Optimization of ELP-intein Purification System

OPTIMAZATION OF ELP-INTEIN PROTEIN PURIFICATION SYSTEM BY COMPARING FOUR DIFFERENT SELF-CLEAVAGE ELP FUSION PROTEINS

BY

HAN LIU, M. A. SC.

A Thesis

Submitted to the School of Graduate Studies

In Partial Fulfillment of the Requirements

For the Degree

Master of Applied Science

McMaster University

MASTER OF APPLIED SCIENCE (2008) McMaster University

(Chemical Engineering) Hamilton Ontario

TITLE: Optimization of ELP-Intein Protein Purification System By Comparing Four Different Self-Cleavage ELP Fusion Proteins

AUTHOR: Han Liu, M.A.S. (Chinese Academy of Sciences)

SUPERVISOR: Dr. Carlos Filipe

NUMBER OF PAGES: xiii, 85

ABSTRACT

Proteins vary tremendously in many of their physical and chemical properties. In order to perform in vitro application or analysis, one protein must be separated from other cellular components. This process is called protein purification. With the advances of modern science and technology, many protein purification schemes have been developed. Among them, the ELP-intein protein purification system has recently attracted an increasing amount of attention because of its positive characteristics: it is simple, inexpensive, scalable, with a high throughput, protease-free, etc. However, although the scientific literature reports all those good aspects of the system, several bad responses to it still exist. In this thesis, through comparing expression and purification of four different self-cleavage ELP fusion proteins, we propose a general solution to these problems for the first time. This makes a significant contribution to increased utility of the method of protein purification using self-cleavable stimulus responsive tags.

When ELP-intein fusion proteins are expressed in bacteria, formation of non-native cytoplasmic aggregates (inclusion bodies) is a common problem which affects the yield of target protein. Inclusion bodies are generally assumed to contain misfolded or partially folded protein through exposure of hydrophobic patches and the consequent intermolecular interactions. Despite a loss of total expression yield and the need for more time, culturing at a lower temperature was reported to promote the expression of genes into soluble proteins and alleviate IB formation. Directly

motivated by previous reports, we have applied a low-temperature expression strategy to solve the problems in this research. As expected, most of the T_4-ELP inclusion bodies disappeared, and were transformed into a soluble expression, when culturing at lower temperatures.

Inverse transition cycling (ITC), as the core method for the system we investigated has proved successful in the past with proteins that were expressed to high levels. However poor level ELP-intein tagged protein expression happens from time to time. It is hypothesized that if an ELP tagged molecule is present in a solution at a very low concentration, adding an excess amount of free ELP to the sample would form hybrid aggregates via the interaction of ELP moieties of the two molecules. We used this efficient and reversible capture system for low yield recombinant protein purification, and found it is perform very well.

ACKNOWLEDGEMENTS

First of all, I would like to express my great appreciation to my supervisor, Dr. Carlos Filipe, for everything he has done for me during the past two years; in particular, for his excellent guidance and valuable suggestions on my research project, his support and encouragement, and his great help on my thesis revision. From him, I have learned a lot, thanks to both his knowledge and his personality. I don't think I could possibly have made it without his supervision and understanding.

I also want to thank all the people who contributed to the completion of this thesis. Thanks especially to Dr. Daniel Yang, from the Department of Biochemistry at McMaster University, for his help in providing the plasmids used in this research, and for his technical support. Many of the members in my lab provided some form of assistance in obtaining the results reported here. In particular, gratitude is due to Xin Ge, Kun Zhang, Jessie Xu and Nilesh Patel. There are so many people in the Chemical Engineering Department that I want to thank, for their kindness that made the past two years a memory that I shall cherish for my whole life.

Last but not the least, a special thanks is reserved for my family: my mother Chunxian Cao, and my father Shijie Liu, for their love and selfless support throughout my pursuit of this degree. Their advice and patience were precious to me. So I hereby dedicate this thesis to them.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATION

aa	Amino acid
Asn (N)	Asparagine
ATP	Adenosine triphosphate
ATPs	Aqueous two-phase system
CBD	Cellulose-binding domain
CNBr	Cyanogen bromide
DLS	Dynamic light scattering
DNA	Deoxyribonucleic acid
dNTP	Deoxynucleotide triphosphate
DSL	Dynamic light scattering
DTT	Dithiothreitol
ELP	Elastin-like polypeptide
GFP	Green fluorescent protein
Gln (Q)	Glutamine
Gly (G)	Glycine
GST	Glutathione S-transferase
His (H)	Histidine
IB	Inclusion body
IMAC	Immobilized metal-affinity chromatography
IPTG	Isopropyl β-thiogalactopyranoside
ITC	Inverse transition cycling

LB	Luria-Bertani broth
mAb	Monoclonal antibody
MBP	Maltose-binding protein
Met (M)	Methionine
MWCO	Molecular weight cut off
NTA	Nitrilotriacetic acid
ORF	Open reading frame
PBS	Phosphate-buffered saline
PCR	Polymerase chain reaction
Pfu I-ELP	His_6-Pfu I -intein-ELP_{90}
Pro (P)	Proline
RNA	Ribonucleic acid
SDS-PAGE	Sodium dodecyl sulphate-poly-acrylamide gel electrophoresis
T_4-ELP	His_6-T_4-intein-ELP_{90}
Taq-ELP	His_6-Taq-intein-ELP_{90}
TB	Terrific broth
TEV	Tobacco etch virus
Trp (W)	Tryptophan
Trx	Thiordoxin
Trx-pfu I-ELP	Trx-thrombin-pfu I - thrombin-ELP_{90}
Val (V)	Valine

1 Introduction

Proteins vary tremendously in many of their physical and chemical properties (Creighton and Thomas, 1940). In order to perform in vitro application or analysis, one protein must be separated from other cellular components. This process is called protein purification. Purification may be preparative or analytical. Preparative purification aims to produce a relatively large quantity of pure proteins for further utilization. Examples include the preparation of commercial products such as enzymes (e.g. cellulase), nutritional proteins (e.g. casein), and certain pharmaceuticals (e.g. streptokinase). Analytical purification produces a relatively small amount of protein for a variety of research or analytical purposes, including identification, quantification, and studies of the protein's function, structure and interactions. Among the earliest purified proteins were Concanavalin A and urease (Allan et al. 1971; Larson and Kallio, 1954).

The protein purification process commonly starts from cell lysis, in which internal crude lysate of the cell will be released through cell membrane disruption. The resulting sample can undergo preparatory purification using centrifugation, which separates soluble proteins from various insoluble impurities. Precipitation by a number of methods (e.g. salting out) can concentrate the proteins from this soluble lysate. Different types of chromatography methods are then applied to isolate the proteins of interest, based on properties such as molecular weight, net charge and binding affinity. The degree of purification can be detected by gel electrophoresis if

the molecular weight and iso-electric point of the target protein are known, by spectroscopy if the protein has distinguishable spectroscopic properties, or by enzymatic activity assays if the target is an enzyme (Angal S and Harris E.L.V, 1989). Further, proteins can also be isolated according their charge, using the electrofocusing approach. This approach allows proteins to migrate in a gel medium with an established pH gradient; and each protein will stop at the site where the pH is equal to its iso-electric point (O'Farrell, 1985). Today, a large number of techniques have been developed for separation of proteins, based on the knowledge of their molecular properties, structure and functions (see Table 1.1).

Table 1.1 Purification methods that can be used to separate proteins

Purification methods	Basis of Separation
Ultrafiltration	size and shape
Dialysis	size and penetrability
Centrifugation	size, shape and density
Precipitation	
salting out (e.g. $(NH_4)_2SO_4$)	solubility
organic solvents (e.g. acetone)	solubility
isoelectric	solubility and pI
high temperature	solubility
Chromatography	
ion-exchange	charge and charge distribution
hydrophobic interaction	hydrophobicity
gel-filtration	size and shape
affinity	ligand-binding site
immobilized metal affinity	metal binding
immunoaffinity	specific antigenic site
reverse-phase HPLC	hydrophobicity and size
chromatogocusing	charge and pI
Electrophoresis	
gel electrophoresis	charge, size and shape
isoelectric focusing	charge and pI

1.1 Frequently used affinity tags for recombinant protein purification

For native proteins, a series of purification steps may be necessary to obtain sufficient purity for subsequent utilization. To simplify this process, genetic engineering is often applied to add biochemical features to proteins that make them easier to purify without influencing their function or structure. Here, a specific peptide sequence called a "tag" is fused to one terminus of the protein, which gives the protein a character (e.g. binding affinity) it would not otherwise have. Using properties given by the fusion tag, recombinant proteins can be purified in a simple manner. When the tags are not needed any more, they can be removed by a site-specific protease or a self-cleavage linker. To date, many fusion tags have been developed for protein purification. In general, it is hard to decide on the best fusion system for a specific protein of interest. This depends on the target protein itself (e.g. stability, hydrophobicity), the expression host and the application of the purified product. Table 1.2 shows some of the most common tags for fusion protein purification (Stevens, 2000).

Table 1.2 Frequently used fusion tags for recombinant protein purification

Tag	Size	Location	Matrix	Elution conditions
His-tag	6-10aa	N, C, internal	divalent metal	imidazole
c-myc-tag	11aa	N, C	anti-myc antibody	low pH
GST-tag	26kD	N, C	glutathione-Sepharose	reduced glutathione
FLAG-tag	8aa	N, C	anti-Flag mAbs (e.g. M_1)	EDTA
3 × FLAG	22aa	N, C	anti-Flag mAbs	low Ca^{2+}, EDTA
Staphylococcal protein A	14kD	N	IgG-Sepharose	low PH IgG-affinity ligand
Staphylococcal protein G	28kD	N, C	albumin	low PH albumin-affinity ligand
SBP-tag	38aa	C	streptavidin	biotin
Chloramphenicol acetyltransferase	24kD	N	chloramphenicol-Sepharose	chloramphenicol
Strep-tag	9aa	C	streptavidin	biotin
Strep-tag II	8aa	N, C	Strep-Tactin	desthiobiotin
MBP-tag	40kD	N, C	amylose resin	maltose
Calmodulin-binding protein	26aa	N, C	calmodulin	EGTA or EGTA with 1M NaCl
S-tag	15kD	N, C, internal	S-protein	denaturing
Chitin-binding domains	51aa	N, C	chitin	thiol reagents
Cellulose-binding domains	107aa 114aa 156aa	C N N	cellulose	denaturant (e.g. urea) containing buffer
Dihydrofolate reductase	25kD	N	methotrexate agarose	folate buffer
Polyarginine-tag	5-6aa	C	SP-Sephadex	NaCl
Polycysteine-tag	4aa	N	thiopropyl-Sepharose	dithiothreitol
Polyphenylalanine	11aa	N	phenyl-Superose	ethylene glycol
lacZ-tag	116kD	N, C	TPEG-Sepharose	borate

1.1.1 Polyarginine-tag (Arg-tag)

The Arg-tag was firstly introduced as fusion partner for recombinant protein purification in 1984 (Sassenfeld and Brewer, 1984), it usually consists of 5-6 arginine residues. Arginine is the most common alkaline amino acid. Proteins fusing with Arg_5-tag can be purified by cation exchange resin SP-Sephadex, and untagged proteins will not bind. Following combination, a linear sodium chloride gradient from 0 to 400 mM at alkaline pH will be used to elute the fusion protein of interest. Polyarginine sequences have been successfully applied as C-terminal tags in bacteria, resulting in recombinant protein with high yield. Five or six consecutive arginine residues on a C-terminal can be single-step cleaved by carboxypeptidase B treatment. Although there are several successful examples, this enzymatic process has often been limited by low removal yield or by unwanted cleavage at the desired position (Nagai and Thogerson, 1987). Arg-tag, as an uncommonly used ion-binding tag, can be an interesting tool for protein purification if incorporated with a second affinity tag.

1.1.2 Polyhistidine-tag (His-tag)

By far the most commonly employed method utilizes immobilized metal-affinity chromatography (IMAC; introduced by Porath et al., 1975) to purify recombinant proteins with a polyhistidine tail. Purification of His-tagged proteins is based on the interaction between a chelated metal ion (e.g. Ni^{2+}, Zn^{2+}, Co^{2+}, Cu^{2+}) immobilized on a matrix, and short oligomeric histidine residues such as hexahistidine. Histidine is the amino acid that shows the biggest interaction with immobilized metal

ion matrices, because electron donor groups on the imidazole side chain of histidine readily form coordination bonds with transition metals. Proteins incorporated with sequences of consecutive histidines are efficiently withheld on IMAC. Peptides containing His-tag can be easily eluted by adding an imidazole gradient from low concentration up to as high as 250 mM in the succedent washing step. Washing with a low concentration of imidazole eliminates nonspecific binding of host proteins with histidines. Effective His_6-tagged protein elution ranges from 20 to 250 mM imidazole (Hefti et al., 2001; Janknecht et al., 1991).

The use of adjacent histidine residues to purify recombinant protein was first described by Hochuli and his co-workers in 1987 (Hochuli et al., 1987). They used nitrilotriacetic acid (NTA) as a ligand for immobilizing metals of nickel, known today as Ni(II)-NTA. In 1988, His-tagged dihydrofolate reductase was successfully purified within this system (Hochuli et al., 1988). The lengths of the polyhistidine and of the solvent system were both found to affect purification efficiency. Another developed system for his-tagged protein purification is TALON, which consists of a Co^{2+} - carboxylmethylaspartate (Co^{2+}-CMA) as a coupled metal ion matrix. The TALON system allows the elution of tagged protein under mild conditions and displays less non-specific protein binding than Ni(II)-NTA, resulting in higher elution product purity (Chaga et al., 1999). Both these His-tag purification systems are able to withstand multiple regeneration cycles under stringent sanitizing conditions, are relatively inexpensive and exhibit high binding efficiency.

Polyhistidine affinity tags can be placed on either the C- or the N-terminus, or even inside the recombinant proteins. A number of expression systems including bacteria (Rank et al., 2001), yeast (Kaslow and Shiloach, 1994), mammalian cells (Janknecht and Nordheim, 1992) and insect cells (Schmidt et al., 1998) successfully enabled purification using poly-histidine tags. Many his-tagged proteins have been identified by now. Although it seems that a relatively small histidine affinity tag would ensure that protein function is hardly affected, no convincing example has yet been found to prove this point. The His-tag also works well under denaturing conditions, adding another dimension to its versatility. Proteins that have aggregated into insoluble complexes can be dissolved in an appropriate denaturing agent (e.g. urea) and purified by IMAC processing. Refolding of the target protein can then be performed without interference from other proteins.

Although applied universally, purification of protein containing metal ions is not recommended because of absorption by the NTA. Purification under the condition of less air is also not recommended due to the reductive property of Ni(II)-NTA (Terpe K, 2003). Nevertheless, purification of proteins with this affinity tag is the most widely used method.

Lu and co-workers have reported an interesting extension of the IMAC technology in 1996 (Lu et al., 1996). The solubilizing *E. coli* thioredoxin protein was used as a presentation scaffold, inserting histidine residues at different positions that satisfied the geometrical limitation for metal ion binding. Such "His-patch

thioredoxin" has found broad application as an affinity tag since it combines the beneficial features of His-tag and thioredoxin, enhancing the solubility as well as the capacity for binding with immobilized metal ions.

1.1.3 FLAG-tag

An affinity gene fusion system that has attracted much attention in the last twenty years is the so-called FLAG system, which uses a short, hydrophilic eight-amino-acid peptide sequence fused with the target protein (Hopp et al., 1988) for immunoaffinity chromatography on immobilized monoclonal antibodies. The FLAG peptide sequence (Asp-Tyr-Lys-Asp-Asp-Asp-Asp-Lys) contains an enterokinase cleavage recognition sequence (Asp-Asp-Asp-Asp-Lys) on its C-terminal (Maroux et al., 1971). Via immunization of mice with a Flag-tag labeled antigen, monoclonal antibodies (mAbs) to this octo-peptide can be plentifully obtained. M_1 is one of the monoclonal antibodies found to combine specifically with the peptide epitope in a calcium-mediated manner, and so the use of a buffer containing a chelating agent (e.g. EDTA) as the eluant makes this system suitable for the recovery of sensitive target proteins (Einhauer and Jungbauer, 2001). M_2 is an alternative mAb that performs a function similar to M_1; however, M_2 interacts with the FLAG peptide in a non-calcium-dependent manner, thereby either lowering the pH or competing with an excess of synthetical FLAG peptide, instead of with a chelating agent for the elution process (Brizzard et al., 1994).

The FLAG-tag can be located at either the C- or the N-terminus of the protein. Different cell types, including examples from bacterial (Su et al., 1992), yeast (Schuster et al., 2000) and mammalian cells (Zhang et al., 1991) have used this system. Recently, it was suggested that a mini-version of the FLAG peptide, comprising only the first four amino acids (Asp-Tyr-Lys-Asp) could also be recognized with similar affinity by M_1 (Knappik and Plückthun, 1994). Compared with the previously introduced His-tag system, a limitation of the FLAG-tag system is that the mAb matrix is not as stable as Ni^{2+}-NTA.

Sometimes 3×FLAG is used for improve the detection of the FLAG-tag. This triplicate FLAG sequence is also hydrophilic, 22 amino acids long (DYKDHDG DYKDHDIDYKDDDDK) and can detect up to 10 fmol of expressed fusion protein.

1.1.4 Glutathione S-transferase-tag (GST-tag)

The glutathione S-transferases (GSTs) are enzymes that perform the function of transferring sulfur from glutathione to the substances (e.g. halogenated compounds or nitro compounds), resulting in their detoxication (Mehler, 1993). A lot of GSTs can be purified by affinity chromatography on immobilized cofactor glutathione, followed by washing down with reduced glutathione. Based on this principle, a one-step purification system of recombinant proteins with GST-tag was first developed by Smith and co- workers, using GST from *Schistosoma japonicum* (Smith and Johnson, 1988). Their work demonstrated that many GST-tagged eukaryotic proteins could be efficiently isolated from a crude *E. coli* lysate by glutathione-affinity chromatography.

In addition, site-specific protease (like thrombin and blood coagulation factor Xa) is recommended to cleave the N-terminal GST fusion partner. Sometimes GST-tagged proteins are not totally soluble. Insoluble GST recombinant proteins can be renatured from inclusion bodies after solubilization using 6M Gu-HCl and subsequently purified by affinity chromatography.

Recently, a series of vectors were developed allowing either C- or N-terminal fusion of GST to the protein of interest for purification in *E. coli*. Besides, a GST-tag can also be used in yeast (Lu et al., 1997), mammalian cells (Rudert et al., 1996), and baculovirus infected insect cells expression system (Beekman et al., 1994). In a dual affinity fusion approach, GST-tag was fused to the N-terminal of the target protein while a His-tag was fused to the C-terminal; this tripartite fusion protein has excellent binding capacity (Christoph et al., 2003). Although the GST fusion system has been widely used in research, a possible complication associated with this system is the utilization of a reducing agent for elution, which influences proteins containing disulfides (Sassenfeld, 1990). Further, the slow bind kinetics of GST to resin is another drawback, since loading cell extracts to the resin is a time consuming process.

1.1.5 Maltose-Binding Protein-tag (MBP-tag)

Maltose-binding protein (MBP) is a 40 kD expression product encoded by the *malE* gene of *E. coli* K12 (Duplay et al., 1988). Proteins fused with an MBP-tag can be purified in a single step by affinity chromatography on resins containing cross-linked amylase followed by competitive elution with 10 mM maltose in a

physiological buffer. This fact, together with gentle purification conditions, low cost for binding resins and the fact that MBP does not contain any cysteine that can form unwanted disulfide bonds, has led to the fast development of expression systems for MBP-tagged protein in *E. coli* applied to the purification field.

MBP-tag is the only affinity tag discovered so far that can enhance the solubility of recombinant proteins in bacteria, especially eukaryotic proteins (Sachdev and Chirgwin, 1999). Ten asparagine residues between the fusion partner and the target protein make complexation much tighter between the binding component and the amylase resin. The MBP can be fused at either the N- or the C-terminus of the protein; here N-terminal fusion would result in high-level translation. It is necessary to cleave the tag with a site-specific protease. Guan and his co-workers had designed vectors by the insertion of a cleavage site for the factor Xa at the C-terminus of MBPs (Guan et al., 1988). In combination with other small affinity tags, the MBP-tag system has achieved wide application. However, not all MBP-tagged proteins can be purified by amylase-based chromatography; it has been reported that MBP may be a vital factor causing protein degradation (Feher et al., 2004).

1.1.6 Strep-tag

The Strep-tag is an artificial 9-amino acid peptide (AWRHPQFGG) that displays inherent binding affinity towards streptavidin and has been used as an affinity tool for corresponding recombinant proteins (Schmide and Skerra, 1993). Presence of the Strep-tag at the C-terminal of over-expressed proteins enables not

only their convenient detection using a streptavidin-enzyme conjugate, but also their single-step affinity purification on immobilized streptavidin. The bound Strep-tagged protein can competitively be eluted with the biotin or analogs (e.g. iminobiotin) under gentle conditions. Washing with the equilibration buffer leads to matrix regeneration.

Recently, an improved version of the octapeptide, Strep-tag II (WSHPQFEK), was described. Specific mutation of streptavidin at position 44, 45 and 47 have a higher affinity for Strep-tag II than for the native form (Voss and Skerra, 1997). This modified streptavidin is called Strep-Tactin. In contrast to the original peptide, Strep-tag II can be engineered to the positions other than the C-terminal of a protein. Another intrinsic advantage of Strep-tag II is its metal ion independence in the purification process, which allows various metal-containing proteins to be studied. Recombinant Strep-tag-hybrids can be produced in many expression systems such as bacteria (Fontaine et al., 2002), yeast (Murphy and Lagarias, 1997) and plants (Drucker et al., 2002).

1.1.7 Cellulose-binding domain (CBD-tag)

Many cellulases consist of a catalytic domain linked to a non-catalytic cellulose-binding domain (CBD, Gilkes et al. 1991). More than 13 different families of proteins with CBDs have been identified so far. A cellulose-binding domain may exist at any position within the polypeptide (C- or N-terminal, internal) with its molecule weight varying between 4 and 20 kDa. Some CBDs bind to cellulose in an irreversible manner and can be used for active enzyme immobilization in this situation

(Xu et al., 2002); other CBDs bind reversibly to cellulose and so are more useful in the field of purification. When fusing the gene fragment encoding the reversible CBD with genes of passengers, the expressed hybrid proteins can bind tightly to the cellulose matrix on column. The affinity of the CBD-tag is so strong that the combined fusion protein can only be eluted with buffers containing urea or guanidine hydrochloride. The denaturing elution makes refolding of the recombinant protein necessary. One of the main benefits of purification with the CBD-tag is that non-specific affinity can rarely be seen. Fusion CBD-hybrids have successfully been produced in bacteria, yeast and mammalian cells (Tomme et al., 1998).

1.2 Effect of fusion tags on protein expression

Expression and purification are the two most essential factors required for the success of producing proteins, whether for structural studies or biochemical analysis. Although protein purification is no longer considered to be a major limiting step because of the continual discovery of novel affinity tags in the past 10 years, the problem of producing soluble proteins in a high yield has gradually become a major bottleneck in this field. When heterologous proteins are overexpressed in bacterial hosts, translated polypeptide chains trend to aggregate together and accumulate in the form of an insoluble "Inclusion Body" (IB). Some efforts have been directed towards finding appropriate expression conditions to assist in making proteins soluble, such as culturing at decreased temperature (Hammarstrom et al., 2002), changing the *E. coli* expression strain (Miroux B and Walker JE, 1996) and increasing the cell chaperone

capacity (De Marco, 2004). Another approach in promoting proper folding of polypeptides during overexpression in bacteria or other host cells is to incorporate the target protein with a solubilising fusion tag.

Today, solubility-enhancing fusion tags commonly used for this purpose include the thioredoxin (LaVallie et al., 1993), NusA (Davis et al., 1999), and MBP-tag (Guan et al., 1988) from *Escherichia coli*, and the GST-tag (Smith et al. 1988) from *Schistosoma japonicum*. Not only can both MBP and GST perform as chaperone-like folding modulators, but they have also been exploited to devise generic protocols for protein purification. The mechanism by which these carrier proteins promote folding of target polypeptides is not fully understood and may not be universal. For instance, it has been suggested that an MBP-tag using a hydrophobic cleft on its surface reversibly binds to exposed hydrophobic regions of nascent passenger polypeptides, steering the polypeptides toward their native conformation by a chaperone-like mechanism (Kapust and Waugh, 1999). In addition, many of these fusion tags do not function equally well with all partner proteins. Recently, many new solubility tags have been found (e.g. SET-tag, Zhang et al. 2004). We believe that producing soluble proteins with the help of fusion tags in different hosts will not be a hit-or-miss affair in future, with further studying of the mechanism of protein folding.

1.3 Cleavage of the tag

Linking with fusion tags may sometimes affect native characteristics or activities of the protein to be studied. Consequently, it is usually desirable to remove the tag. Site-specific cleavage of the affinity handles from a recombinant protein can be accomplished either by harsh chemical treatment or with an enzymatic strategy, and removal should not influence the structure and function of the target protein. Table 1.3 below lists commonly used chemical and enzymatic agents for site-specific cleavage.

Table 1.3 Common chemical and enzymatic agents for site-specific cleavage

Cleavage agent	Cleavage site	Comments
Chemical agents		
CNBr	—XM↓X—	
Hydroxylamine	—XN↓GX—	
Enzymes		
Enterokinase	—XD_4K↓X—	The site will not cleave if followed by a proline residue
		Secondary cleavage sites at other basic residues
Factor Xa	—XIEGR↓X—	Will not cleave if followed by proline and arginine
		Secondary cleavage sites following Gly-Arg sequences
Thrombin	—XLVPR↓GSX—	Secondary cleavage sites
		Biotin labeled for removal of the protease
TEV protease	—XEQLYFQ↓SX—	His-tag for removal of the protease; highly specificity
		Active over a wide range of temperatures
PreScission	—XLEVLFQ↓GPX—	Enables low-temperature cleavage of fusion proteins
		GST-tag for removal of the protease
3C protease	—XETLFQ↓GPX—	GST-tag for removal of the protease
Sortase A	—XLPET↓GX—	Ca^{2+}-induction of cleavage
		His-tag for on column tag removal
Intein	Self-cleavable	Artificial aa left after cleavage in some cases
		On column cleavage
Carboxypeptidase A	Exopeptidase	Cleaves C-ternimal H
Carboxypeptidase B	Exopeptidase	Cleaves C-ternimal R,K

Chemical-based cleavage has a rather low specificity and may result in protein denaturation and side chain modification of amino acids in the target protein (Nilsson et al., 1997). So, although relatively inexpensive and easy to scale up, such methods are primarily used for releasing small peptides from a larger fusion partner (this is only possible for rather small proteins since the chemical cleavage agent recognizes just one or two amino acid residues, (e.g. Trp, Met or Asn-Gly) rather than the removal of fusion tagged proteins.

Most of the available enzymatic methods for affinity tag removal include enzymatic cleavage of the tag followed by specific removal of the process protease by an affinity chromatographic step to yield the tag-free protein. It is more specific than chemical cleavage, and the entire process is usually under gentle conditions. In order to ensure that no contaminating protease activity is present in the protein preparation, elimination of the processing enzyme itself after cleavage is easier using an affinity-tagged protease or a biotinylated protease. An overview of the process for the purification of affinity-tag proteins and subsequently tag removal is shown in Figure 1.1. Non-protease cleavage of the tag is also possible by introducing a self-cleavage linker intein (Chong et al., 1996). Inteins are members of a growing family of autocatalytic enzyme-like elements that excise themselves post-translationally from their protein hosts and ligate the flanking external segments (exteins) through a self-catalyzed protein splicing reaction. The extein function is typically disrupted by the presence of an intein, but restored after protein splicing. This splicing process is shown schematically by Figure 1.2. The most commonly used proteases for tag

splitting are enterokinase, tobacco etch virus (TEV), thrombin and factor Xa. Recovery of the target protein depends on the efficiency of cleavage.

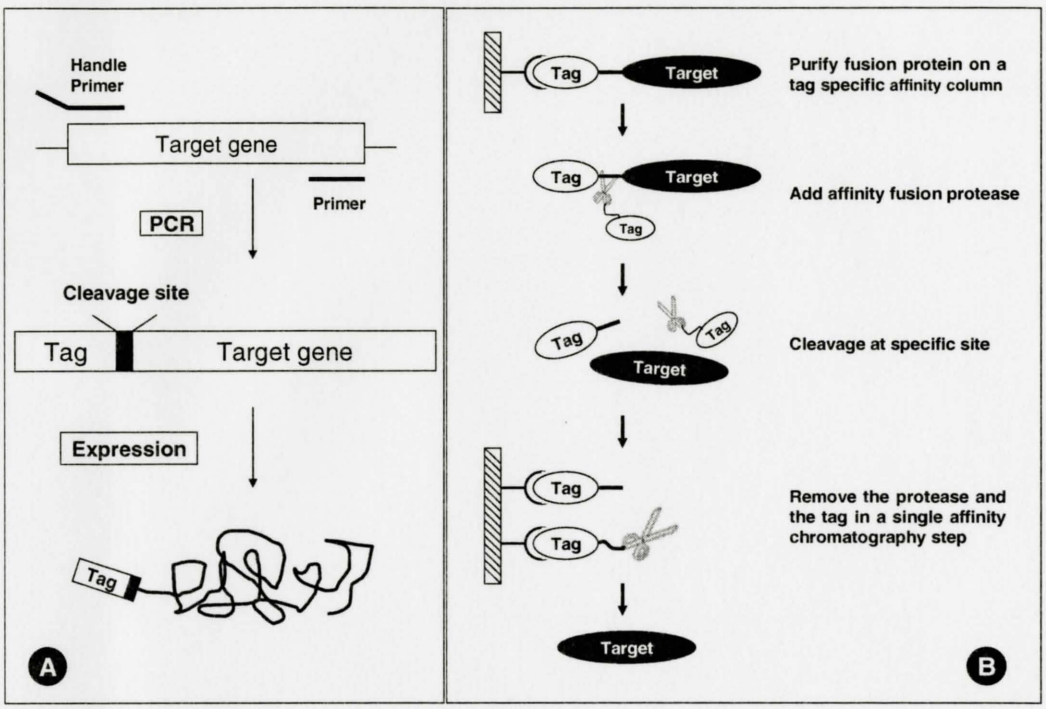

Figure 1.1 (A): Construction of affinity-tagged fusion protein; (B): Purification process of fusion protein using affinity tag and site-specific protease

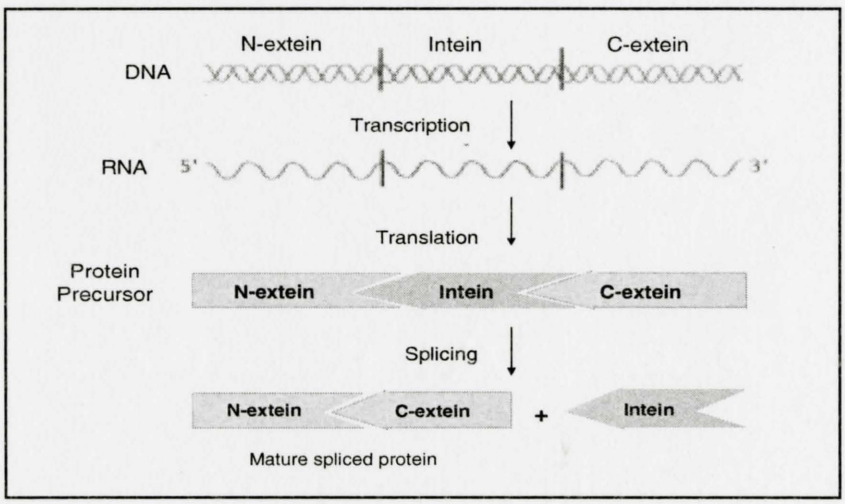

Figure 1.2 The role of protein splicing in gene expression

1.3.1 Enterokinase cleavage

Since enterokinase specifically recognizes a 5-amino-acid sequence (D-D-D-D-K-X_1) and cleaves at the carboxyl site of lysine, it is often the protease of choice for N-terminal fusions. Casual cleavage at other sequences was found to occur at very low levels, depending on the architecture of the protein substrate (Choi et al., 2001). The molecular weight of the light-chain of enterokinase is 26.3 kDa. Biochemical experiments have validated that the amino acid residue X_1 downstream of the D-D-D-D-K sequence affects cleavage efficiency (Hosfield and Lu, 1999). In contrast to other tags, the FLAG-tag (DYKDDDDK) has the enterokinase recognition site inside itself.

1.3.2 TEV protease cleavage

The tobacco etch virus (TEV) protease is a site-specific protease for a 7-amino-acid recognition sequence (E-X-X-Y-X-Q-S). Cleavage occurs between the conserved glutamine and serine (Dougherty et al., 1989). The symbol 'X' in the above sequence can represent various amino acid residues, though not all are suitable. The best cleavage sequence is E-N-L-Y-F-Q-S (Carrington and Dougherty, 1988). When the TEV protease recognition site is placed between two domains, the best cleavage effect will be gained. Efficient cleavage at low temperature, suitable for multiple substrates together with the high specificity, makes TEV protease an ideal tool for removing tags from fusion proteins (Parks et al., 1994). Using TEV protease, one amino acid residue is left at the N-terminal of the processed protein. In this case, further study of the influence of the residual exogenous aa left in the tag-free protein

may be required. The efficiency of cleavage depends on both the tag and the protein fused to the carboxyl terminus of the TEV cleavage site.

1.3.3 Thrombin cleavage

Thrombin is a protease widely employed to remove tags. Cleavage is usually carried out at temperatures between 20°C and 37°C for several hours. As with TEV protease, thrombin cleavage results in the retention of two amino acid residues on the C-terminal side of the cleavage point. The best sequence for α-thrombin has the structure X_1-X_2-P-R (K)-X_3-X_4, where X_1 and X_2 are hydrophobic amino acids and X_3 and X_4 are non-acidic residues (Chang, 1985). L-V-P-R-G-S, L-V-P-R-G-F and M-Y-P-R-G-N are frequently used recognition sites. Cleavage on X_1-X_2-P-R-G-X_4 is more efficient than cleavage on X_1-X_2-P-K-L-X_4. Other short recognition sites are X-R(K)-Y, where X or Y is glycine. Examples are A-R-G and G-K-A, where cleavage occurs after the second residue. Five glycines between the thrombin site and the N-terminal tag improve the cleavage (Guan and Dixon, 1991). Introduction of this "glycine kinker" avoids improper cleavage and saves enzyme necessary to complete digestion. Pure Tris buffer (pH 8) is used during digestion, even though NaCl in the buffer has an inhibitory effect (Haun and Moos, 1992). Thrombin can be liberated from detagged protein by affinity purification on p-amino agarose or gel filtration with a superpose-12 FPLC column (Yu et al. 1995).

1.3.4 Factor Xa cleavage

A factor Xa is another useful tool to completely remove N-terminal affinity tags due to its capacity to recognize a specific sequence between the tag and the target protein. It cleaves at the carboxyl side of the 4-amino-acid peptide I-E(D)-G-R-X, where X can be any amino acid except arginine and proline (Nagai and Thogerson, 1984). Cleavage often occurs in the temperature range 4°C ~ 25°C. Two disulfide-linked chains of 27 kDa and 16 kDa respectively make up the predominant form of 43 kDa factor Xa. Non-specific digestion sometimes occurs with the factor Xa as a site-specific protease (Ko et al., 1993). The reasons may be insolubility of fusion proteins and/or the presence of denaturing conditions. Cleavage can also be enhanced by adding a quinary glycine region (Rodriguez and Carrasco, 1995). Although factor Xa has not been as popular as some other proteases because of its lower efficiency and longer incubation time, there are still several instances of its successful application (Pryor and Leiting, 1997).

In spite of such success, the utilization of site-specific protease has limited the application of many affinity purification systems. First of all, cleavage is not specific all the time and may occur at unwanted sites within the fusion protein. Secondly, the raised temperatures needed for many proteolytic reactions may unfavourably influence properties of the proteins (e.g. stability or activity). Thirdly, because of the inaccessibility of the cleavage site on the fusion protein, cleavage is not always efficient. Lastly, additional chromatographic steps are required to remove the protease

as well as the affinity handle. To eliminate these drawbacks of proteolytic cleavage, a self-splicing protein element, intein, has been developed as a promising alternative for recombinant protein purification (Chong et al., 1997).

1.3.5 Intein self-cleavage

The discovery of a protein splicing element only two decades ago has opened an amazing chapter in the field of biochemistry. In 1990 two research groups independently observed that the open reading frame (ORF) of the gene for *Saccharomyces cerevisiae* vacuolar ATPase (*Sce* VMA) was much larger than the observed protein product. They found that the beginning and end of the encoded protein was very similar to the vacuolar ATPase subunits whose genic sequences had been submitted to the bio-databanks. Nevertheless, 454 aa residues in the central region of the *Sce* VMA bore no resemblance to any known ATPase (Kane et al., 1990, Hirata et al., 1990). It seemed unlikely that this intervening segment was an intron (a non-coding section of precursor mRNA that is spliced out of the RNA before the mature RNA is formed) because it was in-frame with the predicted mature protein - not a necessary condition for RNA splicing. This internal protein element resembling self-splicing intron has been called "intein". Since then, such protein analogs of introns have been repeatedly discovered in all three domains of life (eukaryotes, bacteria, and archaea). More than 200 inteins have been identified to date.

The intein can catalyze its own excision from a precursor protein and concomitantly ligate the flanking regions, the exteins, to form a mature protein. This

internal protein segment auto-processing is a post-translational processing event (Perler et al., 1994; Hirata et al., 1990; Kane et al., 1990). Protein splicing is catalyzed entirely by amino acid residues contained in the intein (no coenzymes and metabolic energy are required) and it is an intramolecular process. The currently accepted mechanism for protein splicing of most inteins is shown in Figure 1.3. It consists of four distinct nucleophilic displacements: (I) formation of a linear (thio)ester intermediate at the N-terminal splice site; (II) formation of a branched intermediate through trans (thio)esterification by the -OH or -SH side chain at the N-terminus of the C-extein; (III) cleavage of amide linkage at the intein C terminus by Asn or Gln cyclization to release the free intein; (IV) formation of a normal peptide bond between the exteins through a spontaneous acyl shift.

Most reported inteins have two domains: a homing endonuclease gene (HEG) domain and a splicing domain. These are known as full inteins. The HEG domain is responsible for DNA double-stranded break repair, which is not necessary for intein splicing, and so it can be removed, forming a mini-intein.

Due to their specific protein splicing feature, inteins have found an important role in some key areas of protein engineering. They have been modified for particular applications such as protein synthesis (Schwarzer and Cole, 2005) as well as for selective labeling of protein segments, which is useful for NMR studies of large proteins (Muralidharan and Muir, 2006). However, the most exciting application is their capacity as self-cleavable tags to facilitate protein purification.

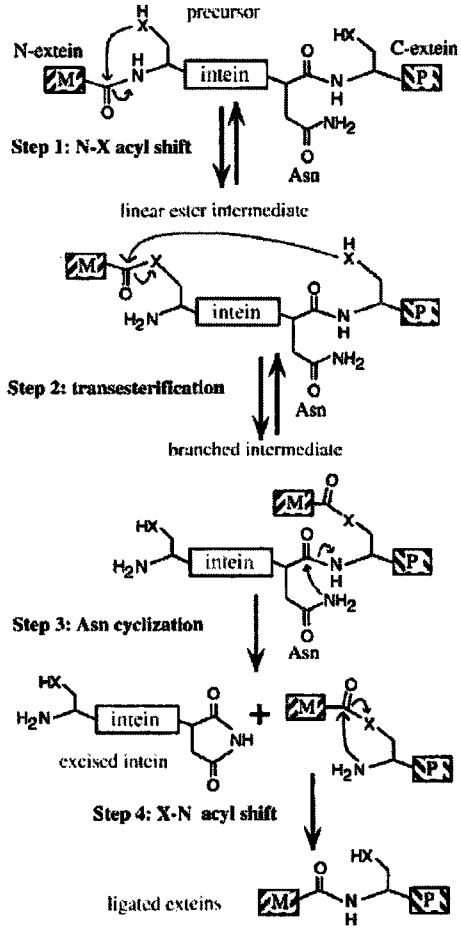

Figure 1.3 The currently accepted mechanism for intein splicing (Christopher et al., 2000)

Using in vitro protein splicing systems, Chong and his co-workers have validated the protein splicing pathway of inteins from the 69 kDa vacuolar ATPase subunit of *Saccharomyces cerevisiae* (Chong et al., 1996). They found an interesting result: that the single aa substitution from Asn^{454} to Ala^{454}, at the C-terminus splicing junction of the intein, blocks splicing and C-terminal cleavage, but has no influence on the N-terminus by the N-S acyl shift at Cysl. Nucleophiles that react with thioesters, such as thiols (e.g. DTT) can effectively shift the N-S equilibrium by attacking the thioester, thereby producing isolated N-terminal cleavage of the intein.

After cleavage, inducers such as DTT can conveniently be removed from the system by an additional desalination step. Figure 1.4 shows intein cleavage for protein purification induced by a reducing agent.

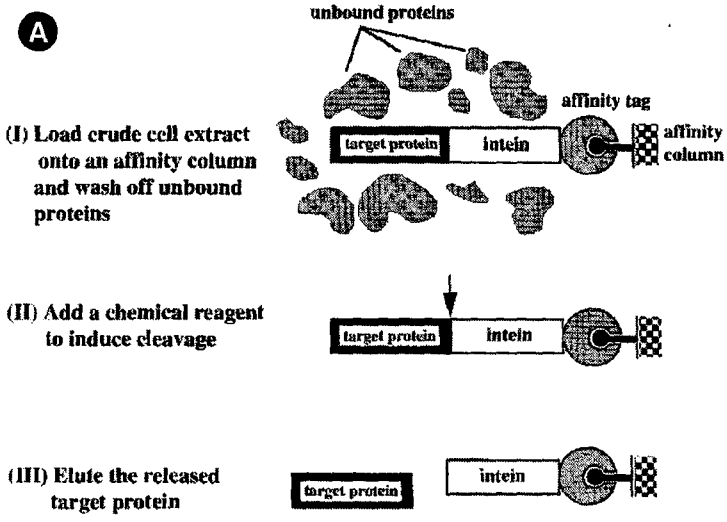

Figure 1.4-A On-column purification of fusion by intein self-cleavage on its N-terminal

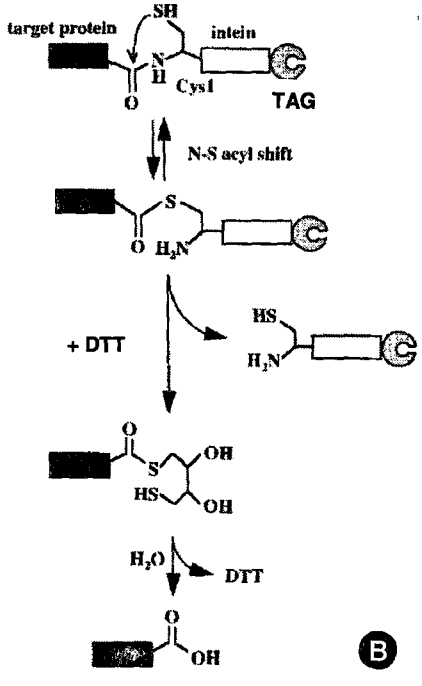

Figure 1.4-B Chemical mechanism for intein N-terminal cleavage (Chong et al. 1997)

In the configuration where the target protein is released by N-terminal cleavage as described above, the cleavage reaction requires the addition of thiol-containing compounds (e.g. DTT) which modify the C-terminus of the product protein. Native target protein is recovered only after subsequent hydrolysis of the cleavage-inducing reagent (Chong et al., 1997). Furthermore, as an unusually strong reducing agent, DTT has a high conformational propensity to open the disulfide bridge between amino acid residues (e.g. cysteine) inside proteins of interest to form a 6-membered ring with an internal disulfide bond (Cleland W, 1964). These factors cause thiol-like agents not desirable for intein cleavage inducement.

Recently, two research groups used rational protein engineering combined with genetic selection to generate novel chimeric inteins whose cleavage activity is regulated by the presence of small molecules (e.g. the human thyroid hormone) in a dose-dependent manner (Skretas and Wood, 2004; Liu et al., 2004). This artificial intein chimera was designed by fusing a ligand-binding domain (LBD) within a previous engineered mini-intein. The insertion of the binding domain blocked the self-cleavage activity of the intein, but allowed its later restoration by combining with an appropriate ligand (conformational changes caused by the ligand binding initiate intein self-cleavage). This small molecule activated process is illustrated by Figure 1.5. Using small molecules instead of traditional stimuli (e.g. DTT) to switch intein self-cleavage "on and off" not only eliminates change of target protein structure but also has potential in many biotechnology applications, since ligand-activated intein splicing out of a strong reporter protein could be used as a highly specific biosensor.

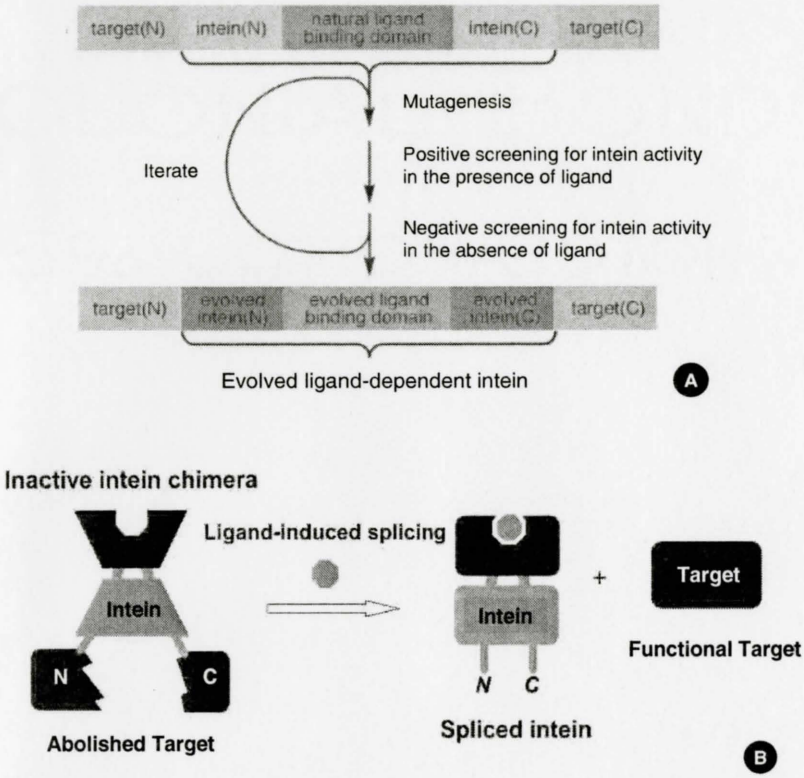

Figure 1.5 (A): Strategy for the directed evolution of a ligand-dependent intein (Liu et al. 2004); (B): Intein splicing tunable behavior with the stimulation of small molecule (Wood et al. 2004)

However, the intein self-cleavage system does have its shortcomings, for example: (I) purification must be done in the absence of reducing agents such as DTT, (II) the slow rate of autosplitting, and (III) the fact that inteins neither increase the solubility nor enhance the stability of the fusion proteins.

1.4 ELP-intein protein purification system

A great number of affinity tag based systems have been developed to facilitate protein purification through the expression of recombinant proteins as fusions to affinity tags. These systems typically allow single-step purification of a fusion protein

from cell lysate by affinity chromatography, using an immobilized moderate-affinity ligand specific to the carrier protein. Though it is useful for lab-scale purification, affinity chromatography on scale-up can represent a major cost of the final protein product at the preparatory level. More technically simple and economical approaches for fusion protein purification are therefore desirable. In contrast with traditional recombinant protein column operation with affinity tags, the ELP-intein system was developed with many attractive features (Ge et al. 2005; Wood et al. 2005).

1.4.1 Thermally-responsive Elastin-like polypeptides (ELPs)

Elastin-like polypeptides (ELPs) are artificial biopolymers consisting of tandem repetitians of the pentapeptide segment Val-Pro-Gly-Xaa-Gly (VPGXG), where the guest residue Xaa can be any amino acid other than Proline (Urry, 1992). This motif is derived from a natural mammalian tropoelastin sequence. ELPs are able to undergo an inverse and switchable hydrophilic-hydrophobic phase transition, within a very narrow temperature range (~2°C). At temperatures below the inverse phase transition temperature (T_t), ELPs are structurally disordered, highly solvated and therefore soluble in aqueous solutions, but when the temperature exceeds T_t, they collapse and aggregate to form coacervate droplets (Urry, 1988; Meyer and Chilkoti, 1999). The aqueous two-phase system (ATPs) formation process for ELPs is shown in Figure 1.6. It has been shown that the inverse phase transition behavior of ELPs holds when they are fused to a variety of proteins (Meyer and Chilkoti, 2003). The commonly accepted explanation of the thermally driven phase transition behavior of

ELPs is that conformational changes in the polypeptide, including loss of waters of hydrophobic hydration, result in the formation of intramolecular contacts between non-polar regions of the ELP.

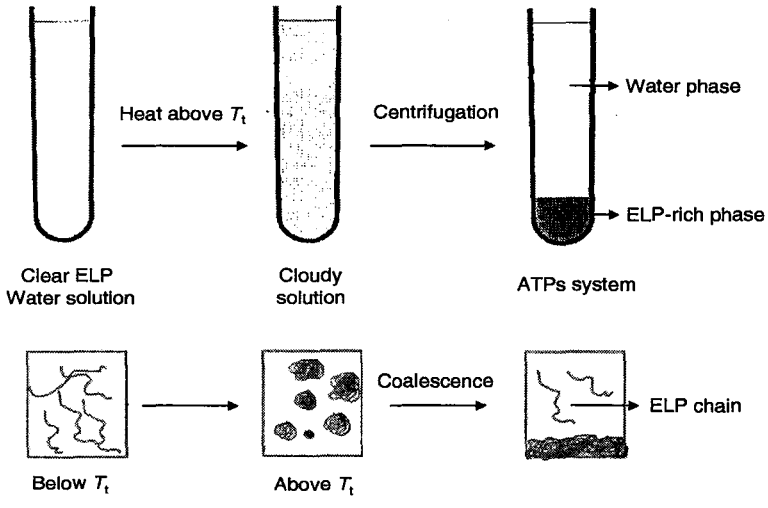

Figure 1.6 Schematic diagram of the ATPs formation process for ELPs

1.4.2 Factors that affect thermal behavior (T_t) of ELP

The phase transition temperature is the most important parameter for evaluating thermal behavior of ELPs. It is very sensitive to the concentration, composition and molecular weight of the ELP, but can also be affected by salt concentration and by the properties of accompanying fusion proteins (Urry et al., 1985; Chilkoti et al. 1999, 2003, 2004). This feature of ELPs is useful because it allows us to design an ELP sequence that will exhibit a desired T_t.

Urry and his colleagues have shown that the T_t values of ELPs are reduced by the incorporation of greater fractions of non-polar residues at the fourth Xaa residue of the repeating VPGXG sequence, and that the extent of this effect is not solely

dependent on the mole fraction but also on the hydrophobicity of the guest residue. This observed correlation between the T_t and guest residue composition within the ELP chain is called the "$\triangle T_t$ effect" (Urry, 1992). Further, exposed hydrophobic patches on the surface of proteins fused to ELPs can also alter the T_t of the ELP in a manner analogous to the $\triangle T_t$ effect (called the "fusion $\triangle T_t$ effect") whereby the T_t is depressed in proportion to the surface hydrophobicity of a fused protein. It is well known that with increasing temperature, the water molecules of hydrophobic hydration solvating the ELP chain, as well as any water molecules closely associated with a hydrophobic surface, are released to bulk solution. The lower T_t of the more hydrophobic ELP (or the ELP fused with a more hydrophobic partner) is simply a result of the release of the additional water of hydrophobic hydration contributed by the hydrophobic amino acid (or hydrophobic region within proximity of the ELP), which offers an excess gain in entropy per ELP chain on its release to bulk (Li et al., 2001).

The effect of concentration and MW (or chain length) on the T_t has been previously reported (Urry et al., 1985). In the two-phase coacervate region, there is a "MW & concentration effect"; that is, although the T_t is relatively invariant at high MW and high concentration, it elevates dramatically at low MW and low concentration. Figure 1.7 shows this trend very clearly. Recently, Meyer and Chilkoti have presented a single equation that quantitatively models the transition temperature as a function of ELP length and concentration (Meyer and Chilkoti, 2003).

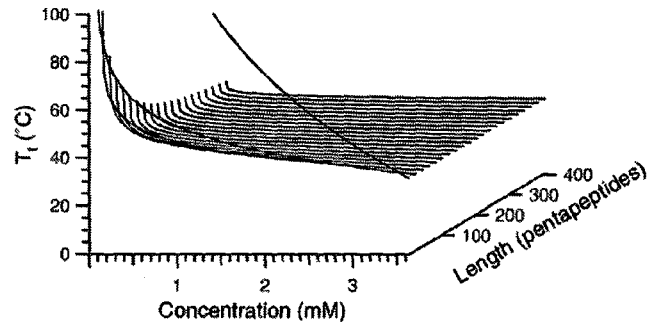

Figure 1.7 Profile of T_t versus concentration for different ELP chain lengths for the ELP [$V_5A_2G_3$] library (Meyer and Chilkoti, 2003)

In addition, regulating the salt concentration of the ELP solution allows the transition temperature to be tuned over a large range (50°C), and thereby offers a means for isothermal triggering of the phase transition (Chilkoti et al. 1999). The T_t of ELP and ELP-fusion is significantly depressed by adding salt into the protein solution. The likely mechanism for this effect of salt is believed to arise due to hydration. A concept of apolar-polar repulsion of hydration free energy, with the emphasis on structural dependence, has been proposed (Urry et al. 1991). The essence of this free energy is that hydrophobic amino acids of ELP and the ions from the salt each have their own hydration structures, and when sufficiently close together, the structures of these two kinds of hydration water are commonly incompatible with each other. The overlap of these two distinct hydration structures increases the free energy of the system. Thereby, the release of these overlaps will be energetically factored. It was found, moreover, that the sensitivity of ELP to salt stimuli could be observably improved by incorporating more ionizable guest residues instead of aliphatic amino acids inside the peptide sequence. The amount of salt needed to initiate the phase

transition was therefore be reduced (Lim et al., 2007)

1.4.3 Effect of ELP tag on the expression of fusion

Unlike its predominant ability in the field of protein purification, the ELP tag is unhelpful for the expression of fusion protein, sometimes even disadvantageous. The influence of ELP on the production of fusion protein depends on the length of the ELP chain. The longer the ELP chain, the lower the fusion protein gain. In order to promote Thioredoxin-ELP expression, Meyer and his colleagues reduced the size of the ELP tag from 36 to 9 kDa and found that the yield of thioredoxin improved 4-fold, to a level comparable to that of expression without an ELP tag (Meyer et al., 2001). Nevertheless, truncation of the ELP also causes a more complicated phase transition behavior. It was shown by dynamic light scattering (DLS) that large droplets with hydrodynamic radii of 2 μm form as the temperature is raised above T_t (for both cases). And they persist at all temperatures above T_t for the large ELP tag fusion. With the short tag, small droplets with hydrodynamic radii of 12 nm begin to form at the expense of the micron-size droplets as the temperature is further raised. For the aforementioned reason, an appropriate selection of salt concentration and solution temperature is required when short chain ELP is fused with target.

1.4.4 Introduction of ELP-intein protein purification system

Due to the specific salt and temperature-sensitive solubility of ELP, ELP-tagged proteins can easily be purified by salt addition or temperature shift. The other

notable recent advance in the field of recombinant protein separation is the development of self-cleavage intein that can be used for the removal of affinity tags under modest conditions. Upon combining an ELP tag with the linker of engineered intein, highly purified native protein can be produced from a total cell lysate by simple mechanical means, without chromatography.

Ge and his colleagues firstly created a tripartite fusion protein to demonstrate proof-of-principle of this methodology (Ge et al., 2005). As Figure 1.8-A shows, thiordoxin (Trx), as a model target protein, was cloned upstream of a mini-intein from the *Mycobacterium xenopi* GyrA gene (Mxe) and an ELP gene, forming an N-terminal fusion to the intein-ELP tag (Trx-Intein-ELP). The intein they chosen has an $Asn^{315} \rightarrow Ala^{315}$ mutation at its C-terminal which allows cleavage only takes place specifically at its N-terminal (since modified Mxe inactives C-terminal cleavage). Three amino acids MRM were added before the mini-intein gene to control the cleavage reaction (Evans et al., 1999). The ELP used in the fusion consisted of 90 repeats of the pentapeptide [VPGXG]-90 (where X is V, A or G, in the ratio of 5:2:3).

Figure 1.8-B shows the non-chromatography and non-protease protein purification process based on the ELP-intein fusion tag. The tagged protein is liberated from insoluble cell debris by centrifugation at a low temperature, at which the ELP fusion is soluble. Then NaCl is added to clarified lysate at room temperature to induce phase transition, which results in the collapse and coalescence of the ELP fusion protein chain to form micrometer-sized droplets. After that, the mixture is gone

through a 0.2 μm membrane by microfiltration, causing the retention of the ELP fusion coacervate and washout of non-tagged and non-aggregated proteins present in the lysate (or using centrifugation and keeping the ELP fusion coacervate on the bottom of the container). Next, the insoluble ELP fusion precursors that have accumulated on the microfiltration membrane surface are redissolved and recovered by eluting with water or low salt concentration buffer (e.g. phosphate-buffered saline, PBS buffer). Then, 20 mM dithiothreitol (that triggers cleavage reaction of N-terminal intein) is added to the eluate, releasing thiordoxin from the Mxe-ELP tag. Finally, a second round of aggregation-filtration-elution is performed to isolate thioredoxin from the Mxe-ELP tag. After ultrafiltration with a 3 kDa MWCO centrifugation filter, the purified thiordoxin can finally be desalted and concentrated.

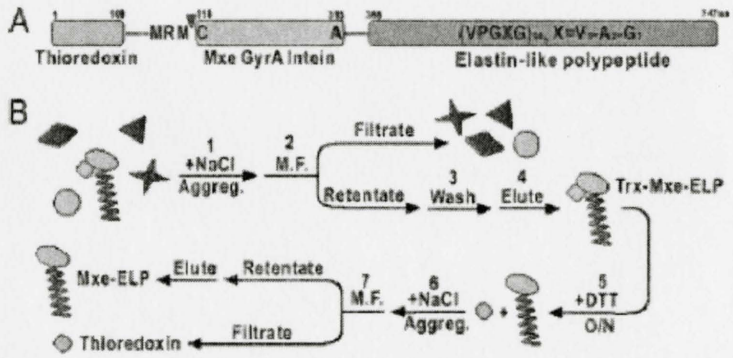

Figure 1.8 Scheme of tripartite self-cleaving environmentally responsive purification tag and purification scheme (Ge et al. 2005)

David Wood and his colleagues at Princeton University almost simultaneously published a similar chromatography-free strategy for fusion protein purification (Wood et al., 2005). In this case, firstly, they constructed a small library of ELPs of

varying lengths and inserted them into the expression vector pET-21(+) under the control of a T_7 promoter. Then they used the dynamic light scattering (DLS) method to evaluate the ability of each ELP to aggregate efficiently in fusion to the highly soluble NusA protein, and determined the approximate T_t in each case. They finally chose [VPGXG] -110 as experimental material due to its moderate T_t (below 30 °C) in high salt buffers. They inserted ten diverse target proteins immediately after the C-terminal dipeptide (His-Asn) of the pH and temperature-sensitive mini-intein (△I-CM, derived from *Mtu* RecA intein) to construct the tripartite fusion protein analog described by Ge. All these proteins can be successfully purified by the aforementioned process.

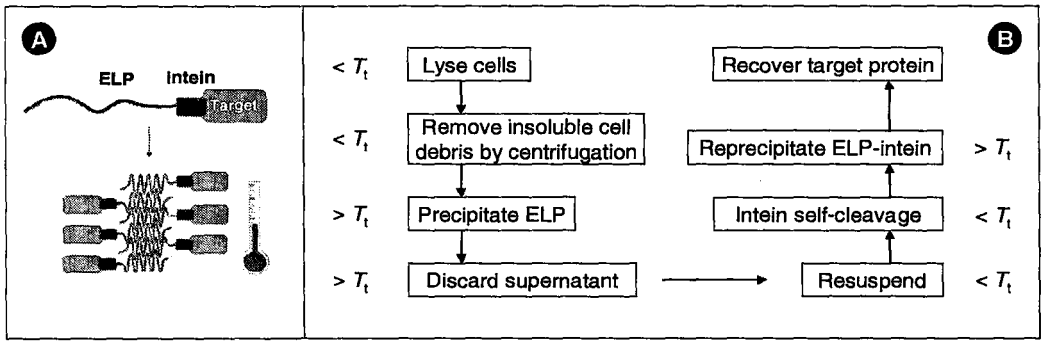

Figure 1.9 (A): The ELP-intein tag is designed to self-associate into an insoluble core upon heating to a temperature above T_t; (B): Scheme of the target protein purification (Wood, 2006)

1.5 Problems with the ELP-intein protein purification system

The simplicity and self-contained nature of the ELP-intein protein purification system promises a breakthrough in the production of purified recombinant proteins. Although possessing many merits over conventional purification approaches, this

system has some defects. Some problems (e.g. buildup of the inclusion body) occur often enough to limit its implementation. These unsolved issues have motivated us to investigate this freshly invented system further.

When recombinant proteins are expressed in bacteria, formation of non-native cytoplasmic aggregates (inclusion bodies) is a common problem which affects yield of the target protein. Inclusion bodies (IBs) are generally thought to contain misfolded or partially folded protein through the exposure of hydrophobic patches and the consequent intermolecular interactions. The mechanism for the formation of IBs is still a controversial issue. Table 1.4 lists several plausible reasons to explain why IBs form. One explanation is that high production rate at high expression temperature may allows insufficient time for the nascent chain to fold properly, causing rapid and incorrect folded proteins to aggregate as an inclusion body (Mitraki and King, 1989). In addition, the absence of cofactors like molecular chaperones (which help properly folding) in the cytoplasm environment is considered to be a possible IB formation pathway (Catherine, 1989). Figure 1.10 displays a hypothetical in vivo maturation process for a dimeric protein. Despite lost part of total expression yield and occupy more time, culturing at lower temperatures was reported to promote the expression of genes into soluble proteins and alleviate IB formation (Hanning and Makrides, 1998). Like most proteins, some ELP-intein fusion proteins suffer from the inclusion body problem at normal expression temperature. Directly motivated by these findings, we have applied a low-temperature expression strategy to solve the problem in this research. As expected, with culturing at lower temperatures, most of the T_4-ELP

inclusion bodies were transformed into soluble expression.

Table 1.4 Suggested explanations for the formation of inclusion bodies in *Escherichia coli*

Less Plausible
(1) Foreign proteins are recognized as such in *E. coli*
(2) High production rate allows insufficient time for the nascent chain to fold properly
More Plausible
(3) High local concentration in the cytoplasm leads to non-specific precipitation
(4) Lack of cellular compartmentation causes the protein to be produced in a reducing environment, preventing formation of S-S bonds necessary for proper folding
(5) Lack of mammalian post-translational modifying enzymes
(6) Lack of proper foldases during production
(7) Aggregation is a function of the limited flexibility of the polypeptide chain and the instability of the native tertiary structure of most proteins

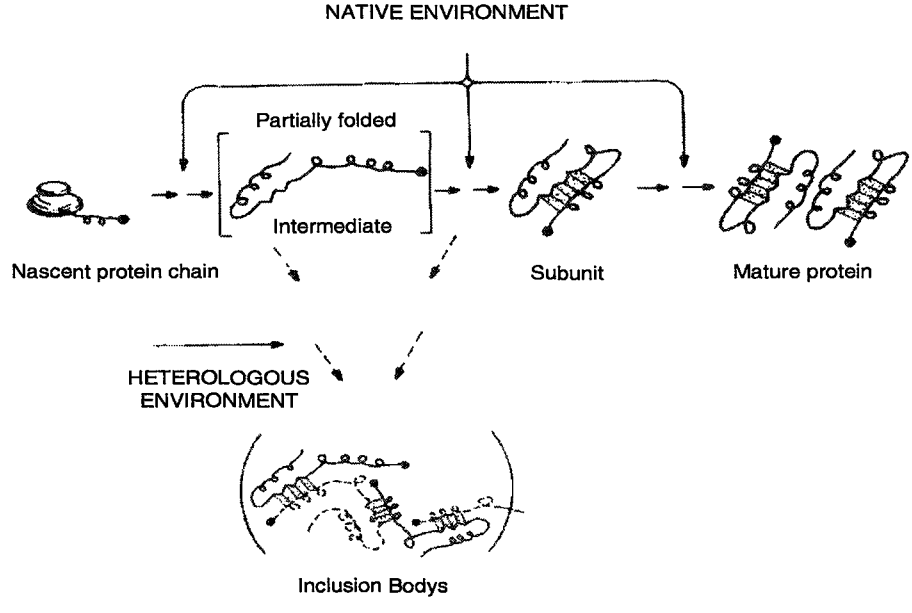

Figure 1.10 Hypothetical folding and maturation pathway for a dimeric protein

Inverse transition cycling (ITC), as a new and simple purification method for ELP-intein tagged proteins (already noted), has been demonstrated in the past with proteins that were expressed to high levels, but poor expression levels were expected for toxic proteins, complex multidomain proteins, or heterelogous protein expression (Chilkoti et al. 1999, 2004; Wood et al., 2005). In this thesis, we also describe how for very low concentrations of ELP tagged proteins, ITC in its original format does not allow recovery of the proteins. We hypothesize that if an ELP tagged molecule is present in a solution at a very low concentration, adding an excess amount of free ELP to the sample and inducing reversible inverse phase transition would form hybrid aggregates via the interaction of ELP moieties of the two molecules. This mechanism is based on the coaggregation of free ELP and ELP tagged proteins that have been reported for the purification of a fusion of a single-chain Fv of an antiatrazine antibody to ELP (Kim et al. 2003, 2005). Later the efficiency of this method was quantitatively shown by Ge and Filipe (2006). We have used this efficient and reversible capture system for low yield ELP-intein tagged protein purification in this study, with very good result.

2 Experimental Procedures

2.1 Preparation of culture medium

2.1.1 Luria-Bertani broth (LB)

An LB broth was prepared in the lab by dissolving 10 g Bacto™ Trypton, 5 g Yeast extract (EMD Chemical Inc.), and 5 g NaCl per liter of distilled water (Milli-Q water system, USA), and mixing it with a magnetic stirrer in a beaker. The medium was poured in 250 mL flasks, 50 ml in each flask. The flasks were closed and sealed with cotton and foil, and then sterilized in an autoclave (HV-50 Hirayama Sterilizer, Japan) for 20 minutes at 121°C and stored for future use.

15 g of BBL™ agar powder was suspended in 1 liter of LB liquid medium. The powder was completely dissolved by boiling for 1 minute with frequent agitation (VWR 375 Hotplate / Stirrer, USA), and then autoclaved at 121°C for 20 minutes. The solution was then chilled to 50°C and aseptically poured to Petri dishes to solidify, using a biosafety cabinet (Napflow Napco, class II , Type A/B3).

2.1.2 Terrific broth (TB)

TB is a highly enriched medium developed by Tartoff and Hobbs to improve yield in plasmid-bearing *E. coli* (Tartoff and Hobbs 1987). Recombinant strains have an extended growth phase in the medium. The addition of extra tryptone and yeast extract in the medium allows higher plasmid yield per volume. The TB medium

contain, per liter of distilled water: 12 g Bacto^{TM} Trypton, 24 g Yeast extract (EMD Chemical Inc.), 17 mM KH_2PO_4, 72 mM K_2HPO_4, and 4 ml glycerol. The medium was autoclaved for 20 minutes at 121°C and stored for future use.

2.2 Preparation of competent cells

There are two major methods for preparation of competent bacterial cells for transformation: the calcium chloride and the electroporation method. In this experiment, we chose the $CaCl_2$ method. The commercial *E. coli* BLR $(DE)_3$ strains from Novagen Company were stocked in an ultra-low-temperature refrigerator (ULT2586-5-A36 Legaci refrigeration system, USA). The experiment proceeded as follows. Thaw -70°C strains in the icebox for 20 min. Pick 200 μL and inoculate it into 20 ml of LB medium. Incubate at 37°C overnight with shaking at 180 rpm (Shel lab incubator, 1575R, Sheldon manufacturing Inc., USA). Transfer 500 μL of the saturated overnight culture to a sterile 500-mL flask containing 100 mL of LB medium. Incubate the cells at 37°C, shaking at 180 rpm, until OD_{600} reach 0.6; this usually takes about 3.0 hours. Check the OD frequently when it gets beyond 0.2 to avoid overgrowth. When the culture reaches an OD_{600} of 0.6, chill the flask on the ice for 20 min, transfer 50 ml of the cells to a sterile polypropylene centrifuge tube, and collect the cells by centrifugation at 5000 rpm for 5 minutes at 4°C in a bench-top centrifuge (Beckman Coulter Allegra 25R centrifuge, Germany). After centrifugation, decant the supernatant and resuspend the cell pellet in 10 mL of cold, sterile 100 mM calcium chloride, incubate in an ice-water bath for 20 minutes, and centrifuge as

before. Decant the supernatant again and gently resuspend in 2 ml of cold, sterile 100 mM calcium chloride to yield the final competent cells suspension. Aliquot each 200 μL of competent cells to a 1.5 mL tube. If they are not immediately used, cells can be stored at 4°C for maximum of 6 hour without significant loss of competence. The same competent cells can also be stored at -70°C for long-term storage.

2.3 Plasmid transformation

Add plasmid ($\leqslant$ 20 μL) to ice cold 200 μL competent cells and incubate on ice for 30 minutes. Heat shock in a 42°C water bath (VWR 1160S, USA) for 90 seconds. After heat shock, immediately put back on ice for 2 minutes. Add 0.8 mL LB medium into the tube. Shake and incubate at 37°C for 45 min. Plate (no more than 200 μL) on the appropriate agar plates which contain antibiotic. Incubate at 37°C, overnight.

2.4 Construction of plasmids

The construction of plasmid that encodes for fusion of thioredoxin and ELP, pTrx-ELP_{90}, is described elsewhere (Ge et al., 2006). It is based on the pET-32b expression vector (Novagen, Madison, WI). The ELP tag in this plasmid is denoted as ELP [$V_5A_2G_3$-90], which comprises 90 repeats of the pentapeptide Val-Pro-Gly-Xaa-Gly, where Xaa is Val, Ala and Gly in the ration of 5:2:3. To obtain a plasmid that encodes for the ELP tag only, the plasmid that encodes the fusion protein was digested with *Nde* I (Fermentas), and the large fragment was separated by agarose gel electrophoresis and purified with DNA extract kits (Qiagen). The purified DNA

fragment was then self-circulated under the catalysis of T_4 DNA ligase (Fermentas) to generate the plasmid that encodes for the expression of the ELP without thioredoxin. This plasmid is called $pELP_{90}$. The other plasmids that encode for the recombinant protein His_6-T_4-intein-ELP_{90}, His_6-pfu I -intein-ELP_{90}, His_6-Taq-intein-ELP_{90}, and Trx-thrombin-pfu I -intein-thrombin-ELP_{90} were a gift from Microstar Biotech Inc.

2.5 Inverse transition cycling (ITC) method

Using environmental stimuli, either increasing the temperature above the critical phase transition temperature (T_t) or adding NaCl into the cleared lysate (or combination of both), ELP-tagged recombinant proteins can be selectively separated from other contaminating *Escherichia coli* biomacromolecules to a high purity. Below the critical phase transition temperature (or present in low-salt buffers), fusion proteins are soluble, whereas above this temperature (or after salt addition), they undergo a reversible inverse phase transition (RIPT) and become insoluble. The aggregated ELP fusion can be isolated from other proteins present in the mixture, either by centrifugation or by microfiltration (Ge et al., 2006). These ELP fusion proteins can be resolubilized by decreasing the temperature lower than T_t or reducing the ionic concentration of the buffer in which the protein is present. This approach has been demonstrated with multifarious target proteins (Chilkoti et al. 1999, 2004). In all cases, the target fusion proteins were obtained with purity levels higher than 95%. The target fusion proteins can achieve even higher purity after several repeat cycles of ITC. This approach has also been shown to yield purified proteins in quantities that are

similar to those obtained using affinity chromatography.

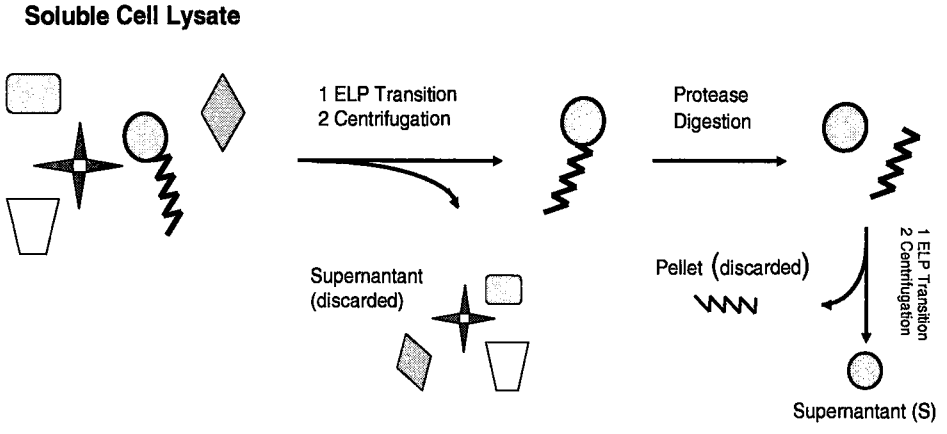

Figure 2.1 Inverse transition cycling (ITC) purification scheme. The ELP fusion protein is isolated from other biomacromolecules in the cell lysate by inducing the inverse temperature phase transition of the ELP. The inverse phase transition of the ELP fusion is switched by an increase in temperature and/or ionic strength, and aggregated protein is separated by centrifugation or microfiltration, leaving the contaminating biomolecules largely in the soluble fraction. The ELP fusion protein is then resolubilized by cooling to below its T_t or in fresh buffer. If desired, the target protein can be removed from the fused ELP tag by self-cleavage of intein tag between the ELP tag and the target protein. The cleaved intein-ELP can be removed by another round of ITC. After centrifugation, the purified target protein is obtained in the supernatant, while the aggregated intein-ELP is discarded in the pellet.

2.6 Protein expression and purification

There are four recombinant ELP fusions that we needed to express and purify for this study. They are Trx-thrombin-pfu I -intein-thrombin-ELP (Trx-Pfu I -ELP), His_6-T_4-intein-ELP (T_4-ELP), His_6-Taq-intein-ELP (Taq-ELP) and His_6-Pfu I -intein-ELP

(Pfu Ⅰ-ELP). T_4 as DNA ligase can link together DNA strands that have double-strand breaks. Pfu Ⅰ and Taq are frequently used DNA polymerase for polymerase chain reaction (PCR). All of them are indispensable tools in modern molecular biology research.

The transformed cells were cultivated in Luria-Bertani (or TB) media, supplemented with corresponding 100 μg/mL antibiotics. The transformed cells of free ELP and Trx-Pfu Ⅰ-ELP used ampicillin as their antibiotics, other fusions chose kanamicine. Shaker flasks were inoculated from a single colony and cultured overnight at 37°C (for Trx-Pfu Ⅰ-ELP and free ELP) or cultured three days at 20°C (for T_4-ELP, Taq-ELP and Pfu Ⅰ-ELP) at a speed of 180 rpm. It has been shown that expression of pure ELP and ELP-tagged proteins is greatly enhanced when the cells are not induced using isopropyl β-thiogalactopyranoside (IPTG), but rather when the cultivation period is extended (Guda et al., 1995); and this approach was used in this study. An additional culture experiment was done with non-transformed *E. coli* BLR (DE_3) strain, to provide cell lysate devoid of ELP or ELP tagged proteins. The cells were harvested by centrifugation at 5,000×g, 4°C, for 10 minutes. The pellet was resuspended in 1×PBS buffer (137 mM NaCl, 2.7 mM KCl, 4.2 mM Na_2HPO_4, and 1.4 mM KH_2PO_4, pH 7.4) in a volume ratio of 1:25 to the initial volume of the culture medium. After washing twice with PBS buffer, the resuspended cells were disrupted by introducing discontinuous ultrasonic pulses (VirTis Company, USA) in an ice-water bath for 10 minutes, at the maximum power at which no substantial amount of foam was generated. The lysate was then centrifuged at 16,000×g, 4°C, for 10

minutes to separate the soluble part away from insoluble particles.

From these samples, free ELP, Trx-Pfu I -ELP and T_4-ELP were purified using three rounds of inverse transition cycling (ITC). Taq-ELP and Pfu I -ELP, the soluble yield of which are usually very low, even when expressed at low temperatures, can be purified by the method of coaggregation with free ELP in PBS (to be described in the later part of this thesis). Except for the free ELP in this experiment, all the recombinant proteins are fused with the intein tag at the C-terminus of the target proteins. In accordance with the specific feature of the intein tag, a reducing agent such as dithiothreitol (DTT) can stimuli its cleavage and release the native target from the purified precursor. The fusing partners at the C-terminus side of intein as well as intein tag itself can be effectively removed by overnight incubation of the samples with 20 mM dithiothreitol (Sigma), followed by one more round of ITC so as to obtain proteins of interest. The purified proteins were then desalted and concentrated by using an ultrafiltration centrifugation filter with a molecular weight cutoff of 3 kDa (Pall Corp, USA). The purity of the proteins was characterized using 10% SDS-PAGE, followed by staining with Coomassie Brilliant Blue. The concentrations of purified ELP and ELP tagged proteins were determined by measuring the UV absorbance at 280 nm (Beckman DU530 UV / Vis spectrophotometer, Germany) and by using the molar extinction coefficients at 280 nm (e.g. 5690 $M^{-1}cm^{-1}$ for ELP) calculated from their amino acid composition with the software program Protean (DNA Star).

2.7 Capture of low concentration ELP fusion proteins

To the cell soluble lysate, obtained from an overnight culture of the transformed host cells, was added a certain amount of free ELP. The amount of target fusion protein was roughly estimated using SDS-PAGE (Bio-Rad, USA). In order to achieve the best coaggregation efficiency, an amount of free ELP 10-20 times the amount of ELP-tagged proteins was added to the pre-treated cell lysate. The protein solution mixture was transferred to 50 mL plastic tube, and mixed thoroughly by repeatedly inverting the tube. The mixture volume was accurately measured, and the same volume of 5M NaCl was added to trigger the phase transition (high salt conditions). For each round of inverse transition cycling, two centrifugation steps were done (16,000×g for 5 minutes at room temperature). After centrifugation using high salt concentrations (high salt spin), the supernatant containing soluble contaminants was discarded, and the pellet was resuspended in cold PBS buffer with the same volume of initial lysate, followed by an additional centrifugation step (low salt spin). The pellet obtained after centrifugation with PBS contained particulate contaminants and the supernatant contained the resolublized free ELP / ELP-tagged mixture, which was then transferred to a new tube. The same procedure was used for the second and third rounds of ITC. Generally, protein solution that consists of only free ELP and ELP-tagged proteins would be obtained after three rounds of ITC. After overnight incubation with 20 mM DTT at 4°C in the refrigerator, one additional round of ITC was performed to separate the native target from ELPs (including intein-ELPs). The supernatant, containing the protein of interest, was desalted and concentrated by

centrifugation using an ultrafiltration membrane filter with a cutoff MW of 3 kDa (Pall Corporation, USA). The purity of the final sample was assessed by 10% SDS-PAGE, and stained with Coomassie Brilliant Blue.

2.8 Nanosep centrifugal device operation

A Nanosep® centrifugal device (Pall Corporation, USA) was used to remove salt contamination (e.g. NaCl and DTT) remaining in the protein solution. It could also be used for protein solution concentration. Because the filter material of this device is a low protein-binding Omega membrane (modified polyethersulfone membrane on polyethylene substrate), typical recoveries are greater than 90%. For the application of concentration, the final concentrate volume must be greater than 15 μL.

The experiment proceeded as follows. Ensure that the sample reservoir is firmly placed in the filtrate receiver. Pipette a certain volume (up to 500 μL) of the protein sample into the sample reservoir. Cap the Nanosep device. Place the Nanosep device into a fixed-angle centrifuge rotor. Spin up to 14,000×g for the required length of time to achieve the desired concentrate volume (Eppendorf 5415D centrifuge, Germany). At the end of spinning, stop the centrifuge and remove the Nanosep device. The concentrated sample is recovered with a micropipette. Table 2.1 below shows the relationship between the MWCO of the Nanosep device and required length of spin time (for 0.5 mL of a 1.0 mg / mL protein solution concentrated to a volume of 15 ~ 60 μL).

Table 2.1 Relationship of MWCO and spin time (Nanosep device technical manual)

MWCO	3K	10K	30K	100K	300K
Spin Time (min)	15	10	8	5	3

2.9 SDS-PAGE preparation

All expression and purification steps were validated by 10% SDS-PAGE in this study. The SDS gel and correlative solutions or buffers were prepared as shown in Table 2.2. The pH values were controlled by a VWR Symphony SB20 pH meter.

2.10 Purification of His-tagged protein with Ni-NTA agarose

For the purification of His_6-T_4-intein-ELP_{90} (to be described in chapter 3), the purity of the end product with 60 kDa molecular weight obtained by ITC is hard to analyze, using only SDS-PAGE. That is due to the fact that both the cleaved target protein part (His_6-T_4) and the other part of the fusion tag (intein-ELP_{90}) have similar molecular weights (about 60 kDa), and so SDS-PAGE cannot distinguish one from the other. In order to demonstrate that the end product is pure target protein without any contaminating fusion tag, affinity chromatography is employed, using the affinity of His-tag in the target protein.

Purification was performed on an immobilized metal ion affinity chromatography (IMAC) system, using a 1.5-ml volume column filled with Ni^{2+}-NTA agarose (Qiagen, USA) and equilibrated with binding buffer (20 mM Na_2HPO_4, 500 mM NaCl, pH 7.4). The sample was then applied to the pre-equilibrated column. Then, the column was

washed with two column volumes of the wash buffer (20 mM imidazole, 20 mM Na_2HPO_4, 500 mM NaCl, pH 7.4). Finally, the His-tagged target protein was eluted with a step-wise imidazole concentration gradient of 100~250 mM in buffer at 1 mL/min. The effluents and eluates were collected for further evaluation.

2.11 Plasmid DNA miniprep

All plasmids performed in this research were amplified and purified, using a QIAprep® miniprep kit (Qiagen, USA). This proceeded as follows. Resuspend pelleted bacterial cells (from 1-5 mL overnight cultures of *E. coli* in LB medium) in 250 µl buffer P_1, and transfer to a microcentrifuge tube. Add 250 µl buffer P_2 and mix thoroughly by inverting the tube 4~6 times (do not vortex). Then add 350 µl buffer N_3 and mix immediately and thoroughly by inverting the tube 4~6 times. Centrifuge for 10 minutes at 13,000×g rpm in a table-top microcentrifuge (Eppendorf 5415D centrifuge, Germany). Apply the supernatants from the last step to the QIAprep spin column by pipetting. Centrifuge for 30~60 s. Discard the flow-through. Wash the QIAprep spin column by adding 0.5 ml buffer PB and centrifuge for 30~60 s. Discard the flow-through. Wash the QIAprep spin column again by adding 0.75 mL buffer PE and centrifuge for 30~60 s. Discard the flow-through, and centrifuge for an additional 1 minute to remove residual wash buffer. Place the QIAprep column in a clean 1.5 mL microcentrifuge tube. To elute DNA, add 50 µL buffer EB or water to the center of each QIAprep spin column, let stand for 1 minute and centrifuge for another minute.

Table 2.2 Preparation of SDS gel and relative buffers

Name	Recipe	Notes
SDS-PAGE Separating Gel Buffer	Tris-base (Promega) 36.3 g 1 M HCl 48 mL Add MilliQ H_2O to 100 mL and Filter via 0.2 μm paper store at 4°C	3.0 M Tris-HCl pH 8.8
SDS-PAGE Stacking Gel Buffer	Dissolve 6.0 g Tris-base in water titrated with HCl to pH 6.8 Add water to final volume of 100 ml Filtration and storage	0.5 M Tris-HCl pH 6.8
10% Separating Gel	MilliQ water 4.21 mL Separating gel buffer 1.00 mL 10% SDS 80 μL 30% Acrylamide / Bis 2.67 mL 10% AP 40 μL, TEMED 4 μL	Total volume : 8.004 mL
Stacking Gel	MilliQ water 1.2 mL Stacking gel buffer 0.5 mL 10% SDS 20 μL 30% Acrylamide / Bis 0.27 mL 10% AP 10 μL, TEMED 2 μL	Total volume : 2.002 mL
SDS-PAGE Running Buffer (10X)	Tris-base 30.3 g Glycine 144 g, SDS 10 g Dissolve in MilliQ water to 1L store at 4°C	Dilute to 1X before use
SDS-PAGE Sample buffer (2X, with DTT)	0.5 M Tris-HCl (pH 6.8) 2.0 mL 10% SDS 2.0 mL, DTT 0.308 g 50% Glycerol 3.0 mL Bromophenol 5.0 mg 2.5 mL MilliQ water	Store at -20°C
Rapid Coomassie Blue Staining Solution	10% (v/v) acetic acid 0.006% (w/v) Coomassie brilliant blue G-250 90% H_20	
De-staining Solution	5% methanol 10% acetic acid 85% H_20	

2.12 Enzyme activity assay

2.12.1 Pfu I and Taq DNA Polymerase activity

The enzyme activity was analyzed by comparing the polymerizing ability of the enzyme produced with the corresponding commercial polymerase for the polymerase chain reaction of green fluorescent protein (GFP). Commercial Pfu DNA polymerase (Fermentas Co., 2.5 unit / μL) and commercial Taq DNA polymerase (Eppendorf Co., 5 unit / μL) were used.

Table 2.3 PCR system for green fluorescent protein

Reagent	Volume
H_2O	16.9 μL
DNA polymerase Buffer	2 μL
10 mM dNTP	0.4 μL
Forward primer	0.2 μL
Reverse primer	0.2 μL
Template ($pcAZ_4N$)	0.2 μL
Polymerase	0.1 μL
Total	20 μL

The thermal cycle was as follows: denaturation at 95°C for 2 mins, 25 amplification cycles of denaturation at 95°C for 1 min, annealing at 61.5°C for 1 min, and extension at 72°C for 2 mins followed by a final extension at 72°C for 5 mins. The PCR product was then analyzed by 1% agarose gel electrophoresis (UVP BioDoc-It system, USA).

2.12.2 T_4 DNA Ligase activity

The weiss unit of T_4 DNA ligase is defined as 100 times the amount of enzyme required to catalyze the ligation of more than 95% of 1 μg of λ / Hind III fragments at 16°C in 20 minutes. The enzyme activity is analyzed by comparing the ligation effect of enzyme produced with the same commercial ligase as control. Commercial T_4 DNA ligase used came from Promega Company (3.0 unit / μL).

The ingredients 1 μL λ / Hind III fragments, 1 μL T_4 DNA ligase buffer, 7 μL H_2O and 1μL T_4 DNA ligase were added in order to the 200 μL microcentrifuge tube, which was then capped. After 20 minutes at 16°C the reaction was complete. 1% agarose gel electrophoresis was then used for inspection of the reaction product.

Table 2.4 Reaction system of λ / Hind III fragments ligation

Volume	Reagent
1 μL	λ / Hind III fragments (New England, USA)
1 μL	T_4 DNA Ligase buffer (Promega, USA)
7 μL	H_2O
1 μL	T_4 DNA Ligase (Promega, USA)
10 μL	Total Volume

3 Results and discussions

Since the invention of the target-intein-ELP system, most of the scientific literature has reported all the good aspects of this purification method (e.g. simple, cheap, protease-free and chromatography-free). In the present study however, through comparing the expression and purification of four different Intein-ELP fusion proteins (Trx-thrombin-pfu I -intein-thrombin-ELP, His_6-T_4-intein-ELP, His_6-Taq-intein-ELP and His_6-Pfu I -intein-ELP), several bad responses of this protein purification system became apparent. For example, we found that Intein-ELP recombinant proteins do not always have a soluble expression. Sometimes they form inclusion bodies instead, so that we cannot use the simple non-chromatography method of inverse transition cycling to purify them. Also, the poor expression level is another factor that limits the application of ITC. By studying this project, we propose an appropriate solution to these problems that makes a significant contribution to further expanding the utility of ELP-based protein purification.

3.1 Trx-thrombin-pfu I-intein-thrombin-ELP$_{90}$ (Trx-pfu I-ELP)

The transformed cells were cultivated in enriched TB media, supplemented with 100 μg/mL ampicillin. Shaker flasks were inoculated from a single colony and cultured overnight at 37°C with 180 rpm (Shel Lab incubator, 1575R, Sheldon Manufacturing Inc., USA). Since Urry and his colleagues have proved that expression of ELP-tagged proteins is greatly enhanced in the absence of inducer IPTG, strain cells grow with a leaky expression in this experiment (Guda et al., 1995).

The bacterial cells were harvested by centrifugation at 5000×g, 4°C, for 10 minutes. The collected pellet was resuspended in cold 1×PBS buffer (137 mM NaCl, 2.7 mM KCl, 4.2 mM Na_2HPO_4, and 1.4 mM KH_2PO_4, pH 7.4) in a volume ratio of 1:25 to the initial volume of the culture medium. After two more cycles of washing with the same 1×PBS buffer, the resuspended cells were then disrupted by introducing discontinuous ultrasonic pulses (VirTis Company, USA) in an ice-water bath for 10 minutes, at the maximum power at which no substantial amount of foam was generated. The lysate was then centrifuged at 16000×g, 4°C, for 10 minutes, to remove insoluble debris.

The volume of soluble lysate was accurately measured and one sample of volume 5 M NaCl stock was added to initiate reversible inverse phase transition at room temperature (ensuring that the amount of salt used was enough for adequate precipitation). The resulting sample was then centrifuged at 16,000×g, 4°C, for 10 minutes, and the supernatant part was discarded to remove contaminating *E. coli* proteins. The phase transition was reversed using cold 1×PBS buffer with the same volume of initial lysate, and the sample was re-centrifuged at 16,000×g, 4°C, for 10 minutes. This time the supernatant part was collected as fusion protein solution, and the pellet part obtained after centrifugation contained particulate contaminants was discarded. We call this method "inverse transition cycling" (ITC). Trx-thrombin-pfu I -intein-thrombin-ELP_{90} was purified after three rounds of ITC.

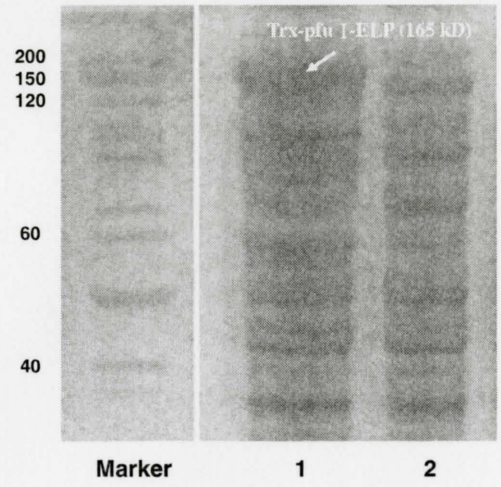

Figure 3.1 Expression of recombinant Trx-thrombin-pfu I -intein-thrombin-ELP_{90} at 37°C. Lane 1 corresponds to the cleared lysate of cultured *E. coli* cells expressing the fusion Trx-pfu I -ELP. Lane 2 corresponds to the insoluble lysate of cultured cells expressing Trx-pfu I -ELP.

Trx-thrombin-pfu I -intein-thrombin-ELP_{90} (165 kDa, referred to below as Trx-pfu I -ELP) was soluble expressed in *E. coli* host cells at 37°C with high yield, as shown in Figure 3.1. The SDS-PAGE results reveal that hardly any inclusion body was formed in the expression process, which means that 37°C is the optimal temperature for Trx-pfu I -ELP expression. The molecular weight of Trx-pfu I -ELP is 165 kDa. Adding reducing agent such as DTT into pure Trx-pfu I -ELP solution could stimulates cleavage at the N-terminal of intein (due to intein molecule modification of $N^{315} \rightarrow A^{315}$, its C-terminal cleavage capacity is blocked) and releases native target from the fusion. After another round of ITC and dilution, the protein of interest was finally obtained. The purity of the target protein was characterized using SDS-PAGE followed by staining with Coomassie Brilliant Blue.

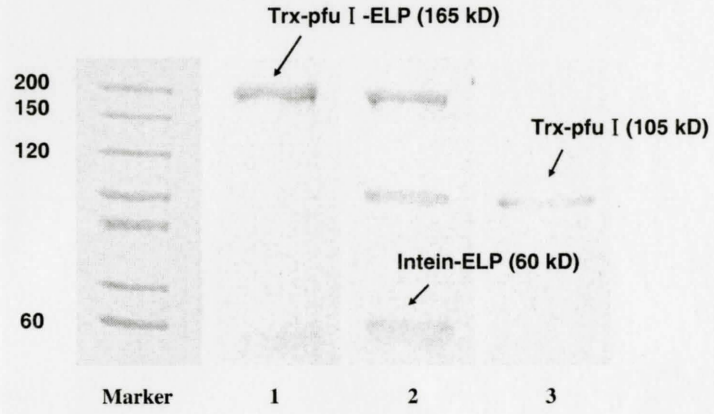

Figure 3.2 Purification of recombinant Trx-thrombin-pfu I -intein-thrombin-ELP_{90} at 37 °C. Lane 1 corresponds to pure Trx-pfu I -ELP solution which was purified by the method of ITC from soluble lysate. Lane 2 corresponds to Trx-pfu I -ELP solution overnight cleavage by thiol addition. Lane 3 corresponds to the final Trx-pfu I solution after one more round of ITC and dilution.

3.2 His_6-T_4-intein-ELP_{90} (T_4-ELP)

T_4 DNA ligase, isolated from the bacteriophage T_4, which can link together DNA strands that have double-strand breaks (a break in both complementary strands of DNA), has become an indispensable tool in modern molecular biology research. In this section, the production of T_4 target ligase from T_4-ELP fusion is described.

From the results shown in Figure 3.3, we can see that the expression of His_6-T_4 -intein-ELP_{90} (called T_4-ELP for short) is also very high following overnight culturing at 37°C. However, there is a severe problem: nearly all T_4-ELP is expressed as insoluble inclusion bodies. Because the inverse transition cycling method is based on the soluble protein lysate, we cannot use ELP specificity to purify T_4-ELP as we did with Trx-pfu I -ELP.

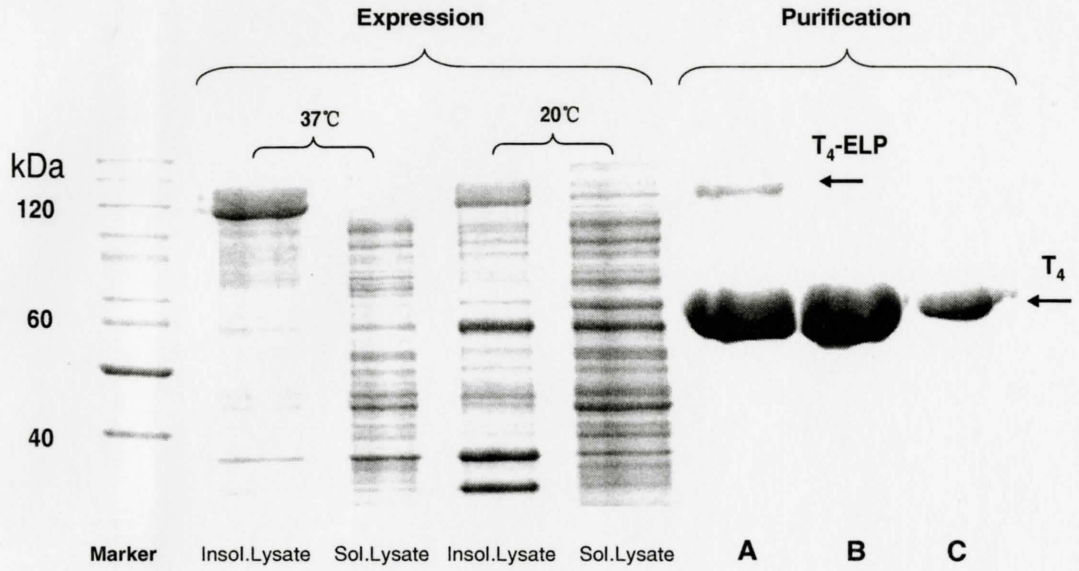

Figure 3.3 SDS-PAGE of expression, and each step of the T_4-ELP purification process by ITC. (A) After three rounds of ITC, we get a mixture of T_4-ELP and intein-ELP from soluble lysate culturing at 20 °C for 3 days; (B) Overnight cleavage with 20 mM DTT, T_4-ELP breaks down to His_6-T_4 and intein-ELP; (C) Another round of ITC is performed to isolate His_6-T_4 from intein-ELP. The purified His_6-T_4 is then desalted and concentrated by an ultrafiltration centrifugation filter with a MW cutoff of 3 kDa (Pall Corp, USA). The amount of protein loaded in lane A and B correspond to 0.15% of the 1-L culture volume. Lane C corresponds to 3% of the 1-L culture volume. The intensity of the protein bands in each lane is proportional to the protein content in that sample.

Inclusion bodies are classically believed to contain misfolded and partially folded proteins. Two factors are generally considered to have a close relationship with the presence of inclusion bodies. One is the rate of protein synthesis, and the other is the rate of protein folding (Catherine, 1989). In general, the cultivation temperature of *E. coli* seems critically important for heterologous gene expression. The higher the cultivation temperature, the faster the recombinant protein produced. However, if the

rate of protein production is too fast compared to the rate of protein folding, target proteins will not have enough time to fold correctly, so that rapid and incorrect folded proteins will aggregate as inclusion bodies. Despite the loss of total expression yield and the increase in time needed, culturing at a lower temperature was reported to promote the soluble expression of genes (Hanning and Makrides, 1998). In the present study, we also applied low-temperature expression strategy to solve the problem of inclusion bodies.

The transformed *E. coli* cells for T_4-ELP expression were cultivated in TB media, supplemented with 100 μg/mL kanamycine. Shaker flasks were inoculated from a single colony and cultured at 20°C with 180 rpm for 3 days. The other experimental protocols are similar to those described in the Trx- pfu I -ELP section.

With overnight culturing at 37°C, all the T_4-ELP proteins (arrow mark at 120 kDa in Figure 3.3) are insoluble expressed; however, after lowering the culturing temperature from 37°C to 20°C and prolonging the culturing time from overnight to three days, the amount of inclusion body formed was significantly decreased, and the lost T_4-ELP inclusion bodies were partially transformed into soluble expression. The soluble T_4-ELP expressed at the lower temperature is sufficient for purification using the method of inverse transition cycling. After three rounds of ITC, we obtained a mixture of T_4-ELP and intein-ELP as indicated in lane A of Figure 3.3. According to the specific feature of the modified intein tag in all four fusion proteins used in our study, adding a reducing agent such as DTT to the solution of lane A can stimuli

cleavage of intein tag at its N-terminus and separate His_6-T_4 from intein-ELP, as shown in lane B of Figure 3.3. Because molecular weight of His_6-T_4 and intein-ELP are both about 60 kDa, only one protein band is shown in lane B. Due to the cleavage behavior of T_4-ELP, the intensity of the protein band of lane B is a little greater than in the same position in lane A. After applying three more rounds of ITC on the solution of lane B, contaminative intein-ELP was removed as pellets from the mixture by centrifugation. The soluble protein solution was then eliminated residual salt content (from DTT and NaCl) and concentrated using an ultrafiltration centrifugation filter with a MW cutoff of 3 kDa (Pall Corp, USA). We finally recovered the native target protein His_6-T_4 with MW of 60 kDa as indicated in lane C of Figure 3.3.

As stated above, because His_6-T_4 and intein-ELP have the same molecular weight of 60 kDa, we cannot prove that the final product of lane C in Figure 3.3 was pure His_6-T_4 with only a single protein band of SDS-PAGE. Under this circumstance, we used the affinity property of His_6 tag to validate the purity of the desired protein. We performed the lane C unknown protein solution of Figure 3.3 on an immobilized metal ion affinity chromatography (IMAC) system, using a 1.5-ml volume column filled with Ni^{2+}-NTA agarose (Qiegen, USA) and equilibrated with binding buffer (20 mM Na_2HPO_4, 500 mM NaCl, pH 7.4). The unknown protein sample was then applied to a pre-equilibrated affinity column. The flow-through was collected for further analysis. After that, the column was washed with two column volumes of the wash buffer (20 mM imidazole, 20 mM Na_2HPO_4, 500 mM NaCl, pH 7.4) to eliminate the nonspecific binding of impurity. Finally, the His_6-T_4 protein was eluted

with a step-wise imidazole concentration gradient of 100 ~250 mM in buffer at 1 mL/min. The eluates were also collected for further evaluation. The entire IMAC process was shown in Figure 3.4-A.

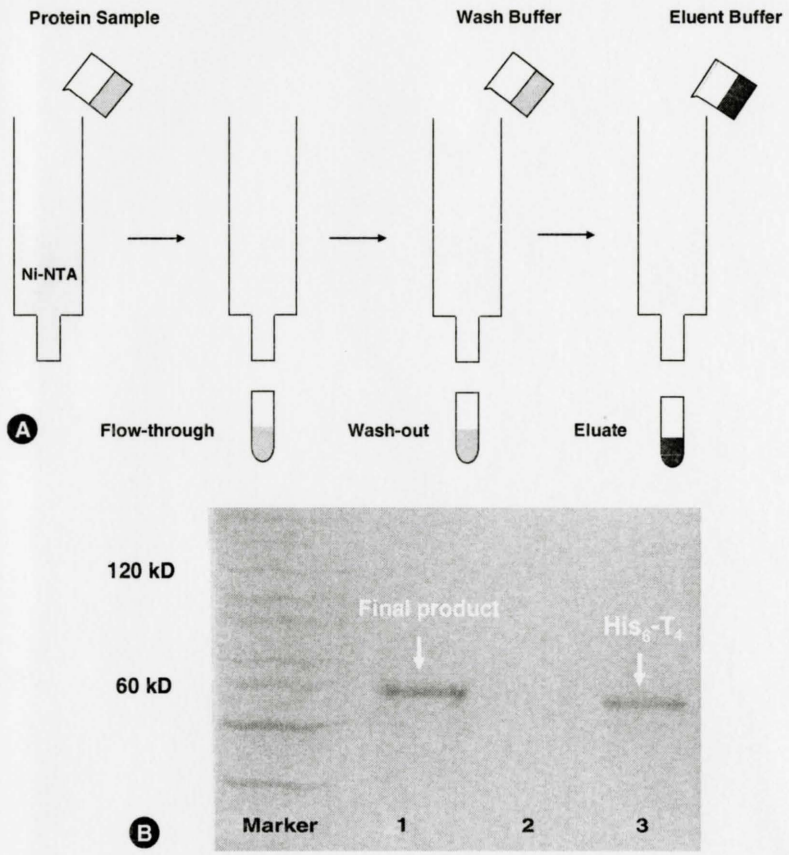

Figure 3.4: (A) His-tagged protein purified by IMAC; (B) SDS-PAGE of His-tagged protein purification by IMAC. Lane 1 is the protein sample (final product) before being loaded to the column. Lane 2 is concentrated flow-through with the same volume of initial sample. Lane 3 is eluted protein with the same volume of initial sample

Figure 3.4-B indicates that the unknown protein sample comes from lane C of Figure 3.3 containing only His_6-T_4. In order to normalize the analytic result, the collected effluents (flow-through) and eluates were concentrated to the same volume as the initial protein sample prior to loading by an ultrafiltration centrifugation filter

(Pall Corp, USA). Same of the original protein sample, concentrated flow-through and concentrated eluates were loaded to 10% SDS-PAGE as shown in lane 1, 2 and 3 of Figure 3.4-B respectively. We can clearly see that no protein band was presented in the flow-through lane, which means that all unwanted intein-ELP was removed from the protein solution through repeated ITC. Moreover, by comparing intensities of protein bands, we noticed that the amount of eluted protein His_6-T_4 from lane 3 was almost the same as the amount of protein contained in the initial sample solution before loading from lane 1. This result confirms the conclusion we obtained previously.

T_4-ELP self-cleavage in the course of culturing at 20°C occurs without any external chemical stimuli, much more seriously than at 37°C. From the SDS-PAGE result of Figure 3.3, we found that the amount of 60 kDa proteins (both in terms of soluble lysate and insoluble lysate) expressed at 20°C was much greater than the amount of 60 kDa proteins expressing at 37°C. After inverse transition cycling treatment of cell free extracts, a large fraction of the ELP-intein fusions does not contain His_6-T_4 (as shown in lane A of Figure 3.3). Much of the already translated T_4-ELPs was self-cleaved before recovery. This results in a loss of yield. To date, we do not quite understand the exact reason why only T_4-ELP (but not other ELP fusions considered later in this study) underwent serious premature cleavage in vivo during expression at low temperature. It would be valuable to find out more about the influence of expression temperature and passenger partners on the self-cleavage behavior of intein tag in future work. We anticipate that the problem of premature cleavage might be solved by the use of ligand-activated inteins mentioned in the first

chapter. Only by solving this problem can the application of self-cleavage stimulus responsive tags become significant for recombinant protein purification in the industrial field.

3.3 Activity assay of His_6-T_4 DNA ligase

T_4 DNA ligase, isolated from bacteriophage T_4, which can link together DNA strands that have double-strand breaks, has become an indispensable tool in modern molecular biology research. The mechanism of DNA ligase is to form covalent phosphodiester bonds between 3' hydroxyl ends of one nucleotide with the 5' phosphate end of another. ATP is required for the ligase reaction. Figure 3.5 below is a pictorial example of how a T_4 DNA ligase works with sticky ends. T_4 DNA ligase will also work with blunt ends, although higher enzyme concentrations and different reaction conditions are required.

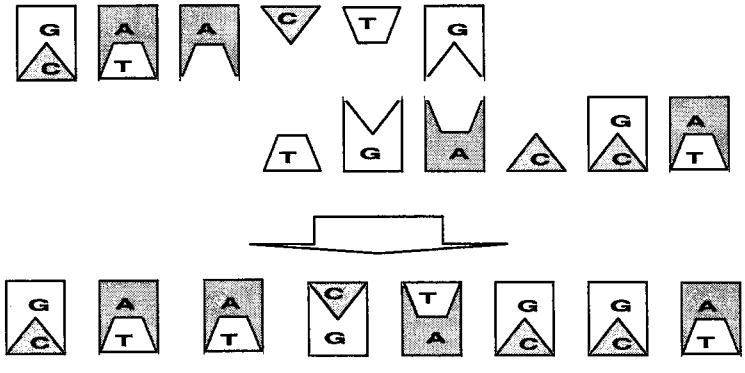

Figure 3.5 Sketch map of how DNA ligase works (with sticky ends)

Note that, 0.01 Weiss unit of T_4 DNA ligase is defined as the amount of enzyme required to catalyze the ligation of more than 95% of 1 μg of λ / Hind III fragments at

16°C in 20 minutes. We put, in order, 1 μL λ / Hind III fragments (New England Lab, 1 μg/μL), 1 μL T_4 DNA Ligase buffer, 7 μL H_2O, and 1 μL T_4 DNA ligase into a 200 μL microcentrifuge tube, cap the tube, put it in 16°C water bath and leave it to react completely for 20 minutes. After reaction, 1% agarose gel electrophoresis was performed for inspection of the product.

Commercial T_4 DNA ligase (Promega, 3 unit/μL) was used in our experiment as control. The left most lane of Figure 3.6-A shows the ligation effect of 0.01 unit commercial T_4 DNA ligase (formed from commercial enzyme diluted 300 times, with 1 μL added to the reacting system), and lanes 1-4 show the ligation effect of 1 μL of enzyme produced with gradient concentration, specifically, their concentrations of 2×10^{-3} μg/μL, 4×10^{-3} μg/μL, 8×10^{-3} μg/μL, and 16×10^{-3} μg/μL respectively. We can see that reaction effect of lane 2 is similar to that of 0.01 unit of commercial enzyme. Accordingly, we calculated that the specific activity of purified enzyme is about 2500 U/mg, consistent with the previously reported specific activity of T_4 DNA ligase in the range 2100~3600 U/mg.

Figure 3.6-B compares the ligation effect of commercial and produced ligase. Both of these can effectively ligate λ/Hind III fragments at a concentration of 3 U/μL. Note that the ligation effect of the commercial sample seems a little better than the produced His_6-T_4. This result reveals that ITC is a simple and convenient method for intein-ELP tagged protein purification with little influence on the activity of the target protein.

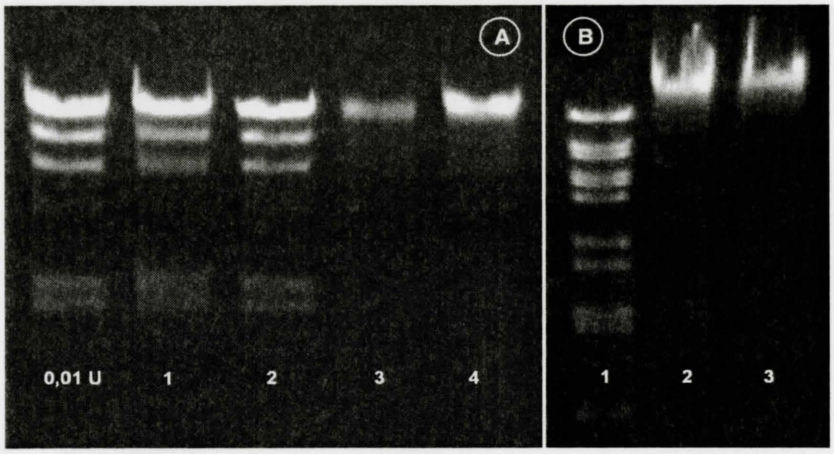

Figure 3.6 : Ligation comparison between commercial and produced ligase (A) The left most lane is λ/Hind III fragments ligated with 0.01U commercial T_4, lanes 1-4 show λ/Hind III fragments ligated with varying concentrations of purified T_4, namely, 2×10^{-3} μg/μL, 4×10^{-3} μg/μL, 8×10^{-3} μg/μL, and 16×10^{-3} μg/μL respectively; (B) Lane 1 is λ/Hind III fragments, lanes 2 and 3 are λ/Hind III fragments ligated with 3U/μL commercial T_4 and purified His_6-T_4 respectively.

3.4 His_6-Pfu Ⅰ-intein-ELP_{90} (Pfu Ⅰ-ELP) & His_6-Taq-intein-ELP_{90} (Taq-ELP)

Pfu and Taq DNA polymerase are thermostable DNA polymerases isolated from thermophilic bacteria living in extreme environments, and frequently used in polymerase chain reaction (PCR). In the present study, we also produced His_6-pfu Ⅰ and His_6-Taq from the ELP-based protein purification system.

Similarly, intein-ELP fusion proteins, incorporated with a DNA polymerase like Pfu Ⅰ or Taq also form inclusion bodies when expressed at 37°C. So naturally, the first idea that came to our mind was to decrease the culturing temperature, as we did with T_4-ELP expression. Unfortunately, it seems that this low-temperature expression strategy does not work well for this situation. When decreasing the culturing

temperature from 37°C to 20°C and prolonging the culturing time from overnight to 3 days, although this indeed promoted the soluble expression of Pfu I -ELP and Taq-ELP slightly, the amount of soluble fusions were far from sufficient for purification, using the original format of ITC.

A poor expression level of heterologous ELP-tagged protein occurs from time to time (Chilkoti et al. 1999, 2004). Researchers have previously reported that if an ELP tagged molecule is present in complex mixtures at a very low concentration, then adding an excess amount of free ELP to the sample and inducing reversible inverse phase transition would form hybrid aggregates via the interaction of ELP moieties of the two molecules. The mechanism is based on the coaggregation of free ELP and ELP tailed proteins, which has been reported in connection with the purification of a fusion of a single-chain Fv of an antiatrazine antibody to ELP (Kim et al., 2003, 2005). Later the efficiency of this method was quantitatively prooved by Ge. (Ge and Filipe, 2006). Through low temperature expression, combined with adding certain amounts of free ELP into the soluble lysate of cultured *E. coli* cells, we successfully set up an efficient system for capturing a low expression yield of Pfu I -ELP and Taq-ELP.

The detailed experimental procedures for expressing of Pfu I -ELP and Taq-ELP at 20°C were the same as we described before. The cultured *E. coli* cells were harvested by centrifugation (5,000×g, 4°C, 10 minutes) and the pellet was resuspended using a cold PBS buffer in 4% of the initial volume of the culture medium. After washing twice with PBS, the supernatant was disrupted by applying discontinuous

ultrasonic pulses (VirTis Company, USA) in an ice-water bath (10 minutes). The lysate was then re-centrifuged (5,000×g, 4°C, 10 minutes) and the clarified lysate was retained for further purification. We could then estimate the amount of target protein contained in the cell free extract by 10% SDS-PAGE. In order to coaggregate throughly, free ELP (10-20 times as much as the target protein) was added to the pre-treated cell lysate. The mixture protein solution was transferred to a fresh centrifuge tube and mixed thoroughly by repeatedly inverting the tube. The purification steps following this were exactly the same as given previously.

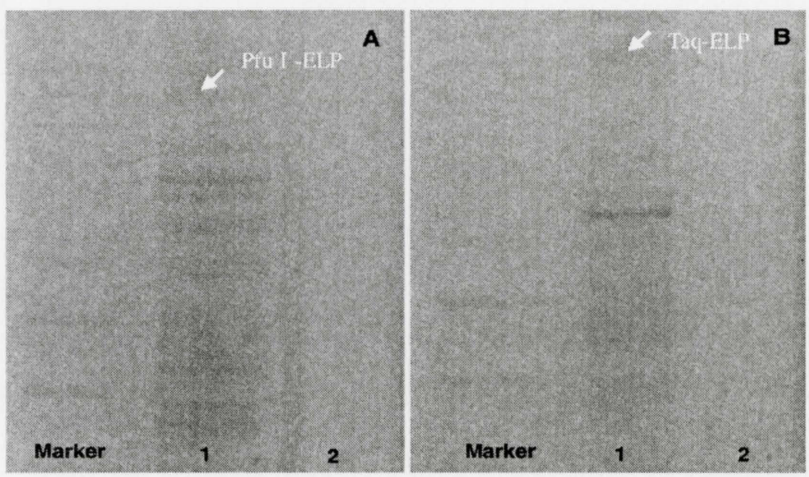

Figure 3.7 Purification without ELP supplementation. (A) Lane 1 is soluble lysate of Pfu Ⅰ-ELP, lane 2 is the lysate solution after one round of ITC. (B) Lane 1 is soluble lysate of Taq-ELP, lane 2 is the lysate solution after one round of ITC.

Figure 3.7 shows the purification of Pfu Ⅰ-ELP and Taq-ELP using ITC without ELP supplementation. Although correlative protein bands could be seen in 10% SDS-PAGE of initial soluble lysate at the correct position (Pfu Ⅰ-ELP ~ 145 kDa, Taq-ELP ~ 155 kDa), barely any protein band appeared when the sample was treated

with one round of ITC as shown above. When adding NaCl to the sample to start the ITC, even though the clear solution still turn turbid, no aggregated Pfu I -ELP (or Taq-ELP) could be effectively pelleted because of the extremely low concentration.

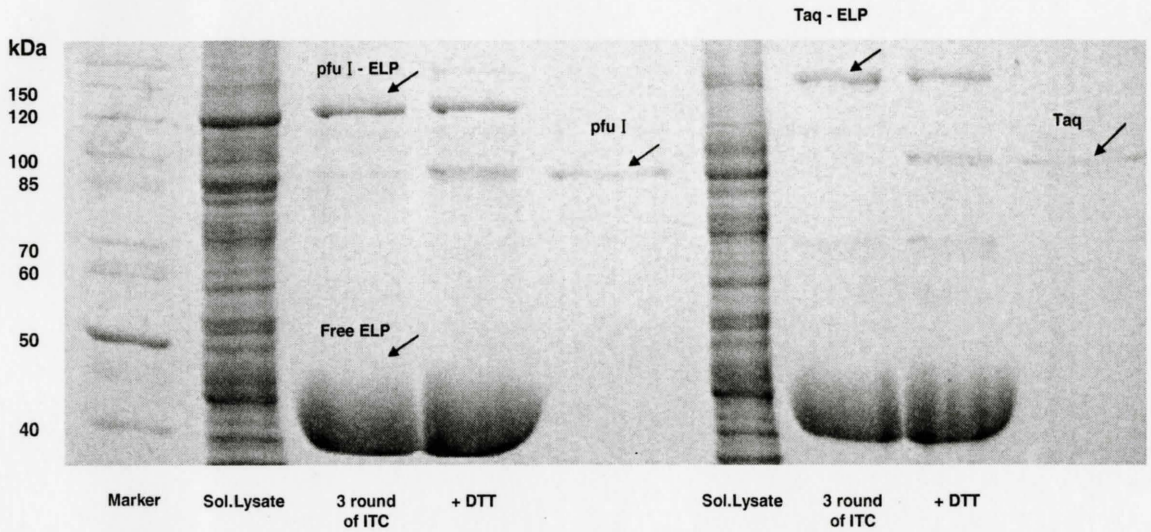

Figure 3.8 SDS-PAGE of each step of the purification process of Pfu I-ELP and Taq-ELP

For the sake of comparison, Figure 3.8 indicates the purification result of each step using the method of ITC under the participation of free ELP. After three rounds of ITC on the mixture of free ELP and soluble lysate, Pfu I -ELP (145 kDa) was captured successfully from hundreds of impurities in the clarified lysate of transformed *E. coli* cells by free ELP. After incubation with 20 mM DTT for a whole night, part of the Pfu I -ELP broke up into His_6-Pfu I and intein-ELP. This explains why the protein band at 145 kDa becomes weak and protein bands represent His_6-Pfu I and intein-ELP were enhanced accordingly. After additional rounds of ITC treatment and dilution, we got pure fusion protein of His_6-Pfu I with the expected molecular weight of 85 kDa. The purification process of His_6-Taq (95 kDa) was

similar. Both expressed at a low temperature, premature cleavage of intein in vivo was not as serious as with T_4-ELP, and not much fusion protein was lost by incorporating the intein in the product. This phenomenon suggests that the premature cleavage of intein is somehow associated with the identity of the target protein.

3.5 Activity assay of His_6-Pfu I and His_6-Taq

Pfu and Taq DNA polymerase are thermostable DNA polymerases isolated from the thermophilic bacteria living in extreme environments, and frequently used in polymerase chain reaction (PCR). Unlike Taq DNA polymerase, Pfu DNA polymerase possesses 3' to 5' exonuclease proofreading activity, meaning that it works its way along the DNA from the 5' end to the 3' end and corrects nucleotide-misincorporation errors. This means that Pfu DNA polymerase-generated PCR fragments will have fewer errors than Taq-generated PCR inserts. As a result, Pfu is more commonly used nowadays for molecular cloning of PCR fragments.

We compared the activity of our two produced polymerases with commercial samples by PCR effect with Green fluorescent protein (GFP): commercial Pfu from the Fermentas Company (2.5 unit / μL) and commercial Taq from the Eppendolf Company (2.5 unit / μL). We put, in order, 16.9 μL H_2O, 2 μL DNA polymerase buffer, 0.4 μL 10 mM dNTP, 0.2 μL forward primer, 0.2 μL reverse primer, 0.2 μL template ($pcAZ_4N$) and 0.1 μL DNA polymerase into a 200 μL centrifuge tube for a polymerase chain reaction. The thermal cycle was as follows: denaturation at 95°C for 2 minutes; 25 amplification cycles of denaturation at 95°C for 1 minute, annealing at

61.5°C for 1 minute, and extension at 72°C for 2 minutes, followed by a final extension at 72°C for 5 minutes. The PCR product was then inspected by 1% agarose gel electrophoresis.

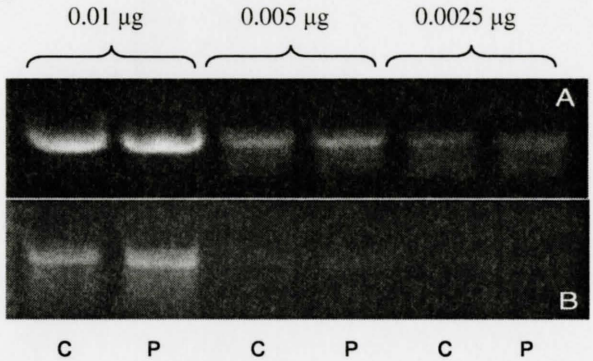

Figure 3.9: (A) PCR of Taq polymerase of varying concentration, from left to right: 0.01 μg, 0.005 μg, and 0.0025 μg; (B) PCR of Pfu polymerase of varying concentration, from left to right: 0.01 μg, 0.005 μg, and 0.0025 μg. 'C' means commercial enzyme and 'P' means produced enzyme.

As Figure 3.9 shows, using the same amount of produced and commercial enzymes, the PCR effects are similar. Their PCR capacities are both attenuated with a decrease in the enzyme used. Using same amount of polymerase, the PCR effect of Taq was a little better than that of Pfu DNA polymerase. The result reveals once more that there is hardly any influence on the activity of the target protein using ITC.

4 Conclusions and Recommendation for future work

Through comparing expression and purification of four different self-cleavage ELP fusion proteins, two common limitations associate with ELP-intein protein purification system were investigated in this research. From the results, several conclusions can be drawn.

Trx-thrombin-pfu I -intein-thrombin-ELP_{90} (165 kD) was soluble expressed in the absence of inducer IPTG at 37 °C. This means we could use the inverse transition cycling (ITC) method to purify this fusion. With overnight expressing, a high yield of Trx-pfu I -ELP was obtained. No inclusion bodies were formed in this case. Based on the ELP-intein system, 105 kD of target protein (Trx-pfu I) was finally obtained.

The expression of His_6-T_4-intein-ELP_{90} (120 kD) is also very high with overnight culturing at 37°C. However, a severe problem occurs here: nearly all T_4-ELPs are expressed as the formation of insoluble inclusion bodies. Because the ITC method is based on soluble lysate, we cannot use ELP property to purify T_4-ELP as we did with Trx-pfu I -ELP. This problem can be solved by a low-temperature expression strategy. After lowering the temperature from 37°C to 20°C and prolonging the culturing time from overnight to 3 days, the amount of IBs formed was decreased significantly, and the lost T_4-ELP IBs were partly transformed into soluble expression. The yield is enough for purification using the ITC method. In this way, 60 kD of target protein was obtained without any contamination. In addition, T_4-ELP underwent serious premature cleavage in vivo during expression at low temperature.

His_6-Pfu I -intein-ELP_{90} (145 kD) and His_6-Taq-intein-ELP_{90} (155 kD) also form IBs when expressing at 37°C. Using a low-temperature strategy alone helped little for this situation, because although it indeed promoted soluble expression of Pfu I -ELP and Taq-ELP slightly, the amount of soluble fusions was far from sufficient for purification using the original format of ITC. However, using a low temperature together with a certain amount of free ELP added to the soluble lysate of cultured *E. coli* cells, we have successfully set up an efficient system for capturing low concentration ELP fusions. For the two fusions, premature cleaving is not as serious as with T_4-ELP. This suggests that this behavior of intein somehow depends on the identity of the particular target protein in the fusion.

Activity assay provided direct evidence that the bio-activities of target proteins purified by the ELP-intein system were well maintained. The specific activities of produced targets were identical to that of commercial enzymes.

Based on the results of the work reported in this thesis, a detailed investigation of the in vivo premature cleavage behavior of intein is recommended for future work. That undesirable occurrence results in a great loss of yield. It would be valuable to find out more about the influence of expression temperature and passenger partners on the self-cleavage of intein tag. We also predict that the premature cleavage might be solved by the use of ligand-activated inteins. Only by solving this problem, can the application of self-cleavage stimulus responsive tags become significant for recombinant protein purification in a general context.

5 References

Allen JM, Cook GM, Poole AR. (1971) Action of concanavalin A on the attachment stage of phagocytosis by macrophages. *Exp Cell Res.* 68 (2), 466-467.

Angal S, Harris EL. (1989) Protein purification methods: a practical approach. *Oxford University Press*, ISBN: 019963002X.

Anna M, Jonathan K. (1989) Protein folding intermediates and inclusion body formation. *Nature.* 7, 690-697.

Beekman JM, Cooney AJ, Elliston JF, Tsai SY, Tsai MJ. (1994) A rapid one-step method to purify baculovirus-expressed human estrogen receptor to be used in the analysis of the oxytocin promoter. *Gene.* 146, 285-289.

Brizzard, Chubet RG, and Vizard DL. (1994) Immunoaffinity purification of FLAG epitope-tagged bacterial alkaline phosphatase using a novel monoclonal antibody and peptide elution. *BioTechniques.* 16, 730–734.

Catherine Schein. (1989) Production of soluble recombinant proteins in bacteria. *Nature.* 7, 1141-1148.

Carrington JC, Dougherty WG. (1988) A viral cleavage site cassette: identification of amino acid sequences required for tobacco etch virus polyprotein processing. *Proc Natl Acad Sci USA.* 85, 3391-3395.

Chaga G, Hopp J, Nelson P. (1999 a) Immobilized metal ion affinity chromatography on Co^{2+}-carboxymethylaspartate-agarose superflow, as demonstrated by one-step purification of lactate dehydrogenase from chicken breast. *Biotechnol Appl Biochem.* 29, 19-24.

Chaga G, Bochkariov DE, Jokhadze GG, Hopp J, Nelson P. (1999 b) Natural poly - histidine affinity tag for purification of recombinant proteins on cobalt (II) - carboxymethylaspartate crosslinked agarose. *J Chromatogr A.* 864, 257-256.

Chang JY. (1985) Thrombin specificity. Requirement for apolar amino acids adjacent to the thrombin cleavage site of polypeptide substrate. *Eur J Biochem.* 151, 217-224.

Chilkoti A, Trabbic-Carlson K, Liu L, Kim B. (2004 a) Expression and purification of recombinant proteins from Escherichia coli: Comparison of an elastin-like polypeptide fusion with an oligohistidine fusion. *Protein Sci.* 13, 3274-3284.

Choi SI, Song HW, Moon JW, Seong BL. (2001) Recombinant enterokinase light chain with affinity tag: expression from *Saccharomyces cerevisiae* and its utilities in fusion protein technology. *Biotechnol Bioeng.* 75, 718-724.

Chong S, Yang, Paulus H, Benner J, Perler FB, Xu MQ. (1996) Protein splicing involving the *Saccharomyces cerevisiae* VMA intein: the steps in the splicing pathway, side reactions leading to protein cleavage, and establishment of an in vitro splicing system. *J. Biol. Chem.* 271, 22159-22168.

Chong S, Mersha FB, Comb DG, Scott ME, Landry D, Vence LM, Perler FB, Benner J, Kucera RB, Hirvonen CA, Pelletier JJ, Paulus H, Xu MQ. (1997) Single-column purification of free recombinant proteins using a self-cleavable affinity tag derived from a protein splicing element. *Gene.* 192, 271-281.

Christoph S, Sievert V, Bussow K. (2003) An automated method for high-throughput protein purification applied to a comparison of His-tag and GST-tag affinity chromatography. *BMC Biotechnology.* 3 (12), 1-8.

Christopher N, Jimin W, Francine BP. (2000) Dissecting the chemistry of protein splicing and its applications. *Angew. Chem. Int. Ed.* 39, 450-466.

Cleland W. (1964) Dithiothreitol, a new protective reagent for SH groups. *Biochemistry*. 3, 480-482.

Creighton, Thomas E. (1940) Proteins: structure and molecular properties. *New York: W.H. Freeman*. ISBN: 016723174.

Davis GD, Elisee C, Newham DM, Harrison RG. (1999) New fusion protein systems designed to give soluble expression in E. coli. *Biotechnol. Bioeng*. 65, 382-388.

De Marco A. (2004) Bacteria co-transformed with recombinant proteins and chaperones cloned in independent plasmids are suitable for expression tuning. *J Biotechnol*. 109, 45-52.

Dougherty WG, Parks TD, Cary SM, Bazan JF, Fletterick RJ. (1989) Characterization of the catalytic residues of the TEV 49-kDa proteinase. *Virology*. 172, 302-310.

Drucker M, German S, Esperandieu P, LeGall O, Blanc S. (2002) Purification of viral protein from infected plant tissues using the Strep-tag. *Botechnol Intl*. June, 16-18.

Duplay P, Bedouelle H, Fowler A, Zabin I, Saurin W, Hofnung M. (1988) Sequence of male gene and of its product, the maltose binding protein of Escherichia coli K12. *J Biol Chem*. 259, 10606-10613.

Einhauer, Jungbauer A. (2001) The FLAG peptide, a versatile fusion tag for the purification of recombinant proteins, *J. Biochem. Biophys. Methods*. 49, 455-465.

Evans T, Benner J, Xu MQ. (1999) Semisynthesis of cytotoxic proteins using a modified protein splicing. *J Biol Chem*. 274, 18359-18363.

Feher, P. Boross, T. Sperka, S. Oroszlan, J. Tozser. (2004) Expression of the murine leukemia virus protease in fusion with maltose-binding protein in Escherichia coli, *Protein Expr. Purif*. 35, 62-68.

Fontaine L, Meynial-Salles I, Girbal L, Soucaille P. (2002) Molecular characterization and transcription analysis of adhE2, the gene encoding the NADH dependent aldehyde dehydrogenase responsible for butanol production in alcohologenic cultures of Clostridium acetobutylicum ATCC 824. *J Bacteriol.* 184, 821-830.

Ge X, Yang D.S, Trabbic-Carlson K, Kim B, Chilkoti A, Filipe C.D.M. (2005) Self-Cleavable stimulus responsive tags for protein purification without chromatography. *J. Am.Chem.Soc.* 127, 11228-11229.

Ge X, Trabbic-Carlson K, Chilkoti A, Filipe C.D.M. (2006) Purification of an elastin-like fusion protein by microfiltration. *Biotech and Bioeng.* 95 (3), 424-432.

Ge X, Filipe C.D.M. (2006) Simultaneous phase transition of ELP tagged molecules and free ELP: an efficient and reversible capture system. *Biomacromolecules.* 7, 2475-2478.

Guan CD, Li P, Riggs PD, Inouye H. (1988) Vectors that facilitate the expression and purification of foreign peptides in E. coli by fusion to MBP. *Gene.* 67, 21–30.

Guan K, Dixon JE. (1991) Eukaryotic proteins expressed in E. coli: an improved thrombin cleavage and purification procedure of fusion proteins with glutathione S-transferase. *Anal Biochem.* 192, 262-267.

Guda C, Zhang X, McPherson DT, Xu J, Urry DW. (1995) Hyper expression of an environmentally friendly synthetic polymer gene. *Biotechnol. Lett.* 17, 745-750.

Hammarstrom M, Hellgren N, Van Den Berg S, Berglund H, Hard T. (2002) Rapid screening for imporved solubility of small human proteins produced as fusion proteins in E. coli. *Protein Sci.* 11, 313-321.

Hanning G, Makrides SC. (1998) Strategies for optimizing heterologous protein expression in Escherichia coli. *Tibtechnology.* 16, 54–60.

Haun RS, Moss J. (1992) Ligation-independent cloning of glutathione fusion genes for expression in Escherichia coli. *Gene.* 112, 37-43.

Hefti MH, Caroline JG, Der Toorn V, Dixon R, Vervoort J. (2001) A novel purification method for histidine-tagged proteins containing a thrombin cleavage site. *Anal Biochem.* 295, 180-185.

Hochuli E, Döbeli H, Schacher A. (1987) New metal chelate adsorbent selective for proteins and peptide containing neighbouring histidine residues. *J Chromatogr.* 411, 177-184.

Hochuli E, Bannwarth W, Döbeli H, Gentz R, Stüber D. (1988) Genetic approach to facilitate purification of recombinant proteins with a novel metal chelate adsorbent. *Bio / Technology.* 6, 1321-1325.

Hopp TP, Pricket KS, Price VL, Libby RT, March CJ, Ceretti DP, Urdal DL, Conlon PJ. (1988) A short polypeptide marker sequence useful for recombinant protein identification and purification. *Bio / Technology.* 6, 1204-1210.

Hosfield T, Lu Q. (1999) Influence of the amino acid residue downstream of (Asp)4Lys on enterokinase cleavage of a fusion protein. *Anal Biochem.* 269, 10-16.

Janknecht R, de Martynoff G, Lou J, Hipskind R, Nordheim A, Stunnenberg HG. (1991) Rapid and efficient purification of native histidine-tagged protein expressed by recombinant vaccina virus. *Proc Natl Acad Sci USA.* 88, 8972-8976.

Janknecht R, Nordheim A. (1992) Affinity purification of histidine tagged proteins transiently produced in HeLa cells. *Gene.* 121, 321–324.

Kane PM, Yamashiro CT, Wolczyk DF, Neff N, Goebl M, Stevens TH. (1990) Protein splicing converts the yeast TFP1 gene product to the 69 kD subunit of the vacuolar H^+-adenosine triphosphatase. *Science.* 250, 651–657.

Kapust RB, Waugh DS. (1999) E. coli MBP is uncommonly effective at promoting the solubility of polypeptides to which it is fused. *Protein Sci.* 8, 1668-1674.

Kaslow DC, Shiloach J. (1994) Production, purification and immunogenicity of a malarian transmission-blocking vaccine candidate: TBV25H expressed in yeast and purified using nickel-NTA agarose. *Bio / Technology.* 12, 494-499.

Kim JY, Mulchandani A, Chen W. (2003) An immunoassay for atrazine using tunable immunosorbent. *Anal. Biochem.* 322, 251-256.

Kim JY, O'Malley S, Mulchandani A, Chen W. (2005) Genetically engineered elastin - protein a fusion as a universal platform forhomogeneous, phase-separation immunoassay. *Anal. Biochem.* 77, 2318-2322.

Knappik A, Plückthun A. (1994) An improved affinity based on FLAG peptide for the detection and purification of recombinant antibody fragments, *Biotech.* 17, 754-761.

Ko YH, Thomas PJ, Delannoy MR, Pedersen PL. (1993) The cystic fibrosis transmembrane conductance regulator. Overexpression, purification, and characterization of wild type and delta F508 mutant forms of the first nucleotide binding fold in fusion with the MBP. *J Biol Chem.* 268, 24330-24338.

Larson AD, Kallio RE. (1954) Purification and properties of bacterial urease. *J Bacteriol.* 68 (1), 67-73.

LaVallie ER, Diblasio EA, Kovacic S, Grant KL, Schendel PF, McCoy JM. (1993) A thioredoxin gene fusion expression system that circumvents inclusion body formation in the E. coli cytoplasm. *Biotechnology.* 11, 187-193.

Li Bin, Darwin Alonso, Daggett Valerie. (2001) The molecular basis for the inverse temperature transition of elstin. *J. Mol. Biol.* 305, 581-592.

Lim DW, Kimberly TC, Mackay A, Chilkoti A. (2007) Improved non-chromatographic purification of a recombinant protein by cationic ELP, *Biomacromolecules.* 8, 1417-1424.

Liu D, Allen RB, Gartner ZJ. (2004) Directed evolution of ligand dependence: Small - molecule - activated protein splicing. *PNAS.* 101, 10505-10510.

Lu Q, Bauer JC, Greener A. (1997) Using Schizosaccharomyces pombe as host for expression and purification of eukaryotic proteins. *Gene.* 200, 135-144.

Lu Z, DiBlasio-Smith EA, Grant KL, Warne NW, LaVallie ER, Collins-Racie LA, Follettie MT, Williamson MJ, McCoy JM. (1996) Histidine patch thioredoxins. Mutant forms of thioredoxin with metal chelating affinity that provide for convenient purifications of thioredoxin fusion proteins. *J. Biol. Chem.* 271, 5059-5065.

Maroux S, Baratti J, Desnuelle P. (1971) Purification and specificity of procine enterokinase. *J Biol Chem.* 246, 5031-5039.

Mehler AH. (1993) Glutathione S-transferases function in detoxification reactions, in ''Textbook of Biochemistry'', *Wiley-Liss*, New York. 523–524.

Meyer DE, Chilkoti A. (1999) Purification of recombinant proteins by fusion with thermally-responsive polypeptides. *Nat Biotechno.* 17, 1112-1115.

Meyer DE, Kimberly T, Chilkoti A. (2001) Protein purification by fusion with an environmentally responsive ELP: effect of polypeptide length on the purification of thioredoxin. *Biotechnol. Prog.* 17, 720-728.

Meyer DE and Chilkoti A. (2003) Quantification of the effects of chain length and concentration on the thermal behavior of ELP. *Biomacromolecules.* 5, 846-851.

Miroux B, Walker JE. (1996) Over-production of proteins in E. coli: mutant hosts that allow synthesis of some membrane proteins and globular proteins at high levels. *J Mol Biol.* 260, 289-298.

Muralidharan V, Muir TW. (2006) Protein ligation: an enabling technology for the biophysical analysis of proteins. *Nat. Methods.* 3 (6), 429–38.

Murphy JT, Lagarias JC. (1997) Purification and characterization of recombinant affinity peptide-tagged oat phytochrom A. *Photochem Photobiol.* 65, 750-758.

Nagai K, Thogersen HC. (1987) Synthesis and sequence specific proteolysis of hybrid proteins produced in Escherichia coli. *Methods Enzymol.* 153, 461-481.

Nagai K, Thogersen HC. (1984) Generation of β-globin by sequence-specific proteolysis of a hybrid protein produced in Escherichia coli. *Nature.* 309, 810-812.

Nilsson S, Lundeberg, Uhlén, Nygren. (1997) Affinity fusion strategies for detection, purification and immobilization of recombinant proteins, *Prot Expr. Purif.* 11, 1-16.

O'Farrell. (1985) Separation Techniques Based on the Opposition of Two Counteracting Forces to Produce a Dynamic Equilibrium. *Science.* 227, 1586-1589.

Parks TD, Leuther KK, Howard ED, Johnston SA, Dougherty WG. (1994) Release of proteins and peptides from fusion proteins using a recombinant plant virus proteinase. *Anal Biochem.* 216, 413-417.

Perler FB, Davis EO, Dean GE, Gimble FS, Jack WE, Neff N, Noren CJ, Thorner J, Belfort M. (1994) Protein splicing elements: inteins and exteins a definition of terms and recommended nomenclature. *Nucleic Acids Res.* 22, 1125-1127.

Pryor KD, Leiting B. (1997) High-level expression of soluble protein in E. coli using His_6-tag and MBP double-affinity system. *Prot Expr Purif.* 10, 309-319.

Rank KB, Mildner AM, Leone JW, Koeplinger KA, Chou KC, Tomasselli AG, Heinrikson RL, Sharma SK. (2001) [W206R]- procaspase 3: an inactivateable substrate for caspase 8. *Protein Expr Purif.* 22, 258-266.

Raymond Stevens. (2000) Design of high-throughput methods of protein production for structural biology. *Structure.* 8, 177-185.

Rodriguez PL, Carrasco L. (1995) Improved factor Xa cleavage of fusion proteins containing maltose binding protein. *Biotechniques.* 238, 241-243.

Rudert F, Visser E, Gradl G, Grandison P, Shemshedini L, Wang Y, Grierson A, Watson J. (1996) pLEF, a novel vector for expression of glutathione S-transferase fusion proteins in mammalian cells. *Gene.* 169, 281-282.

Sachdev D, Chirgwin JM. (1999) Properties of soluble fusions between mammalian aspartic proteinases and bacterial MBP. *J Protein Chem.* 18, 127-136.

Sassenfeld H. (1990) Engineering proteins for purification. *Trends Biotech.* 8, 88–93.

Sassenfeld H, Brewer SJ. (1984) A polypeptide fusion designed for purification of recombinant proteins. *Bio / Technology.* 2, 76–81.

Schmidt M, Tuominen N, Johansson T, Weiss SA, Keinamen K, Oker-Blom C. (1998) Baculovirus - mediated large - scale expression and purification of a polyhistidine - tagged Rubrella virus capsid protein. *Protein Expr Purif.* 12, 323-330.

Schuster M, Wasserbauer E, Einhauer A, Ordner C, Jungbauer A, Hammerschmidt F, Werner G. (2000) Protein expression strategies for identification of novel target proteins. *J Biomol Screen.* 5, 89-97.

Schwarzer D, Cole PA. (2005) Protein semisynthesis and expressed protein ligation: chasing a protein's tail. *Curr Opin Chem Biol.* 9 (6), 561-569.

Skretas G, Wood D. (2005) Regulation of protein activity with small - molecule - controlled inteins. *Protein Sci.* 14, 525-532.

Smith DB, Johnson K. (1988) Single-step purification of polypeptides expressed in Escherichia coli as fusions with glutathione S-transferase. *Gene.* 67, 31-40.

Su XA, Prestwood A, McGraw RA. (1992) Production of recombinant porcine tumor necrosis factor alpha in a novel E. coli expression system. *BioTech.* 13, 756-762.

Terpe K. (2003) Overview of tag protein fusion: from molecular and biochemical fundamentals to commercial systems. *Appl Microbiol Biotechnol.* 60, 523-533.

Tomme P, Boraston A, McLean B, Kormos J, Creagh AL, Sturch K, Gilkes NR, Haynes CA, Warren RA, Kilburn DG. (1998) Characterization and affinity applications of cellulose-binding domains. *J Chromatogr B.* 715, 283-296.

Urry DW, Trapane TL and Prasad KU. (1985) Phase-structure transitions of the elastin poypentapeptide water-system within the framework of composition temperature studies. *Biopolymer.* 12, 2345-2356.

Urry DW. (1988) Entropic elastic processes in protein mechanisms. Elastic structure due to a temperature transition due to internal chain dynamics. *J. Prot. Chem.* 7, 1-34.

Urry DW. (1992) Free energy transduction in polypeptides and proteins based on inverse temperature transitions. *Prog Biophys Mol Biol.* 57, 23-57.

Urry DW, Luan CH, Parker TM. (1991) Differential scanning calorimetry studies of NaCl effect on the inverse temperature transition of some elastin-based polytetrapepides, polypentapeptides and polynonapeptides. *Biopolymer.* 31, 465-475.

Voss S, Skerra A. (1997) Mutagenesis of a flexible loop in streptavidin leads to higher affinity for the Strep-tag II peptide and improved performance in recombinant protein purification. *Protein Eng.* 10, 975-982.

Wood DW, Mahmoud RB, Liang F. (2005) Simple bioseparations using self-cleaving elastin-like polypeptide tags. *Nature Methods*. 2 (9), 659-662.

Xu Z, Bae W, Mulchandani A, Mehra K, Chen W. (2002) Heavy metal removal by novel CBD-EC_{20} sorbents immobilized on cellulose. *Biomacromolecules*. 3, 462-465.

Zhang XK, Wills KN, Hunsmann M, Hermann T, Pfahl M. (1991) Novel pathway for thyroid hormone receptor action through interaction with jun and fos oncogene activities. *Mol Biol.* 11, 6016–6025.

Zhang YB, Howitt J, Mccorkle S, Lawrence P, Springer K, Freimuth P. (2004) Protein aggregation during overexpression limited by peptide extensions with large net negative charge. *Protein Expr. Purif.* 36, 207-216.

Appendices

Appendix I : ELP_{90} Sequences

MHHHHHHSSG LVPRGSGKGP GVGVPGVGVP GGGVPGAGVP

GVGVPGVGVP GVGVPGGGVP GVGVPGGGVP GVGVPGVGVP

GGGVPGAGVP GVGVPGVGVP GVGVPGGGVP GAGVPGGGVP

GVGVPGVGVP GGGVPGAGVP GVGVPGVGVP GVGVPGGGVP

GAGVPGGGVP GVGVPGVGVP GGGVPGAGVP GVGVPGVGVP

GVGVPGGGVP GAGVPGGGVP GVGVPGVGVP GGGVPGAGVP

GVGVPGVGVP GVGVPGGGVP GAGVPGGGVP GVGVPGVGVP

GGGVPGAGVP GVGVPGVGVP GVGVPGGGVP GAGVPGGGVP

GVGVPGVGVP GGGVPGAGVP GVGVPGVGVP GVGVPGGGVP

GAGVPGGGVP GVGVPGVGVP GGGVPGAGVP GVGVPGVGVP

GVGVPGGGVP GAGVPGGGVP GVGVPGVGVP GGGVPGAGVP

GVGVPGVGVP GVGVPGGGVP GAGVPGGGVP GWP

Appendix II : His_6-T_4 Sequences

HHHHHHMILK ILNEIASIGS TKQKQAILEK NKDNELLKRV

YRLTYSRGLQ YYIKKWPKPG IATQSFGMLT LTDMLDFIEF

TLATRKLTGN AAIEELTGYI TDGKKDDVEV LRRVMMRDLE

CGASVSIANK VWPGLIPEQP QMLASSYDEK GINKNIKFPA

FAQLKADGAR CFAEVRGDEL DDVRLLSRAG NEYLGLDLLK

EELIKMTAEA RQIHPEGVLI DGELVYHEQV KKEPEGLDFL

FDAYPENSKA KEFAEVAESR TASNGIANKS LKGTISEKEA

QCMKFQVWDY VPLVEIYSLP AFRLKYDVRF SKLEQMTSGY

DKVILIENQV VNNLDEAKVI YKKYIDQGLE GIILKNIDGL

WENARSKNLY KFKEVIDVDL KIVGIYPHRK DPTKAGGFIL

ESECGKIKVN AGSGLKDKAG VKSHELDRTR IMENQNYYIG

KILECECNGW LKSDGRTDYV KLFLPIAIRL REDKTKANTF

EDVFGDFHEV TGL

Appendix III : His_6-Pfu I Sequences

HHHHHHMILD VDYITEEGKP VIRLFKKENG KFKIEHDRTF

RPYIYALLRD DSKIEEVKKI TGERHGKIVR IVDVEKVEKK

FLGKPITVWK LYLEHPQDVP TIREKVREHP AVVDIFEYDI

PFAKRYLIDK GLIPMEGEEE LKILAFDIET LYHEGEEFGK

GPIIMISYAD ENEAKVITWK NIDLPYVEVV SSEREMIKRF

LRIIREKDPD IIVTYNGDSF DFPYLAKRAE KLGIKLTIGR

DGSEPKMQRI GDMTAVEVKG RIHFDLYHVI TRTINLPTYT

LEAVYEAIFG KPKEKVYADE IAKAWESGEN LERVAKYSME

DAKATYELGK EFLPMEIQLS RLVGQPLWDV SRSSTGNLVE

WFLLRKAYER NEVAPNKPSE EEYQRRLRES YTGGFVKEPE

KGLWENIVYL DFRALYPSII ITHNVSPDTL NLEGCKNYDI

APQVGHKFCK DIPGFIPSLL GHLLEERQKI KTKMKETQDP

IEKILLDYRQ KAIKLLANSF YGYYGYAKAR WYCKECAESV

TAWGRKYIEL VWKELEEKFG FKVLYIDTDG LYATIPGGES

EEIKKKALEF VKYINSKLPG LLELEYEGFY KRGFFVTKKR

YAVIDEEGKV ITRGLEIVRR DWSEIAKETQ ARVLETILKH

GDVEEAVRIV KEVIQKLANY EIPPEKLAIY EQITRPLHEY

KAIGPHVAVA KKLAAKGVKI KPGMVIGYIV LRGDGPISNR

AILAEEYDPK KHKYDAEYYI ENQVLPAVLR ILEGFGYRKE

DLRYQKTRQ GLTSWLNIKK S

Appendix IV : His_6-Taq Sequences

HHHHHHMRGM LPLFEPKGRV LLVDGHHLAY RTFHALKGLT

TSRGEPVQAV YGFAKSLLKA LKEDGDAVIV VFDAKAPSFR

HEAYGGYKAG RAPTPEDFPR QLALIKELVD LLGLARLEVP

GYEADDVLAS LAKKAEKEGY EVRILTADKD LYQLLSDRIH

VLHPEGYLIT PAWLWEKYGL RPDQWADYRA LTGDESDNLP

GVKGIGEKTA RKLLEEWGSL EALLKNLDRL KPAIREKILA

HMDDLKLSWD LAKVRTDLPL EVDFAKRREP DRERLRAFLE

RLEFGSLLHE FGLLESPKAL EEAPWPPPEG AFVGFVLSRK

EPMWADLLAL AAARGGRVHR APEPYKALRD LKEARGLLAK

DLSVLALREG LGLPPGDDPM LLAYLLDPSN TTPEGVARRY

GGEWTEEAGE RAALSERLFA NLWGRLEGEE RLLWLYREVE

RPLSAVLAHM EATGVRLDVA YLRALSLEVA EEIARLEAEV

FRLAGHPFNL NSRDQLERVL FDELGLPAIG KTEKTGKRST

SAAVLEALRE AHPIVEKILQ YRELTKLKST YIDPLPDLIH

PRTGRLHTRF NQTATATGRL CCCDPNLQNI PVRTPLGQRI

RRGFIAEEGW LLVALDYSQI ELRVLAHLSG DENLIRVFQE

GRDIHTETAS WMFGVPREAV DPLMRRAAKT INFGVLYGMS

AHRLSQELAI PYEEAQAFIE RYFQSFPKVR AWIEKTLEEG

RRRGYVETLF GRRRYVPDLE ARVKSVREAA ERMAFNMPVQ

GTAADLMKLA MVKLFPRLEE MGARMLLQVH DELVLEAPKE

RAEAVARLAK EVMEGVYPLA VPLEVEVGIG EDWLSAKE