulfonylfluoride</td>
</tr>
<tr>
<td>PTGS</td>
<td>Post-transcriptional gene silencing</td>
</tr>
<tr>
<td>PVDF</td>
<td>Polyvinylidene fluoride</td>
</tr>
<tr>
<td>PVPP</td>
<td>Polyvinylpolypyrrolidone</td>
</tr>
<tr>
<td>RE</td>
<td>Restriction enzyme</td>
</tr>
<tr>
<td>RecA</td>
<td>Recombinase A from <em>Mycobacterium tuberculosis</em></td>
</tr>
<tr>
<td>RNA</td>
<td>Ribonucleic acid</td>
</tr>
<tr>
<td>SDS-PAGE</td>
<td>Sodium-dodecyl sulfate polyacrylamide gel electrophoresis</td>
</tr>
<tr>
<td>SPGK</td>
<td>Single peptide GFP-KDEL</td>
</tr>
<tr>
<td>TBS</td>
<td>Tris buffered saline</td>
</tr>
<tr>
<td>TBS-T</td>
<td>Tris buffered saline with 0.1% v/v Tween-20</td>
</tr>
<tr>
<td>T-DNA</td>
<td>Transfer DNA</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>------------</td>
<td>-------------</td>
</tr>
<tr>
<td>TEB</td>
<td>Tris-HCl protein extraction buffer</td>
</tr>
<tr>
<td>TEV</td>
<td>Tobacco etch virus</td>
</tr>
<tr>
<td>TGS</td>
<td>Tris-glycine saline</td>
</tr>
<tr>
<td>TSP</td>
<td>Total soluble protein</td>
</tr>
<tr>
<td>T&lt;sub&gt;t&lt;/sub&gt;</td>
<td>Transition temperature for ITC</td>
</tr>
<tr>
<td>YEB</td>
<td>Yeast extract broth</td>
</tr>
</tbody>
</table>
CHAPTER 1: INTRODUCTION

1.1 Plants as a Platform for Recombinant Protein Production

With the expanding global market for recombinant proteins in pharmaceutical and therapeutic applications, the need for cost-effective and efficient ways for the production of these proteins becomes more prominent. Sales of $36 billion in 2002 grew to $163 billion in 2012 and these numbers are expected to keep rising. In this same time span, the level of biotechnology product sales as a percentage of total drug sales rose from 8 to 19% (Evens and Kaitin, 2014). Most commercial human therapeutic proteins are produced in mammalian and bacterial expression systems (Boehm, 2007). Bacterial systems are often chosen for their simple and low-cost maintenance, but they lack several post-translational modifications that are strictly eukaryotic, such as glycosylation and disulfide bond formation (Gomord and Faye, 2004). Mammalian cultures can provide these necessary modifications. However, they require costly media components and can carry human and animal pathogens (Streatfield, 2007).

Plants have recently received attention for the production of recombinant proteins due to their numerous advantages. One advantage is that plants do not have the potential to carry animal and human pathogens. They also require less maintenance and costly media components and are more scalable than other systems such as mammalian cells (Dingermann, 2008); scalability is an important characteristic when it comes time to commercialize products. In contrast to bacterial systems, plants can also perform the necessary post-translational modifications required for many complex eukaryotic proteins (Pogue et al., 2010; Xu et al., 2011). Post-translational modifications between plants and
humans are highly conserved with only few minor differences during glycosylation (Cabanes-Macheteau et al., 1999). These differences in glycosylation are added as the protein passes through the Golgi (Gomord and Faye, 2004). To avoid passing through the Golgi, a C-terminal motif consisting of lysine, aspartic acid, glutamic acid and leucine (KDEL) is added for endoplasmic reticulum (ER) retrieval and retention (Lord and Roberts, 1998). As well, it was found that levels of accumulation for recombinant proteins increased when the proteins were retained in the ER (Menassa et al., 2001).

Although the advantages of plant platforms are many, several limitations still exist. These include low protein accumulation levels and inefficient and costly purification methods. Typical levels of recombinant protein yielded in plants range from 0.1% to 1.0% of total soluble protein (TSP; Hood et al., 2002). The most common purification procedures require the use of affinity chromatography. The regeneration of the resins for chromatography in addition to the use of proteases for the removal of a fusion tag increases the cost of purification (Chapman, 2005). In a material cost comparison study by Banki and Wood (2005), the cost to purify 1 kg of protein product with a fusion tag, including the removal of it, was upwards of $9.6 million, where approximately $8.4 million was for the cost of the protease alone. Therefore, the search for more effective and economical purification methods is ongoing.

1.2 Transient Expression of Proteins in Nicotiana benthamiana

Tobacco has become a favourite plant for recombinant protein production because of several distinct benefits including a large amount of leaf biomass, simple transformation and established transient expression methods currently used commercially. The relative of
tobacco, *Nicotiana benthamiana*, is the main model used for transient expression of recombinant proteins through infiltration of *Agrobacterium* carrying the gene of interest into the leaves of the plant. This method is more commonly called “agroinfiltration”. Infection begins when *Agrobacterium tumefaciens* containing a binary expression vector is infiltrated into the leaves of the host plant by vacuum infiltration or syringe injection (Kapila *et al.*, 1997). Single-stranded T-DNA then moves from *Agrobacterium* into the plant cells, where it is trafficked into the nucleus *via* chaperones (Zambryski, 1988). Although much of the T-DNA does not integrate into the host chromosomes in the short time frame of transient transformation, the free T-DNA copies are still transcriptionally active (Voinnet *et al.*, 2003). This allows for tissue samples to be harvested 2-5 days post infiltration, whereas with stable transformation, regeneration of transgenic plant lines can take up to 3-6 months depending on the species (Kapila *et al.*, 1997; Sparkes *et al.*, 2006). Also, as expression is not dependent on transcriptional activity of a chromosome insertion site, insertion position effects are not observed. As well, multiple genes can be expressed simultaneously within the same host (Kapila *et al.*, 1997). This can be used to achieve production of complex proteins such as antibodies. The speed of transient expression also allows screening of multiple genes in a short time span.

### 1.3 Fusion Tags for Increased Accumulation and Simplified Purification

Advances for purification of proteins without the use of chromatographic methods include the addition of fusion tags to a protein of interest. Elastin-like polypeptides (ELP) and hydrophobin I (HFBI) are both fusion tags which have been shown to increase accumulation of recombinant proteins (Patel *et al.*, 2007; Conley *et al.*, 2009; Joensuu *et
al., 2010), while aiding in non-chromotographic purification. ELPs are a synthetic polypeptide composed of an elastin-derived repeating pentapeptide motif: VPGXG, where X is any amino acid except proline (Urry et al., 1985; Urry, 1992). They have the property of being able to undergo a reversible inverse phase transition (Urry 1992; 1997) and this property can be transferred to a fused partner (Meyer and Chilkoti, 1999). With this knowledge, purification of ELP-fused proteins could be carried out using a simple method called inverse transition cycling (ITC; Fig. 1). This method works by raising the temperature above the ELP’s transition temperature (T_t) and/or increasing the salt concentration changes the ELP’s conformation and causing aggregation of the ELP fusion protein. The ELP aggregates can then be isolated and resolublized in an aqueous solution at a temperature lower than the T_t. According to Urry et al. (1985), when phase transition is induced, aggregation does not denature the fusion protein as only the ELP portion aggregates. Also, these aggregates have a radius larger than one micron (Meyer et al., 2001; Meyer and Chilkoti, 2002a), allowing these fusion proteins to be captured by membrane microfiltration ITC (mITC; Ge et al., 2006). All these properties give ELP-fused proteins a special advantage in biotechnological fields where low cost and high efficiency large scale purification is desirable.

Transition temperature is mainly dependent on two factors: the concentration of the ELP fusion protein and the solution composition, especially its ionic strength. As the concentration of ELP increases, the T_t decreases logarithmically (Meyer and Chilkoti, 2002b). Increasing the concentration of NaCl, the most commonly used salt, also decreases the transition temperature. As a rule of thumb, the T_t is lowered by ~15°C.
Figure 1. Schematic of inverse transition cycling and membrane inverse transition cycling. In centrifugation ITC (A) and membrane ITC (B), heat and/or salt is added to induce aggregation of the ELPs. The aggregates are then isolated either by centrifugation (A) or membrane filtration (B) and resolubilized prior to intein cleavage. To complete the purification process, after cleavage, aggregation of the ELP is induced again to separate out the ELPs from the protein of interest.
for every 1 M NaCl increase. However, the efficiency of the aggregation and precipitation of the ELP fusion proteins also depends on the effects of the ions in the solution. This effectiveness follows the Hofmeister series, which ranks ions based on their kosmotropic nature and their effect on protein solubility (Cacace et al., 1997). Kosmotropic ions, or ions that contribute to the stability of water-water interactions, exceed chaotrophic ions, which disrupt the hydrogen bonding network between water molecules, in their ability to precipitate proteins. Chloride and sodium are both ranked in the middle of the Hofmeister series. Fong et al. (2009) conducted a comparative study of different salts under the assumption that a more kosmotropic salt such as ammonium sulfate would be a better precipitant of proteins than sodium chloride. Their results showed that ammonium sulfate concentrations could be used at 27% of sodium chloride to obtain the same effect on ELP precipitation. As well, this reduces the Tc from 37°C to about 23°C, eliminating the need for additional energy to heat the solution.

The levels of accumulation and efficiency of purification of ELP-fused proteins also depends on the size of ELP, as determined by the number of pentapeptide repeats. More efficient purification can be achieved with longer ELP segments, but results in lower levels of protein accumulation (Conley et al., 2009). Conversely, shorter ELP segments generally result in higher accumulation levels but less efficient purification. Currently, the efficiency of purification of a protein of interest from plant extracts using shorter ELPs is low and affinity chromatography methods are still required to obtain high purity (Joensuu et al., 2010). With ELP110 fused to either β-lactamase or to green fluorescent protein (GFP) expressed in E. coli resulted in recoveries of 80% and 50%, respectively, demonstrating that the fusion partner may also influence purification efficiency (Fong et al., 2009).
Hydrophobins are globular proteins produced by filamentous fungi that are able to form layers at water-air surfaces (Whiteford and Spanu, 2002). In this study, I used hydrophobin I (HFBI), which originates from *Trichoderma reesei* and is about 7.5 kDa in size. Although there are different classes of hydrophobins, one common characteristic is the eight Cys residues that form four intramolecular disulfide bridges, which increase the stability of the protein structure (Hakanpaa *et al.*, 2004). Their unique amphipathic property allows hydrophobins to be purified using a liquid-liquid extraction in an aqueous two-phase system (ATPS) and this property can be transferred to their fused partners (Linder *et al.*, 2004; Lahtinen *et al.*, 2008). Purification using ATPS allows the phase separation of hydrophobic and hydrophilic proteins through the addition of a surfactant (Fig. 2). When HFBI-fused proteins form micelles, the fusion protein is pulled into the surfactant (hydrophobic) phase. Isobutanol is added to remove the surfactant and the purified fusion protein settles in the aqueous phase. Recovery of GFP fused to HFBI with ATPS resulted in about 91% (Joensuu *et al.*, 2010).

Despite the improvements to accumulation levels and purification methods, the addition of fusion tags has its disadvantages. If a fusion tag is used, it must be cleaved prior to any usage where a purified untagged product is required. The Tobacco Etch Virus (TEV) protease site consisting of seven amino acids ENLYFQG is commonly inserted between a protein of interest and its fused tag to allow for separation following purification. Cleavage with the TEV protease occurs between the Q and G, leaving either six (ENLYFQ) or one amino acids (G) on the resulting fragments. The potential issue with extra amino acids is demonstrated in recombinant antibody/antigen production, where a single amino acid can
Figure 2. Conceptual schematic of aqueous two-phase system purification of a GFP-intein-HFBI fusion protein. (A) In ATPS, the first phase separation is induced by the addition of a surfactant. Isobutanol is added to the micelle-rich (surfactant) phase to remove the surfactant. In the second phase separation, the purified fusion protein is separated from the isobutanol-surfactant mixture. (B) The purified fusion protein is then treated with a reagent to induce intein cleavage. Following cleavage, GFP is isolated from intein-HFBI by the addition of surfactant. Modified from Yan (2011).
alter the entire interaction between antigen and antibody (Colman, 1994). The cost of using a protease site for the removal of a fusion tag is very high and large scale purification using this method is, therefore, infeasible. As an alternative, the TEV protease (or any other protease) site can be replaced by an intein, and the N- and/or C-terminal integrity can be maintained when the intein is spliced. This means that the termini regain a normal peptide group (i.e. an amine or carboxyl group), leaving no residual amino acids.

1.4 Inteins

In a comparative study in 1987 between Daucus carota (carrot) and Neurospora crassa vacuolar ATPases and a putative calcium ion pumping ATPase, the first intein was discovered (Shih et al., 1988). This putative ATPase was isolated when it was found that a mutation in its gene made yeast resistant to trifluoperazine. When this gene was compared to the vacuolar ATPase subunits, it was found that the beginnings and the ends of the sequences were similar, but the middle portion contained an endonuclease-like insertion. Hirata et al. (1990) isolated cDNA for a vacuolar ATPase from Saccharomyces cerevisiae and found a sequence similar to the trifluoperazine resistant gene inserted within the gene. Interestingly, the molecular weight of the yeast V-ATPase was found to be the same as a subunit which contained no insertion. Kane et al. (1990) showed that this insertion was included in the mRNA and translated, but spliced post-translationally.

Inteins are found in all three domains of life: bacteria, eukaryotes and archaea. Few inteins have been identified in eukaryotes and the majority of inteins discovered have been found in bacteria and archaea. They get their name from “internal” or “interrupting” proteins and they function analogously to introns in RNA. However, unlike introns, inteins
are translated with their host protein prior to splicing. There are many studies which suggest that, since there is similarity between the endonuclease motifs of inteins and mobile introns, inteins may be involved in intein homing just as mobile introns are involved in intron homing (Belfort et al., 1995; Colston and Davis, 1994; Cooper and Stevens, 1995).

The vast majority of reported inteins consist of two domains: an endonuclease domain and a self-splicing domain where the C- and N-termini have the necessary characteristics to allow for splicing (Liu, 2000). Smaller inteins called mini-inteins consist of only the self-splicing domain. Found in almost 80% of known inteins, the most common endonuclease motif is a dodecapeptide LAGLIDADG/HNH (Perler, 2000). Although the deletion of the endonuclease does not affect protein splicing in the large inteins (Chong and Xu, 1997), it is proposed that the endonuclease plays an important role in the spreading of inteins (Liu, 2000). This supports the idea of inteins being mobile genetic elements.

Although inteins are found in diverse proteins such as metabolic enzymes, proteases, and DNA and RNA polymerases, most are found in enzymes related to DNA replication and repair (Liu, 2000). Chute (1998) gives two reasons why inteins appear to be biased towards these types of host proteins. The first stems from the theory of intein homing. Intein homing means that in a cell which has an intein-containing allele and an intein-less allele, the endonuclease will identify and cleave the intein-less allele at the insertion/homing site. The double-stranded DNA breaks and triggers the repair system. This allows the intein to be inserted into the intein-less allele using the intein-containing allele as template. To prevent cleavage of the new intein-containing DNA, the insertion is made at a homing site which disrupts the original endonuclease recognition site (Liu, 2000). Chute (1998) suggests that inteins found in host proteins related to DNA replication
and repair are more adapted for intein homing. The host protein and intein are produced at the same time, ensuring that the endonuclease is produced during periods of active DNA replication and repair. This means that when a double-stranded DNA break occurs at the homing site and the intein is introduced to the intein-less allele, replication and repair proteins are able to fix the break promptly and complete the homing process.

The second explanation why inteins are biased towards replication and repair proteins is the potential hazard the endonuclease has on the host cell’s chromosome. If the intein and the host protein are not yet an established pair, then a new intein may accidentally recognize and break the chromosome at regions other than the intended homing site. Chute (1998) suggests that the cell’s response to reduce this risk is to restrict the production of intein endonucleases only to times of active DNA replication and repair. This can minimize and quickly rectify the damage done to the chromosome without affecting chromosome integrity and cell viability.

### 1.5 Intein Cleavage

After translation, the intein splices and leaves the two parts of the host protein (exteins or “external proteins”) to rejoin *via* a peptide bond and resume normal function (Cooper and Stevens, 1995; Dujon, 1989; Perler *et al.*, 1994). A quick overview of intein splicing (Chong *et al.*, 1997) is as follows (Fig. 3):

- **Step 1.** At the N-terminus, a spontaneous N-O or N-S acyl shift forms an ester or thioester intermediate. This will lead the N-extein to bind to the oxygen of the serine or sulfur of the cysteine residue at the N-terminus of the intein.
Step 2. The nucleophilic residue at the C-terminus will attack the (thio)ester bond during transesterification, cleaving it. This leaves a branched protein intermediate.

Step 3. Immediately following transesterification, the asparagine at the C-terminus undergoes cyclization, cleaving the intein through a succinimide intermediate.

Step 4. A spontaneous N-S acyl shift reforms the peptide bond between the exteins.

Because of their self-catalyzed splicing nature, inteins are used in biotechnological applications as a substitute to proteases but it can still result in unwanted *in vivo* splicing prior to purification. Hence, a regulated intein was designed for controlled splicing. The *Mycobacterium tuberculosis* RecA intein was modified by inserting a human estrogen receptor ligand binding domain (LBD) to disrupt intein folding and prevent splicing (Buskirk *et al.*, 2004). This LBD has a high affinity for the synthetic molecule 4-hydroxytamoxifen (4-HT). When 4-HT is bound to the LBD, intein structure is restored and splicing can occur. This engineered intein was termed a “small-molecule dependent” intein. However, Buskirk *et al.* (2004) found that the insertion of control splicing at either the C-terminus or the N-terminus. By changing asparagine (Asn) to alanine (Ala) at the C-terminus of the intein, asparagine cyclization is prevented, thereby inhibiting splicing at that terminus (Fig. 4C). Substituting Ala for cysteine (Cys) at the N-terminus disrupts the N-S acyl shift required for splicing the intein from the N-extein (Evans *et al.*, 2005; Fig. 4B). A third modification was designed to enhance the rate of the LBD did not result in ligand-dependent splicing but only reduced *in vivo* splicing activity.
Figure 3. Diagram of steps involved in intein cleavage. Cleavage starts with a spontaneous N-S acyl shift, allowing the intein to undergo a transesterification to displace the N-extein onto the C-extein. Following that, cyclization of asparagine at the C-terminus cleaves the joined exteins, forming succinimide in the process. Finally, an O-N acyl shift forms the peptide bond and the intein is successfully cleaved from the two exteins. Modified from Mathys et al. (1999).
To adapt this intein to recombinant fusion protein production and purification in plants, Hendy (2009) modified the intein further. Terminal amino acids were modified to cleavage at the C-terminus. This was achieved by replacing the aspartic acid (Asp) at position 394 with glycine (Gly) of the C-intein to prevent hydrogen bonding at the C-terminus, allowing for premature succinimide formation. In addition to these changes, the two exteins (Fig. 4A) were replaced by an intact protein of interest at one terminus and a fusion tag at the other. Hendy (2009) experimented with the three different intein constructs expressed in tobacco BY-2 (bright yellow) cell suspensions: the N-intein for N-terminal cleavage, the C-intein for C-terminal cleavage, and the S-intein for enhanced C-terminal cleavage, each tagged with an ELP 30, 60 or 90 pentapeptide repeats in length.

Research from Chong et al. (1997) showed that by adding a reducing agent such as dithiothreitol (DTT), the N-S equilibrium can be shifted by attacking the thioester and inducing the N-terminal cleavage. Wood et al. (1999) discovered that the His before the Asn at the C-terminus allowed for pH-based cleavage at that terminus. Histidine also acts as a proton donor to promote the formation of succinimide, thereby resulting in C-terminal cleavage. Hendy (2009) noted that 4-HT was not required in the splicing of the inteins, but a change in pH to 5.5 was sufficient for C-terminal cleavage. No cleavage at the N-terminus was observed. In a study by Linder et al. (2001), different pH of extraction buffers were tested for determining efficiency of purification by ATPS. They found that pH did not significantly hinder the partitioning of HFBI during ATPS. Therefore, a change in pH when cleaving the intein should not affect the overall purification process of HFBI. There is no literature for the effects of DTT on ATPS as of yet. However, since HFBI is
Figure 4. Modifications of an intein to promote ease in separation of target protein and fusion tag. (A) The original intein with the native Cys and Asn terminal amino acids. The intein is flanked by the N-extein and C-extein. (B) N-cleaving intein: the Asn at the N-terminus is replaced by Ala and the two exteins are replaced by the target protein and the fusion tag. Addition of 50 mM DTT aids in cleavage of the target protein at the N-terminus from the intein-fusion tag. (C) C-cleaving intein: the Cys at the C-terminus is replaced by Ala and the two exteins are replaced by the fusion tag and target protein. A change in the pH to 5.5 results in the cleavage of the target protein from the intein-fusion tag complex.
held together by four disulfide bonds, it is expected that DTT would reduce these bonds and unfold the HFBI protein.

Using all of this background knowledge, a protein fused to an intein and either ELP or HFBI is predicted to result in increased accumulation levels and easier purification. The simple steps of both ITC (Fig. 1) and ATPS (Fig. 2) should allow for quick purification of the fusion protein, followed by cleavage of the intein-tag fusion and one additional step to segregate the protein of interest from the tag. This project intends to test the functionality and efficiency of intein-HFBI and intein-ELP fusions in the purification of the target protein, GFP, when transiently expressed in *Nicotiana benthamiana*. This could create a more rapid and cost-efficient method for purifying recombinant proteins from crude plant extracts.

1.6 Hypotheses

Based on previous literature about the expression of GFP-ELP and GFP-HFBI in plants, GFP-HFBI with an accumulation level of up to 51% TSP (Joensuu *et al.*, 2010) outperforms the 21% TSP accumulation level of GFP-ELP$_{10}$ (Conley *et al.*, 2009). As the intein incorporated between the GFP and the fusion tag (ELP or HFBI) is very large, it can be predicted to accumulate at lower levels than a fusion protein without the intein. As well, based on purification recovery reported by these two studies, the 91% recovery of GFP-HFBI through ATPS (Joensuu *et al.*, 2010) greatly surpasses the 60% recovery of GFP-ELP$_{20}$ using ITC (Conley *et al.*, 2009). Therefore, I hypothesize that the GFP-intein constructs fused to HFBI will accumulate at higher levels and result in higher recovery of the fusion protein than when fused to ELP.
Preliminary results for intein cleavage conditions for this modified small-molecule dependent intein were previously tested in BY-2 tobacco cells (Hendy, 2009). It was discovered that cleavage of GFP could be induced by a change in pH to 5.5 for the C-terminus cleaving intein. However, cleavage at the N-terminus could not be induced. I hypothesize that both cleavage conditions will work for their respective inteins but the efficiency of the N-terminus cleaving intein with DTT will not be as high as for the C-terminus cleaving intein with a pH change.

1.7 Objectives and Approaches

1. Determine accumulation levels of intein-ELP constructs (Hendy, 2009) using transient expression in *N. benthamiana* through agroinfiltration.

2. Design GFP-intein-HFBI constructs and determine the accumulation levels in *N. benthamiana* through agroinfiltration.

3. Determine optimal conditions for purification of ELP-fused and HFBI-fused proteins using ITC and ATPS, respectively.

4. Determine efficiency of intein cleavage in crude extract of ELP and HFBI fusion proteins for removal of intein-ELP or intein-HFBI.

5. If successful, fusion proteins will be purified, cleaved and GFP will be recovered *via* another partial round of the respective purification.
CHAPTER 2: MATERIALS AND METHODS

2.1 Common Media, Solutions and Reagents

2.1.1 Media and Media Components

Antibiotics

Ampicillin (100 mg/ml) and kanamycin (100 mg/ml) stocks were prepared by dissolving the antibiotic in milliQ water and filter sterilizing. Rifampicin stock (10 mg/ml) was made by dissolving antibiotic in 100% methanol and filter sterilizing. Aliquots were stored at -20°C.

Gamborg’s Solution

Per 100 ml: 0.32 g Gamborg’s B5 + vitamins (Research Products International Corp., G20200), 2 g sucrose, 1 ml MES (1 M), 100 µl acetosyringone (200 mM).

Infiltration Medium

Per 100 ml: 100 ml YEB, 1 ml MES (1 M), 50 µl acetosyringone (200 mM), 50 µl kanamycin (100 mg/ml), 100 µl rifampicin (10 mg/ml).

200 mM Acetosyringone

Per 100 ml: 0.392 g acetosyringone was dissolved in ethanol. Aliquots were stored at -20°C.
1 M Morpholinoethanesulfonic Acid (MES) pH 5.6

Per 100 ml: 19.50 g MES was dissolved in milliQ water. Adjusted pH to 5.6 with KOH. Aliquots were stored at -20°C.

Luria-Bertani (LB) Broth

Per L: 10 g NaCl, 10 g tryptone and 5 g yeast extract were added to milliQ water. Adjusted pH to 7.0. For solid media, a pre-mixed LB agar mixture (QIAgen) was used.

Yeast Extract Broth (YEB)

Per L: 5 g beef extract, 1 g yeast extract, 5 g peptone, 5 g sucrose and 2 mM MgSO₄ were added to milliQ water. Adjusted pH to 7.2. For solid media, 14 g of agar was added.

2.1.2 Solutions and Reagents

10X Phosphate Buffer Solution (PBS)

Per L: 80 g NaCl, 2 g KCl, 14.4 g Na₂HPO₄ and 2.4 g KH₂PO₄ was dissolved in milliQ water. The pH was adjusted to 7.4.

1X PBS-Tween (PBS-T)

Per L: 100 ml 10X PBS was diluted in 899 ml milliQ water and 1 ml Tween-20 was added.

50% Glycerol (v/v)

Per 100 ml: 50 ml glycerol was added to 50 ml milliQ water. Solution was filter sterilized and stored at room temperature.
Protein Extraction Buffer (PEB)

Per 10 ml: 9.6 ml 1X PBS-T, 0.2 g polyvinylpolypyrrolidone (PVPP), 20 µl EDTA (0.5 M), 100 µl PMSF (100 mM), 10 µl leupeptin (1 mg/ml), 0.2 g sodium ascorbate. PVPP was allowed to hydrate for 2-24 h before use. Leupeptin and PMSF were added immediately before use.

0.5 M Ethylenediamine-tetraacetic acid (EDTA) pH 8.0

Per 50 ml: 9.3 g EDTA disodium salt dehydrate was added to 30 ml milliQ water. NaOH was used to adjust pH to 8.0 and topped to 50 ml with milliQ water. Solution was stored at room temperature.

1 mg/ml Leupeptin

Per 100 ml: 0.1 g leupeptin was dissolved in milliQ water. Aliquots were stored at -20°C.

100 mM Phenylmethanesulfonyl Fluoride (PMSF)

Per 100 ml: 1.74 g PMSF was dissolved in 100 ml isopropanol. Aliquots were stored at -20°C.

50 mM Tris-HCl, 0.1% v/v Triton X-100 Extraction Buffer (TEB)

Per L: 6.07 g Trizma-base was dissolved in MilliQ water. The pH was adjusted to 8.0 and 1 ml Triton X-100 was added. The buffer was supplemented with 100 µl PMSF (100 mM), 10 µl leupeptin (1 mg/ml).
Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis (SDS-PAGE) Solutions

5X Reducing SDS Sample Buffer
Per 10 ml: 3 ml 1 M Tris-HCl pH 8.0, 0.5 g SDS, 5 ml 50% glycerol, 1 ml DTT (1M), 5 mg Phenol Red were added to milliQ water. Aliquots were stored at -20°C.

10% Ammonium Persulfate (APS)
Per 1 ml: 0.1g ammonium persulfate was dissolved in milliQ water.

10% Sodium Dodecyl Sulfate (SDS)
Per 100 ml: 10 g SDS was added to milliQ water and stored at room temperature.

0.5 M Tris-HCl pH 6.8 (Stacking Gel Buffer)
Per 100 ml: 6 g Tris base was dissolved in milliQ water. The pH was adjusted to 6.8 with HCl and stored at room temperature.

1.5 M Tris-HCl pH 8.8 (Separating Gel Buffer)
Per 100 ml: 18.15 g Tris base was dissolved in milliQ water. The pH was adjusted to 8.8 with HCl and stored at room temperature.

5X Tris-Glycine SDS (TGS) Running Buffer
Per L: 15 g Tris base, 72 g glycine and 5 g SDS were added to milliQ water and stored at room temperature.
1X TGS Running Buffer
Per L: dilute 100 ml 5X TGS Running Buffer in 900 ml milliQ water and stored at room temperature.

Immunoblot Solutions

Western Transfer Buffer (10% Methanol)
Per L: 3.03 g Tris base, 24.41 g glycine were added to milliQ water, topped to volume of 800 ml. Then 200 ml methanol was added.

5% Blocking Solution
Per 50 ml: 2.5 g Carnation skim milk power was dissolved in 1X TBS-T and stored at 4°C.

10X Tris-Buffered Saline (TBS)
Per L: 24.2 g Tris base and 175.3 g NaCl were added to milliQ water. The pH was adjusted to 7.5 and stored at room temperature.

1X TBS-Tween (TBS-T)
Per L: 100 ml 10X TBS and 1 ml Tween-20 were added to 899 ml milliQ water and stored at room temperature.
Intein Cleavage Solutions

4-Hydroxytamoxifen (4-HT)

Per 5 mg 4-HT (Sigma-Aldrich, H7904-5MG): dissolve in 250 µl 95% ethanol at 50°C. Stock concentration was diluted to 10 mM with 95% ethanol and stored in the dark at -20°C. The working concentration is 10 µM.

1 M Dithiothreitol (DTT)

Per 100 ml: 15.45 g DTT was dissolved in milliQ water. Aliquots were stored at -20°C.

2.2 Strains and Plasmids

2.2.1 Strains and Growth Conditions

*Escherichia coli* strain XL1-blue (Agilent Technologies, Cat. No. 200301) was used for the amplification of plasmid DNA. The bacteria were grown at 37°C on LB plates or in liquid LB broth supplemented with antibiotics. *Agrobacterium tumefaciens* strain EHA105 (Hood *et al.*, 1993) was chosen for the infiltration of the gene of interest into *N. benthamiana*. *Agrobacterium tumefaciens* was grown at 28°C on YEB plates or in YEB medium supplemented with appropriate antibiotics.

2.2.2 Plasmids

The cloning vector used in this study was pBluescript II KS+ (Strategene, Cat. No. 12207). The order of restriction enzyme cut sites in the multiple cloning site (MCS) made it an ideal candidate for the purpose of cloning. It is also a high copy plasmid, thus optimal
in amplifying plasmid DNA. The construct containing the gene of interest was then cloned into pCaMterX (Harris and Gleddie, 2001), the plant binary vector used for plant transformation. This vector consists of a double enhanced constitutive cauliflower mosaic virus 35S promoter, a multiple cloning site and a nopaline synthase (NOS) terminator (Odell et al., 1985).

2.3 Design of Constructs and Cloning Procedures

2.3.1 Design of Intein-HFBI Constructs

Based on the design of the intein-ELP constructs, the elastin-like polypeptide was replaced by a hydrophobin I sequence. Differing from the intein-ELP constructs, (GGGS)₃ and/or Aspergillus niger glucoamylase G1 linkers (SGSVTSTSKTTATASKTSTST) were placed between the HFBI and the intein to allow for more space for the HFBI to fold properly.

These constructs were synthesized commercially by Eurofins Genomics in the pUC57 vector. The resuspended plasmid DNA was used to transform E. coli cultures which were grown overnight at 37°C in an orbital shaker at 250 rpm.

2.3.2 Cloning of HFBI Constructs

Plasmid DNA from the overnight E. coli cultures were isolated using the QIAprep Spin Miniprep Kit (QIAGen, Cat. No. 27106). After plasmid DNA was digested with the appropriate restriction enzymes (RE; Invitrogen and New England Biolabs) to isolate the construct, it was inserted into one of two vectors: pBluescript II KS+ for plasmid DNA
amplification; or pCaMterX for agroinfiltration. Ligation was mediated by the T4 DNA ligase (Invitrogen, Cat. No. 15224-017) following the manufacturer’s protocol.

2.4 *Escherichia coli* and *Agrobacterium tumefaciens* Transformation

Constructs were transformed into electrocompetent *E. coli* XL1-blue or *A. tumefaciens* EHA105 cells using electroporation. A 0.2 cm Gene Pulser Cuvette (Bio Rad) was placed on ice prior to transformation. Once cooled, 1 µl of plasmid was added to 50 µl of electrocompetent cells and transferred to the cuvette. Following the manufacturer’s instructions, an electrical pulse was administered to the cuvette containing the cell suspension. One millilitre of medium was immediately added to the cuvette and transferred to a sterile 1.5 ml Eppendorf tube. The tube was incubated at 37°C for *E. coli* or at 28°C for *A. tumefaciens* in a horizontal shaker at 250 rpm for one hour. A portion of the cells (50 and 100 µl) were then plated onto LB or YEB plates with the appropriate antibiotics.

Positive transformants were identified and grown in 5 ml of LB media containing the appropriate antibiotic. Stocks were created using 750 µl of the overnight culture and 250 µl of 50% glycerol, mixed in a cryogenic vial and flash frozen in liquid nitrogen before being stored at -80°C.

2.5 Transient Expression in *N. benthamiana* Leaves

Plates were streaked using the *A. tumefaciens* stock cultures and grown at 28°C for two nights. Colonies were picked and grown in liquid medium (YEB supplemented with the appropriate antibiotics) overnight at 28°C. This culture was used to inoculate the infiltration liquid culture. These cultures were grown overnight in infiltration medium at
28°C in an incubator shaker at 250 rpm. Cells were pelleted at room temperature at 3000 rpm for half an hour. The supernatant was discarded and the pellet was resuspended in Gamborg’s solution to an OD₆₀₀ = 0.6. This cell suspension was incubated for an hour at room temperature.

After incubation, the *Agrobacterium* suspension containing the gene of interest was mixed at equal volumes with an *Agrobacterium* suspension of p19, a suppressor of post-transcriptional gene silencing (PTGS). The p19 is from the *Cymbidium* ringspot virus (Silhavy *et al*., 2002). Because transcription levels of foreign DNA are reduced due to PTGS, it is necessary to co-infiltrate the gene of interest with a suppressor of PTGS (Voinnet *et al*., 1999). This allows for increased expression of recombinant proteins through agroinfiltration (Voinnet *et al*., 2003). This mixture was infiltrated into the leaves with a needleless 1mL syringe (Santi *et al*., 2008). Using a needle, a scratch was made on the underside of the leaf and the mixture was infiltrated from that point. Samples (whole leaves or leaf disks using a size 4 cork borer, approx. 8 mm in diameter) were collected four days post infiltration (dpi). Leaf disks were placed into 2 ml Eppendorf tubes with three zircon beads (BioSpec Products Inc., Cat. No. 11079125z) while whole leaf tissue was placed into a plastic bag. Samples were flash frozen using liquid nitrogen and stored at -80°C.

2.6 Protein Extraction

Method A: Leaf disks

Samples were homogenized using the Tissue Lyser (QIAGEN) for 60 s, turned over and homogenized for another 60 s. The homogenized tissues were centrifuged at
4000 rpm for a minute. Then 200 µl of protein extraction buffer was added per three leaf disks. This mixture was vortexed for 15 s and centrifuged down at 14,000 x g for five min at 4°C. The supernatant was transferred to a new 1.5 ml tube and centrifuged again. The supernatant was transferred to a new tube for further analysis. The TSP in the samples was quantified using the Bradford assay (Bradford, 1976).

Method B: Whole leaf tissue

For experiments requiring larger amounts of crude extract, this method was used. Samples were homogenized manually using a pre-cooled mortar and pestle. The samples were removed from the freezer and quickly crushed. A pinch of sand was added to ease the grinding. Protein extraction buffer was added in a 1 g tissue to 3 ml buffer ratio. Samples were centrifuged at 15,500 x g for 15 min at 4°C. The supernatant was transferred to a new tube and centrifuged, then transferred to a new tube for further analysis. Total soluble protein was determined by the Bradford assay.

2.7 Protein Quantification

2.7.1 SDS-PAGE and Staining

TSPs were separated by sodium dodecylsulfate-polyacrylamide gel electrophoresis (SDS-PAGE; Laemmli, 1970) on a 4% stacking and 10% separating SDS-PAGE gel. Gels were run at 120 V until the gel front reached the bottom of the gel.

Gels were stained for an hour in GelCode Blue (Thermo Scientific) before destaining overnight in milliQ water. Known concentrations of bovine serum albumin (BSA) were run alongside the samples for quantification.
2.7.2 Immunoblotting

Gels were transferred to a PVDF membrane (Bio Rad, Cat. No. 162-0184) using a Trans-blot Semi-dry Transfer Cell (Bio Rad) or a Trans-Blot Turbo Transfer System (Bio Rad). Membranes were blocked overnight at 4°C in a 5% blocking solution. Afterwards, the membranes were incubated in 0.5% blocking solution with the primary antibody, monoclonal mouse anti-GFP (Living Colours, Cat. No. 632381), at a 1:5000 dilution at room temperature for an hour. Membranes were washed prior to incubation with the secondary horseradish peroxidase (HRP)-conjugated goat anti-mouse antibody (Bio Rad, Cat. No. 170-6515) in a 1:5000 dilution for another hour. The membrane was detected with ECL chemiluminescent detection solutions (GE Healthcare, Cat. No. RPN 2106) according to the manufacturer’s protocol and visualized on developing film or by a MicroChemi chemiluminescent system (Neutec Group Inc., Farmingdale, USA).

2.7.3 Fluorometry

Using a 96-well black Costar system (Corning, Cat. No. 07-200-627), samples expressing GFP were quantified using fluorometry instead of the traditional immunoblotting. Standards were made from a sample with a known concentration of GFP-HFBI in 1X PBS. A dilution series for each sample was prepared, usually a 1/10 and 1/50 dilution. For each sample, three technical replicates were analyzed. The plate(s) were placed inside a Synergy 2 Microplate Reader (BioTek Instruments Inc., Winooski, USA). Excitation and emission were set at 485 nm and 516 nm, respectively. The results from fluorometry were recorded as the % GFP/TSP.
2.8 Purification

2.8.1 Aqueous Two-Phase System

Following protein extraction (section 2.6) with PEB, samples were incubated at 23°C for 10 min. The warmed crude extract was added to a clear tube with 4% w/v ratio of surfactant (Triton X-114 or Agrimul NRE 1205). The samples were incubated for an hour at room temperature or until distinct phases were seen, up to two hours. Samples were then centrifuged at 3000 rpm for 15 min. The surfactant phase was extracted and transferred to another tube. Isobutanol was added to the surfactant phase at a 10:1 ratio to the amount of surfactant added, left to settle, and centrifuged at 3000 rpm for 15 min before the purified protein layer was recovered.

2.8.2 Inverse Transition Cycling

Following protein extraction (section 2.6) with TEB, the extract was added to a tube containing pre-weighed (NH$_4$)$_2$SO$_4$ in concentrations varying from 0.2 M to 1.0 M. The salt was dissolved prior to incubation at one of three temperatures for 30 min: 23°C, 30°C and 37°C. After incubation, the sample was centrifuged at 20,000 x g for 40 min at the appropriate temperature (i.e. same as incubation). After removal of the supernatant from the “hot spin”, the pellet was resuspended in 1/8 of the sample volume with cold TEB. On ice, the pellet was resuspended using the Sonic Dismembrator Model 100 (Fisher Scientific, Ontario, Canada) at a setting of 5 for 15 s twice. This was placed on a horizontal shaker at 4°C for 30 min to allow for resolubilization. The sample was then centrifuged for 10 min at 20,000 x g at 4°C. The purified product from the “cold spin” was collected. The pellet was resuspended in 1/8 of the sample volume in TEB for analysis.
2.8.3 Membrane Inverse Transition Cycling

After protein extraction (2.6) with TEB, the extract was pre-filtered using vacuum filtration with 0.22 µm PVDF Durapore Membrane Filters (Millipore Ltd, Cat. No. GVWP04700). Ammonium sulfate was added to the extract at the correct concentration and dissolved completely before continuing onto the next step. The samples were incubated at room temperature for 30 min before being filtered again. This step leaves the ELP-fused proteins trapped above the membrane. The membrane was then washed with ammonium sulfate of the same concentration and temperature. Elution by gravity at 4°C was performed as an overnight step with half the sample volume of cold TEB to obtain the purified product.

2.9 Separation and Recovery of GFP

2.9.1 Intein Cleavage

Crude extract samples were treated with different cleaving agents according to the intein. The C-terminus cleaving intein samples were subjected to three conditions: pH change to 5.5, addition of 4-HT to a concentration of 10 µM, and a combination of both. For the N-terminus cleaving intein samples, the three conditions were: addition of DTT to a concentration of 50 mM, addition of 4-HT to a concentration of 10 µM, and a combination of both. A negative control of just the crude extract was also analyzed. All samples were incubated overnight (18-24 hours) in a water bath at room temperature. Samples were properly mixed prior to taking a portion for SDS-PAGE. After determination of the most successful treatment, intein cleavage of the purified samples was induced using that agent.
2.9.2 Cleavage of ITC-Purified ELP Fusions and Recovery of GFP

After determining the conditions necessary for intein cleavage, ELP fusion proteins were purified using ITC and subjected to the conditions described above (section 2.9.1) for the respective intein. Following the overnight incubation period for intein cleavage, another round of ITC was performed to separate the intein-ELP tag and recover the cleaved GFP. Aggregation of the intein-ELP fragments were induced by the addition of (NH₄)₂SO₄ to 1.0 M at room temperature for 30 min. The sample was centrifuged at 20,000 x g for 40 min at room temperature to separate the intein-ELP pellet and aqueous GFP. Samples taken pre- and post-ITC were analyzed using immunoblots.
CHAPTER 3: RESULTS

This study aimed to find an easier and more cost efficient way to purify a recombinant protein of interest from crude plant extract. By utilizing fusion tags such as ELPs and HFBI and replacing protease cleavage sites with an intein, transient expression of the recombinant proteins was expected to result in increased accumulation levels as well as simpler purification of the target protein.

3.1 Transient Expression of ELP Constructs

It was previously shown that short ELP tags of 40 repeats or less allow the production of higher levels of recombinant protein, while larger ELP tags of 100 repeats or more allow for more efficient purification of the fusion protein (Conley et al., 2009). Therefore, constructs containing three different lengths of ELP fused to GFP and intein (Fig. 5A, 5B) were designed and cloned into the pCaMterX vector. The resulting expression vector was then transformed into electrocompetent EHA105 A. tumefaciens.

The constructs were infiltrated into N. benthamiana plants for transient expression via agroinfiltration. Samples were collected four days post infiltration and extracted using TEB supplemented with leupeptin and PMSF to prevent protein degradation. A construct containing GFP with a KDEL ER retrieval motif (SPGK, Fig. 5F) was used as positive control. Immunoblot analysis of tissue expressing the N-terminus-cleaving intein constructs (N30, N60 and N90; Fig. 5A), revealed that only the N30 construct produced a detectable protein band (Fig. 6A). The accumulation level of this fusion protein
Figure 5. Constructs for transient expression of intein-GFP in *N. benthamiana*.

(A) and (B) represent the ELP constructs. (C) and (D) represent the HFBI constructs NHFBI and CHFBI, respectively. (E) represents the HFBI-GFP construct, pJJJ161. (F) represents the positive control SPGK. Common to all constructs are the tCUP translational enhancer, the pr1b secretory signal peptide, the protein of interest (GFP), and a KDEL motif for ER-retrieval. A StrepII tag is incorporated for alternative purification purposes. The TEV in (E) represents the protease cleavage site. In (A) and (B), the “n” represents the number of ELP pentapeptide repeats, and in this study, n = 30, 60 or 90 (shown above is n = 30). The linkers used in (C) and (D) are L1, a flexible linker (GGGS)₃ and/or L2, an *A. niger* glucoamylase G1 linker.
Figure 6. Expression of ELP constructs in *N. benthamiana*. (A) Immunoblot with two replicates confirming expression of the GFP-intein-ELP constructs with the N-terminus-cleaving intein and the various sizes of ELP. (B) Immunoblot with two replicates showing expression of the C-terminus-cleaving inteins fused to GFP and various sizes of ELP. These were hybridized with an anti-GFP antibody. The expected sizes for ELP$_{30}$, ELP$_{60}$ and ELP$_{90}$ constructs are 90 kDa, 102 kDa, and 114 kDa, respectively. SPGK was used as positive control; the expected size is 27 kDa.
was 0.40 ± 0.01% of TSP as determined by fluorometry against a concentration curve of purified GFP. For the C-terminus-cleaving intein constructs (C30, C60 and C90), both C30 and C60 produced detectable protein bands by immunoblot analysis (Fig. 6B). The accumulation levels for these two constructs were 0.44 ± 0.01% GFP/TSP and 0.19 ± 0.01% GFP/TSP, respectively. For the purpose of this study, only the higher accumulating N30 and C30 were chosen for use in ITC optimization.

3.2. Design and Transient Expression of HFBI Constructs

The constructs used in this portion of the study were designed similarly to the ELP constructs. The ELP tag was replaced with the HFBI tag and its corresponding linker(s). The HFBI tag was flanked by one (for NHFBI) or two (for CHFBI) linkers to reduce chances of improper folding by the intein and hydrophobin. The two linkers were (GGGS)$_3$ linker (Fig. 5; L1) and the A. niger glucoamylase G1 linker (Fig. 5; L2). Joensuu et al. (2010) had shown these two linkers to aid in the proper folding of HFBI. Because these constructs were synthesized commercially, it was desirable to reduce the number of restriction digest sites to allow for correct orientation in one step. For this reason, only the EcoRI and the BamHI sites were incorporated into the gene (Fig. 5C and D) and were used for directional cloning into the binary vector pCaMterX.

The HFBI constructs were transiently expressed in N. benthamiana plants using agroinfiltration. Samples (leaf disks or whole leaf tissue) were collected four dpi and extracted using PEB supplemented with protease inhibitors leupeptin and PMSF. Accumulation levels were found to be 0.27 ± 0.02% GFP/TSP for NHFBI and 0.12 ± 0.01% GFP/TSP for CHFBI (Fig. 7) through fluorometry.
Figure 7. Transient expression of HFBI constructs in *N. benthamiana*. Samples were collected four days post infiltration and analyzed through immunoblotting. The blots were hybridized with an anti-GFP primary antibody. The expected sizes for NHFBI and CHFBI are 86 kDa and 88 kDa, respectively. The size of spliced GFP is 27 kDa. The number corresponds to the volume of extract (µl) loaded in each lane.
3.3. Aqueous Two-Phase System with HFBI Constructs

To study the purification of the HFBI constructs using ATPS, plant tissue samples were collected four dpi and proteins were extracted with PEB* (PEB minus leupeptin and PMSF). The crude extract was then subjected to ATPS purification with the surfactant Triton X-114 as described in section 2.8.1. To determine the success of the purification, samples were collected throughout the process (refer to Fig. 2A): the crude extract (E); the aqueous layer (A) after the addition of surfactant; and the purified product in the second aqueous phase after the removal of surfactant by the addition of isobutanol (P). These samples were analyzed by immunoblotting. The results showed the hydrophobin fusions (NHFBI and CHFBI) were not detected in the aqueous phase nor in the purified product (Fig. 8A). To determine if this result was due to a technical error or to the inability of this protein to purify using ATPS, a GFP-HFBI construct called pJJJ161 was used as a positive control. This was previously shown to accumulate at very high levels and to purify very well by ATPS (Joensuu et al., 2010). Aqueous two-phase system carried out with GFP-HFBI showed that most of the protein was recovered in the purified fraction (data not shown). Purification using another surfactant, Agrimul NRE 1205, did not lead to a separation in phases following a two hour incubation period (data not shown).

It was possible that the GFP-intein-HFBI proteins were not purified by ATPS since the accumulation levels were 100 times lower than GFP-HFBI, which is shown to accumulate at 20-40% TSP (Joensuu et al., 2010; Saberianfar et al., 2015). To examine if a higher concentration of HFBI could help pull the HFBI fusions into the surfactant
Figure 8. Aqueous two-phase system for HFBI constructs and pJJJ161. Immunoblots detected with anti-GFP antibody were used to determine the phases in which GFP and GFP fusions resided during ATPS with Triton X-114. (A) ATPS with NHFBI and CHFBI alone. (B) ATPS under spiked conditions: WT (wild type), NHFBI and CHFBI extracts were mixed with equal volumes of pJJJ161 extract, labeled 161, N161 and C161, respectively. The expected size of pJJJ161 is 35 kDa (arrowhead). The expected sizes of NHFBI and CHFBI are 86 kDa and 88 kDa (arrow). E = crude extract; A = aqueous (non-surfactant) phase; P = second aqueous phase (purified product).
phase, extracts of tissue infiltrated with the HFBI fusion constructs were spiked with equal volumes of extracts of tissue infiltrated with GFP-HFBI (pJJJ161) and the effectiveness of spiking was compared among the different fusion proteins. The wild type (WT) extract spiked with an equal volume of GFP-HFBI was used as a control. Results showed that pJJJ161 was successfully recovered in the purified fraction by ATPS (Fig. 8B), although it was not 100% efficient as some of the protein was still detected in the aqueous phase. However, when the HFBI constructs were spiked with GFP-HFBI, although the aqueous phase showed similar results to WT extract spiked with GFP-HFBI, the purified fractions did not detect any of the intein-HFBI fusion constructs. All three samples (Fig. 8B) showed a band at 60-75 kDa, possibly a dimer or trimer formed from either the GFP or GFP-HFBI. Thus, the addition of more HFBI did not help increase the odds of successful purification of HFBI constructs NHFBI and CHFBI.

In an attempt to track where the HFBI fusions were located throughout the steps of ATPS, photos were taken under UV light during the purification process. The visual difference in fluorescence intensity between WT and pJJJ161 under UV light can be seen in Figures 9A (right) and 9B (right), respectively. With the addition of surfactant, it became difficult to track GFP-HFBI because chlorophyll was also pulled into the surfactant phase (Fig. 9C). However, the addition of isobutanol successfully separated GFP-HFBI from chlorophyll and other proteins (Fig. 9D). While the mixed samples (N161 and C161) also showed fluorescence in the purified phase, the results from the immunoblot (Fig. 8B) showed that this fluorescence was from GFP-HFBI as opposed to the intein-HFBI fusions. In conclusion, the NHFBI and CHFBI fusion proteins were unable to be purified via ATPS.
Figure 9. Aqueous two-phase system under visible and UV light. Images were taken under visible (left) and UV light (right). (A) Wild type extract before addition of surfactant. (B) Crude extract of equal volumes of pJJJ161 and WT. (C) Phase separation after the addition of Triton X-114 to pJJJ161/WT. Upper layer is the non-surfactant layer. (D) Lower phase of (C) was removed and isobutanol was added. Phase separation shows GFP-HFBI fusions in the lower phase and other plant proteins in the upper phase.
3.4 Optimal Conditions for Inverse Transition Cycling

To isolate the ELP fusions from the crude extract, ITC was used to salt out the ELP aggregates prior to resolubilization of the fusion proteins. Using 0.4 M (NH₄)₂SO₄ at 37°C as a starting point, which are the conditions set out by Duvenage et al. (2013), this study was aimed to optimize both the temperature and salt concentration needed to obtain the highest recovery of the ELP fusion protein, as well as the purified GFP after intein cleavage. The process of ITC was carried out in the following steps: the crude extract (E) was subjected to a salt and heat treatment and was centrifuged; the supernatant (S1) was then removed – S1 should be void of any fusion protein; the pellet was resuspended in a cold buffer to resolubilize the fusion protein; after it was centrifuged again, the supernatant containing the solubilized fusion protein (S2) was recovered and the remaining pellet was resuspended (P) for analysis. The starting conditions of 0.4 M (NH₄)₂SO₄ at 37°C outlined by Duvenage et al. (2013) resulted in the loss of the fusion protein into the pellet (Fig. 10A), which could indicate that the salt concentration was too high, and possibly denatured the fusion protein in the extract. Therefore, the salt concentration was decreased to 0.2 M. I found that most of the fusion protein was still in the pellet fraction under reduced salt concentration, although a small amount of protein was now present in S2. This result suggests that lowering the salt concentration itself was not enough to improve resolubilization of this particular ELP fusion at 37°C. To examine if the temperature at 37°C may have caused irreversible precipitation, the same salt conditions (0.2 M and 0.4 M ammonium sulfate) were tested at 30°C (Fig. 10B) and 23°C (Fig. 10D). As the temperature was reduced to 30°C, then to 23°C, the amount of
Figure 10. Testing of ITC conditions for the purification of C30 using (NH₄)₂SO₄.

Immunoblots of (A) ITC performed at 37°C, (B) and (C) ITC performed at 30°C, and (D) and (E) ITC performed at 23°C. Blots were hybridized with an anti-GFP antibody. All lanes were loaded with the equal volumes of sample. E = crude extract; S1 = supernatant after hot spin; S2 = supernatant after cold spin; P = resolubilized pellet after completion of ITC.
unaggregated protein remaining in S1 increased and the band in S2 was fainter than in S1 at 0.2 M salt, and fainter than in the pellet at 0.4 M salt. These results meant the conditions were not enough to induce full aggregation of the fusion proteins. With the focus of industrial purposes in mind, treatments with increased concentrations of salt were performed only at 30°C and 23°C to decrease the amount of heat that would be required. Similar results were noted when performing ITC at the three temperatures with the various salt concentrations with N30 (data not shown).

At increased salt concentrations of 0.8 M and 1.0 M, complete aggregation of the ELP fusion protein was seen, as noted in the absence of a band in the S1 phases (Fig. 10C and E). At 23°C, the purification efficiency showed that ~50% of the ELP fusion protein was recovered in the S2 fraction and the other half remained insoluble in the pellet. For industrial purposes, purification at room temperature is more desirable and energy efficient than at higher temperatures. Thus, the conditions at room temperature and with 1 M (NH₄)₂SO₄ were used in the next study: purification through membrane ITC.

3.5. Purification of ELP Fusions Using Membrane ITC

Membrane ITC (Fig. 1B) is a modified version of the regular centrifugation ITC (Fig. 1A). It also uses salt and heat to precipitate the fusion proteins. However, instead of centrifuging the aggregates, the sample is passed through a membrane to catch the aggregates while allowing other proteins to pass through. Using the conditions set out in section 3.4, aggregation of the ELP fusions for the pre-filtered crude extract was induced at room temperature with 1.0 M (NH₄)₂SO₄. The initial filtering of the extract showed no loss in the protein of interest (data not shown). Filtration of the aggregated ELP fusions
showed complete aggregation (Fig. 11). The wash through showed none of the fusion protein, as expected. Elution by gravity of the purified fraction was carried out overnight at 4°C in 50 mM Tri-HCl with 0.1% v/v Triton X-100. A very minimal amount of the purified and resolubilized ELP fusion was observed in the P1 fraction. Much of the fusion protein was found in the P2 fraction for C30, unable to pass through the membrane. This fraction was recovered by taking a small volume of elution buffer and washing the membrane with it. Although it seems only a small amount of unsolubilized protein was retained in the P2 fraction of N30, the amount of protein detected in P1 and P2 does not sum up to the initial amount found in the extract. This indicates that the fusion proteins were possibly trapped in the pores of the membrane, thus not making mITC a feasible choice for these particular proteins.

3.6 Intein Cleavage in Crude Extract

Although purification with mITC and ATPS both proved unsuccessful, intein cleavage could still be induced in crude leaf extracts of N30 and C30 for ITC purposes. The N30 cleaves at the N-terminus and this splicing is mediated by the addition of DTT to 50 mM and/or 4-HT. A pH change to 5.5 and/or the addition of 4-HT induces cleavage at the C-terminus of the C30 construct. The efficiency of intein splicing was determined by the amount of cleaved GFP detected after the 18-24 hour incubation period. Intein cleavage of N30 with 50 mM DTT was detected by shown to be very inefficient. Cleavage with 4-HT and DTT/4-HT was not detected (Fig. 12). The effect of the pH change, without the addition of 4-HT, in C30 samples was sufficient for inducing cleavage of
Figure 11. Membrane ITC with N30 and C30 at room temperature with 1.0 M (NH₄)₂SO₄. Immunoblot of mITC were hybridized with anti-GFP antibody. After inducing aggregation of the ELPs, the suspension was passed through a 0.22 µm PVDF membrane to capture all the aggregates while allowing background plant proteins to pass through. Cold 50 mM Tris-HCl with 0.1% v/v Triton X-100 was used for resolubilization of ELP fusions. The expected sizes of N30 and C30 are 90 kDa with purified product in P1. Equal volumes of sample were loaded into each lane. E = crude extract; F = filtrate after incubation in (NH₄)₂SO₄; W = wash with 1.0 M (NH₄)₂SO₄; P1 = purified product left overnight to elute; P2 = product left on top of membrane.
Figure 12. Intein cleavage treatments for crude extracts of N30 and C30. An immunoblot detected with anti-GFP antibody was used to determine efficiency of intein cleavage. Samples were incubated for 24 hours in a water bath at room temperature. The conditions set for N30 were 50 mM DTT alone (D), 10 µM 4-HT (HT) alone and a combination of both. For C30, the cleavage conditions included a change in pH to 5.5, addition of 10 µM 4-HT and a combination of the two. Control treatments = hyphen. Equal volumes of crude extract (E) and samples of different treatments were loaded into each well.
GFP from the intein-ELP fusion tag. The treatment with 4-HT alone could not induce as much splicing as with a pH change. Thus, only a pH change is required for the cleavage of GFP from the C30 fusion protein. Crude extract with no treatments for the same incubation period showed no detectable cleavage.

3.7 Recovery of GFP from Cleaved Intein-ELP Fusions

The final step was to string together every successful method to create a complete purification process. I first used centrifugation-based ITC to purify N30 and C30 extracts. Then, based on the results from intein cleavage in crude extract, cleavage was induced by adding DTT to 50 mM for the purified N30 fraction, while cleavage for the C30 was induced by changing the pH to 5.5. After the overnight incubation period, a second round of ITC was performed at room temperature by adding (NH₄)₂SO₄ to a concentration of 1.0 M. From the immunoblots, spliced GFP was not detected in the expected aqueous phase (Fig. 13; lane “S3”). This showed that the intein cleavage step was not efficient enough to cleave the GFP from the intein-ELP tag (Fig. 13; lane “pH” or “D”). Therefore, the second round of ITC was only able to purify the fusion protein in the resulting pellet (Fig. 13; lane “P2”).
Figure 13. Recovery of cleaved GFP through ITC, post intein cleavage. Immunoblots of A) N30 and B) C30 through ITC, 24 hour intein cleavage with the respective cleaving reagents and recovery of spliced GFP via another round of ITC, using anti-GFP antibody. The expected sizes of N30, C30 and spliced GFP were 86 kDa, 88 kDa and 27 kDa, respectively. Equal volumes of sample were loaded into each lane. E = crude extract; S1 = supernatant after hot spin; S2 = supernatant after cold spin; P1 = resuspended pellet after cold spin; D/pH = DTT or pH treatment for intein cleavage; S3 = recovered cleaved GFP after second hot spin; P2 = resuspended pellet after hot spin.
CHAPTER 4: DISCUSSION

4.1 Transient Expression of ELP and HFBI Fusions in *N. benthamiana*

This study aimed to purify recombinant proteins fused to intein-ELP and intein-HFBI fusion tags produced through transient expression using agroinfiltration in *N. benthamiana*. Following the successful purification of the recombinant proteins, the fusion tag was cleaved with the help of inteins. This provides a method that is more cost effective than traditional purification strategies such as chromatography, and also reduces the need for expensive proteases required for the cleavage of the tag at a protease site. Green fluorescent protein is often used as a biological marker due to its stability and its ability to be fused to another protein without alteration of function (Zimmer, 2002). The plants were able to produce both ELP- and HFBI-GFP fusions though the accumulation levels varied between the different constructs. Consistent with the work of Conley *et al.* (2009), it was found that the fusion proteins containing fewer repeats of ELP (i.e. 30 repeats) resulted in higher accumulation levels than those with 60 or 90 ELP repeats, which accumulated at much lower levels or not at all (Fig. 6). The HFBI constructs containing the intein accumulated at much lower levels than if only GFP or GFP-HFBI was expressed. Accumulation levels for these are reported to be 18% and 38% GFP/TSP, respectively (Joensuu *et al.*, 2010).

As stated by Gräslund *et al.* (2008), in a study performed with *E. coli*, the expression levels of larger proteins versus smaller ones in a soluble form was greatly reduced. This could explain the lack of the larger fusion proteins, which range from 90 kDa to 114 kDa, during detection through immunoblotting – the proteins either accumulated at
very low levels or they were lost in the insoluble fraction of the crude extract. However, for this particular study, the focus is on the purification of these fusion proteins and not the accumulation levels of these proteins. Therefore, as N30 and C30 both accumulated to typical levels (Hood et al., 2002) of 0.4-0.5% GFP/TSP, these two constructs were chosen for ITC purification.

4.2 Purification of ELP-intein- and HFBI-intein-GFP Fusion Proteins

In this study, GFP was fused to intein-ELP or intein-HFBI, allowing the fusion to take on the unique property of each tag. Fusion to the ELP tag passed on the ability to perform a reversible aggregation/solubilization as required in the ITC purification process. Traditional ITC calls for the use of NaCl. However, Duvenage et al. (2013) found that they could precipitate their protein of interest at 37°C with 0.4 M (NH₄)₂SO₄. This agrees with the results of Fong et al. (2009) who improved their previous methods (which used NaCl) by testing various combinations of the Hofmeister salts and their efficiency in the precipitation of their ELP-tagged proteins. They found that they could achieve the same result with a lower T and a lower molarity of (NH₄)₂SO₄ versus using a higher molarity of NaCl at a higher temperature. Thus, the starting condition used for ITC was 37°C with 0.4 M (NH₄)₂SO₄. The temperatures and salt concentrations were experimented with until a feasible combination was determined. It was found that ITC could be performed using 1.0 M (NH₄)₂SO₄ at room temperature (23°C; Fig. 10E). This yielded both complete aggregation of the fusion protein as well as ~50% recovery of the resolubilized protein. Fong et al. (2009) also reported recovery of approximately 50% for their ELP-fused GFP.
This improvement is useful for industrial purposes as high concentrations of sodium chloride can corrode stainless steel and other equipment used in manufacturing. As well, because ammonium sulfate-based precipitation systems are common practice, integrating this method is not farfetched. In addition, because this method works at room temperature, it reduces the amount of energy required to make this purification process work, thus reducing the overall cost.

Although Phan and Conrad (2011) reported promising findings for the use of the purification method of membrane ITC, this method does not appear to be appropriate for every protein. They were able to recover the majority of their protein of interest. However, following their methods, I was unable to obtain the same results. During the overnight elution, very little of the ELP fusions were resolubilized. Washing the membrane did not reveal the missing unsolubilized proteins. This could be due to the proteins being caught in the membrane pores, unable to pass through or to be released. The difference between Phan and Conrad’s (2011) proteins and the ELP fusions I studied is the presence of the intein. This leads me to believe that the intein, whether it be shape, size, or another characteristic, could be the reason why the ELP fusion proteins could not pass through the membrane.

Aqueous two-phase system is a purification system that allows separation of proteins into different phases. Because of the amphipathic nature of hydrophobins and their capability for transferring this property to their fusion partner, the whole fusion protein is able to undergo the process of ATPS upon the addition of a surfactant. Hydrophobins assemble into micelles, pulling the fused protein along with them into the surfactant phase, while most other proteins remain in the aqueous phase. Keeping this in mind, the improper
folding of the four disulfide bridges can disrupt the stability of the molecule (Hakanpaa et al., 2004) and result in the hydrophobins not forming proper micelles that can be pulled into the surfactant phase. In this study, the designing of the constructs took this into consideration with the addition of linkers between the intein and HFBI tag to allow for proper folding of HFBI. However, my results showed that the fusion proteins are not pulled into the surfactant phase after the addition of Triton X-114 (Fig. 8A). This could be the result of the very large intein (~50 kDa) interfering with the folding process, suggesting that the linker(s) may not have been long enough to give the HFBI adequate space to fold properly.

Another possibility is that a minimum accumulation level of HFBI may be needed to induce micelle formation. Protein bodies are one of several specialized protein storage organelles that are commonly found in seeds (Khan et al., 2012). Formation of protein bodies is thought to help increase accumulation levels by storing large amounts of recombinant proteins in a stable and safe environment (Conley et al., 2011). Gutiérrez et al. (2013) found that protein bodies only form when the accumulation levels of recombinant protein are over the threshold of 0.2% per TSP. The accumulation level of the positive control GFP-HFBI is much higher than those of the HFBI constructs (NHFBI and CHFBI) used in this study, which were 0.27% and 0.12% GFP/TSP, respectively. Therefore, like the formation of protein bodies, there is the possibility that there must be enough HFBI proteins, in other words, high enough accumulation level, so that the proteins can interact with each other to form micelles, and the levels achieved by the constructs in question were too low for this to happen. To test if this theory was true, the HFBI constructs were spiked with pJJJ161 (GFP-HFBI) to see if the increased concentration of HFBI aids
in micelle formation. The results from this experiment showed that only the GFP-HFBI was recovered in the purified fraction and the CHFBI and NHFBI are not be detected beyond the crude extract (Fig. 8B).

Although the above reasons are all plausible, there is also the possibility that HFBI is simply not compatible with these particular protein fusions. The accumulation levels and purification processes of ELP and HFBI were compared in a recent study of ELP and HFBI by Phan et al. (2014). Because of their known association with enhancing the accumulation levels of recombinant proteins and their ease during purification (Conley et al., 2009; Joensuu et al., 2010), the fusion tags were added to a sequence coding for hemagglutinin subtype 5 (H5). The results from Phan et al. (2014) showed that while the ELP fusions showed a significant increase in expression levels compared to H5 alone, the addition of HFBI did not affect the expression level at all. Because these levels were very low (less than 0.05% per TSP), purification by ATPS was not attempted. This further supports the idea that a certain accumulation level of HFBI fusion proteins is needed to induce the formation of micelles.

4.3 Controlled Intein Cleavage in N30 and C30

The controlled induction of splicing of a protein of interest and its fusion tag mediated by an intein can be a more cost effective method since this separation can be catalyzed by common reagents rather than expensive proteases. When undergoing the process of intein splicing, the reducing agent DTT attacks the thioester bond at the N-terminus, thus inducing cleavage (Chong et al., 1997). C-terminal cleavage is determined by the penultimate His residue which acts as a proton donor in the presence of
an acidic pH change. This exchange of protons induces cyclization of asparagine, forming succinimide as the intermediate (Wood et al., 1999). Therefore, as predicted, addition of DTT and a pH change were enough to induce cleavage in the crude extracts of N30 and C30, respectively (Fig. 12). However, cleavage at the N-terminus was not as efficient as the pH change and was barely detected. Similar results were noted by Hendy (2009) where only cleavage at the C-terminus by pH change was detected when he performed a similar experiment using extracts from BY-2 cells. This being said, when working with ITC-purified N30 and C30, the efficiency of either intein cleavage was not high enough and after the second ITC, only the intact fusion protein was detected.

This particular intein was modified to cleave only in the presence of 4-HT, but was discovered to only reduce the amount of in vivo splicing (Buskirk et al., 2004). My results also indicate, similar to those of Hendy (2009), that the addition of this synthetic estrogen-like molecule did not enhance cleavage at either terminus (Fig. 12). This could be due to the intein not being dependent on the 4-HT to cleave since one terminal residue in each intein was altered. This waives the requirement of 4-HT for the proper conformational change in the intein, thus rendering it ineffective.

4.4 Summary and Future Work

Plants can be used as an alternative platform for the production of recombinant proteins. They are cheaper to maintain because they do not require the same amount of care and nutrients as mammalian cultures, but at the same time, they can perform complex post-translational modifications required for many of these eukaryotic proteins. Nevertheless, there are drawbacks to plants. They have low accumulation levels for many proteins as
well as low recovery during purification. Fusion tags such as ELP and HFBI can counter the low accumulation levels often found in plants, as well as aid in simpler forms of purification of the recombinant proteins. In this study, GFP fused to intein-ELP and intein-HFBI were transiently expressed in *N. benthamiana* and accumulated at levels of 0.2% to 0.5% GFP/TSP. The tags must then be cleaved to yield a purified product. An *M. tuberculosis* RecA intein was modified to cleave only in the presence of the small molecule 4-HT (Buskirk *et al.*, 2004) and then further modified to cleave only at one terminus (Hendy, 2009). This intein’s role in purification is to reduce the need for a protease cleavage site, which would otherwise result in very high costs. Contrary to my hypothesis, mITC and ATPS did not seem to work with the specific proteins designed for this study, but ITC showed promising results. About 50% of the fusion protein was recovered in a soluble form at room temperature with a lower concentration of ammonium sulfate than would be required with the more commonly used sodium chloride. As well, inducing intein splicing with simple reagents such as DTT and HCl resulted in the cleavage of GFP from the fusion tag and intein. These low cost reagents can be an advantage in an industrial setting. Confirming my hypothesis, the pH cleavage of C30 resulted in more efficient cleavage than N30 with DTT. However, the efficiency of intein cleavage was not enough to induce complete removal of the intein-ELP tag and therefore, during the second round of ITC, cleaved GFP was not detected. Thus, between decreased accumulation levels and poor cleavage results to obtain the purified protein, modifications would have to be made to allow this intein to be biotechnologically effective.

From this study, it appears that the added estrogen receptor site is not necessary for the cleavage of a protein of interest at either end of the intein. As stated by Buskirk *et al.*
(2004), it was found that the receptor site only decreased the splicing frequency but did not moderate the splicing of the intein. Because the termini of this intein were altered so that it would only cleave on one end, as well as the fact that 4-HT does not play a role in splicing, removing the estrogen receptor can be tested to see if a smaller intein with these modified termini results in higher accumulation levels. This would result in a mini intein gene about 600 bp in length versus the one used, which was over 1.2 kb. In turn, this could possibly increase the accumulation levels of the intein fusions and result in more efficient splicing.

In regard to hydrophobins and ATPS, it appears that HFBI is not folding properly, so modifications would have to be made to the construct itself. For instance, if the intein’s folding is affecting HFBI’s folding, then a longer linker, or more repeats of the GGGS linker, can be added to ensure more space between the two to increase chances of proper folding. This could also increase the success of ATPS downstream as proper micelles would form. As well, HFBI can be replaced with another hydrophobin to determine if it was specifically this hydrophobin’s folding that is incompatible with the intein’s.
References


Gutiérrez SP, Saberianfar R, Kohalmi SE and Menassa R. 2013. Protein body formation in stable transgenic tobacco expressing elastin-like polypeptide and hydrophobin fusion proteins. BMC Biotechnol 13, 1-10


Hendy B. 2009. Recombinant protein purification in plants utilizing small-molecule dependent inteins and elastin-like polypeptides. Masters thesis. Department of Biology, University of Western Ontario


Urry DW, Trapane TL and Prasad KU. 1985. Phase-structure transitions of the elastin polypentapeptide-water system within the framework of composition-temperature studies. Biopolymers 24, 2345–2356


Voinnet O, Pinto YM and Baulcombe D. 1999. Suppression of gene silencing: A general strategy used by diverse DNA and RNA viruses of plants. PNAS 96, 14147-14152


Yan R. 2011. “Expression of active fungal xylanases in *N. benthamiana* for hemicellulose degredation”. Masters thesis. Department of Biology, University of Western Ontario

Curriculum vitae

Education

Sept. 2012 – Present  Master of Science in Biology – Expected graduation date: August 2015 from the University of Western Ontario, London, Canada

Sept. 2008 – April 2011  Honours Bachelor of Science – University of Toronto, Scarborough, Canada

Teaching and Work Experience

Sept. 2012 – Dec. 2014  Teaching Assistant with the University of Western Ontario
BIOL 1001/1002, 1201/1202 – General Biology I & II
BIOL 2290 – Scientific Methods in Biology

May – Sept. 2010  Summer Student (FSWEP) at Agriculture and Agri-Food Canada, London, Canada

June – Sept. 2009  Summer Student (FSWEP) at Agriculture and Agri-Food Canada, London, Canada

Awards

2012 – 2014  Western Graduate Research Scholarship