ABSTRACT

Title of Dissertation: CYCLICAL, CELL-PENETRATING, PEPTIDE-PROTEIN FUSIONS

John Forbes Kerwin
Doctor of Philosophy, 2017

Dissertation directed by: William E. Bentley, PhD.

The expression of recombinant proteins often exploits amino acid motifs that can provide unique properties for recognition, post-translational modification, binding capacity, and translocation potential. Implementing a poly-histidine tag, for example, creates a unique binding site for efficient purification of a target protein from unwanted impurities. This tag can later be cleaved rendering the functional 'native' protein. Inteins are natural, "protein ligators", facilitating the formation of a covalent bond between two flanking protein fragments (exteins) of a translated protein sequence. The intein excises itself upon ligation of the flanking exteins. Split-inteins can flank a target protein and upon an excision event, they enable a post-translational mechanism that cyclizes the protein by cleaving out the flanking intein sequences. Cell-penetrating peptides (CPPs) permit the translocation of exogenous

macromolecules across the cell membrane, while maintaining membrane integrity. Nucleic acids, proteins, and small molecules can be linked to the CPPs through a covalent bond or a non-covalent interaction to facilitate their translocation into the cell. The uptake of a CPP-protein fusion is typically rapid and follows first-order transduction kinetics.

CPP-mediated uptake can be rapidly measured by quenching extracellular fluorescence with trypan blue. Trypan blue will penetrate cells with compromised membranes that could transduce proteins unmediated. In this study, eGFP is expressed alongside the transactivator of transcription (TAT) peptide derived from HIV-1 that can function as a cell-penetrating peptide. The TATeGFP fusion protein is also expressed alongside an artificial split-intein system. The cyclized TATeGFP protein exhibits enhanced *in vitro* stability during protein expression, resistance against exopeptidase digestion, and resistance against chaotropic agent degradation. A cell-penetrating peptide-protein fusion can transduce a cell membrane to deliver an intracellular protein with low cytotoxic effects. A cyclical, cell-penetrating, peptide-protein fusion is a novel tool for fluorometric reporting studies and a backbone for the delivery of intracellular therapeutic proteins with enhanced thermal, enzymatic, and chemical stability.

CYCLICAL, CELL-PENETRATING, PEPTIDE-PROTEIN FUSIONS

by

John Forbes Kerwin

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2017

Advisory Committee:

Professor William E. Bentley, Chair
Professor Panagiotis Dimitrakopoulos
Professor Amy J. Karlsson
Professor Nam Sun Wang
Professor Gregory F. Payne (Dean's Representative)

Dedication

To Dr. Yanto Lunardi-Iskandar—a friend and colleague and the most generous and selfless man I've ever met.

Acknowledgements

My wife, Ashley, has offered relentless support, patience, and love. Thank you

My son, Tyler, and my (future) daughter, Sophia, have offered reason and drive. Thank you

My Dad has taught me to never stop learning. Thank you

My Mom has taught me humility and persistence. Thank you

Bill has offered guidance and intellectual challenge. Thank you

Dr. Zvi Kelman functioned as a co-advisor for the thesis and provided instruction on all things science and fashion. Our chalk talks over coffee and ice cream were instrumental in my experience. Thank you

Table of Contents

List of Tables

List of Figures

List of Abbreviations

293	Human embryonic kidney cells
6-His	Six-histidine tag
AIDA-1	Adhesion-involved-in-diffuse-adherence autotransporter
AIDS	Acquired immune deficiency syndrome
ATCC	American Type Culture Collection
ATP	Adenosine triphosphate
ßME	2-mercaptoethanol
CHMP	Committee for Medicinal Products for Human Use
CHO	Chinese hamster ovary cells
Cint	C_{intein} fragment
CpA	Carboxypeptidase-A
CPP	Cell-penetrating peptide
CpY	Carboxypeptidase-Y
CV	Column Volume
(C)eGFP	Cyclized eGFP mutant
(C)Maternin-eGFP	Cyclized Maternin-eGFP mutant
(C)mCherry	Cyclized mCherry mutant
(C)TATeGFP	Cyclized TATeGFP mutant
(C)TATmCherry	Cyclized TATmCherry mutant
DMD	Duchenne muscular dystrophy
DNA	Deoxyribonucleic acid
DnaE	Genetic product of dnaE gene (enzyme)
DSC	Differential scanning calorimetry
DTT	Dithiothreitol
E. coli	*Escherichia coli*
EMA	European Medicines Agency
FBS	Fetal bovine serum
FDA	Food and Drug Administration
FHV	Flock-house virus cell-penetrating peptide
FLAG	polypeptide protein affinity tag
F/T	Flowthrough
g	Grams
eGFP	Enhanced green fluorescent protein
GST	TAG
GuHCl	Guanidinium chloride
HCl	Hydrochloric acid
HEK	Human Embryonic Kidney cells
HIV	Human immunodeficiency virus
HRP	Horseradish peroxidase
HTB-123	metastatic mammary gland, *Homo sapiens*
IMAC	Immobilized metal ion affinity chromatography
Int	Intein
IPTG	Isopropyl ß-D-1-thiogalactopyranoside

kDa	Kilodalton
KSY-1	Neoplastic AIDS-associated Kaposi's sarcoma, *Homo sapiens*
L	Liter
MA	Maternin
min	Minutes
NF	Naphthofluorescein
Nint	N_{intein} fragment
PBS	Phosphate buffer
PC-3	metastatic prostate, *Homo sapiens*
PE	Plating efficiency
PTH	Parathyroid hormone
Rcf	Relative centrifugal force
RNA	Ribonucleic acid
RPM	Rotations per minute
Scid	Severe combined immune deficiency
SF	Survival fraction
sHPT	Secondary hyperparathyroidism
SICLOPPS	Split-intein circular ligation of peptides and proteins
ßhCG	beta-Human chorionic gonadotropin
Ssp	*Synechocystis* species
St Dev.	Standard deviation
TAT	Transactivator of transcription
TCEP	Tris(2-carboxyethyl)phosphine
TEV	Tobacco Etch Virus
Thr	Thrombin
wt	Wild-type

Chapter 1: Introduction and Literature Review

1.1 Problem Description and Motivation

Therapeutic proteins have revolutionized modern health care. The immense potential to influence diagnostic studies and therapeutic care are limitless. Proteins catalyze *in vitro* and *in vivo* diagnostic studies to advance the treatment for diseases and conditions. Maintaining function while retaining stability is a primary challenge for early drug development studies. Proteins are susceptible to degradation—from enzymatic digestion to unfolding by thermal and chemical inducers. Monitoring protein degradation requires a short-time scale and utilizing fluorometric reporters offer a real-time, on-line platform for rapid measurement and analysis.

Exogenous proteins primary interaction with cells exists only in the outer membrane. The cell membrane an effective barrier of the inner membrane and strictly regulates the transduction of foreign molecules. Exogenous protein transduction can be facilitated by small peptides, penetrating the outer membrane while retaining the native state of the protein. A fluorometric reporter system, capable of cell transduction, engineered with enhanced stability, is a unique platform for therapeutic protein research. This chapter reviews a broad framework of scientific theory and published results that support the experimental method of this dissertation.

1.2 Literature Review

1.2.1 Recombinant Protein Expression in *E. coli*

E. coli is the most prevalent host organism for recombinant protein expression. The highly-characterized organism is widely used for the study of protein transcription, translation, and protein folding. It is an attractive substrate for research platforms based on its rapid doubling time, inexpensive growth media, well-characterized genetics, and the ease to culture at high-cell densities [1].

For protein expression in *E. coli*, heterologous DNA are introduced through a recombinant plasmid. The target DNA sequence is introduced into the plasmid by flanking the sequence with restriction enzymes that allow for complimentary sticky ends to facilitate recombination. The DNA sequence can also be introduced through amplifying oligomer sequences through a polymerase chain reaction (PCR) [2]. The recombinant plasmid often is encoded with an antibiotic-resistance marker. When the growth media is supplemented with the specific antibiotic, plasmid-free cells either die or fail to propagate. Antibiotic resistance genes do cause limitations for expression and target applications. The selective pressure applied by the antibiotic resistance gene will be reduced or eliminated from the degradation or inactivation of soluble antibiotics [3]. Moreover, the desired recombinant protein could become contaminated with the soluble antibiotic. The Food and Drug Administration (FDA) places strict regulations on the usage and acceptable concentrations of antibiotics for protein expression [3-5].

Control of the protein expression can be modulated by utilizing a repressor-protein gene on the recombinant plasmid. The repressor-protein gene, or promoter, is inserted in the plasmid upstream of the target protein. Typically driven by a DNA-binding

protein, the promoter is not transcribed when the DNA-binding protein is present. Small molecules can induce the release of the DNA-binding protein allowing for transcription from the promoter and into the target DNA sequence to proceed. The most prevalent promoter system in *E. coli* is the bacteriophage T7 late promoter. Protein expression is controlled by the T7 RNA polymerase binding to the *lacUV5* promoter [6]. Under normal conditions, the *lac* operon is upstream of the target DNA sequence on the recombinant plasmid. The transcription of the *lac* operon (and thus the target DNA sequence) are inhibited as the *lac* repressor binds to the *lac* operon DNA sequence [7]. The *lac* repressor is a DNA-binding protein that, when bound to the *lac* operon, enhances the affinity of the RNA polymerase for the promoter sequence. This prevents the RNA polymerase in continuing transcription along the plasmid. This mechanism is functional when lactose is not present in the cell media. When present, lactose is metabolized into allolactose, inhibiting the binding ability of the *lac* repressor [7-10]. IPTG can mimic the function of allolactose and is widely use to induce protein expression of heterologous proteins in *E. coli* [11-13].

The recombinant plasmid can encode DNA sequences for genes that encode proteins that can help facilitate and enhance protein expression, purification, and analysis. The additional genes encode a fusion protein that is non-inhibitory on the target protein structure and function. Small poly-peptide sequences can enhance the selectivity of the target protein for affinity chromatography media. The poly-histidine sequence (six, repeated histidine residues) has a strong affinity for immobilized metal affinity resins. This 6-His sequence can be introduced at the 5' or 3' end of the target protein[14, 15]. Green Fluorescent Protein (GFP) is an effective tool for the rapid

visualization and analysis of protein expression level and relative location: expression in the periplasm, cytoplasm, or secreted into the culture media [16]. A FLAG-tag (amino acid sequence, DYKDDDDK) can be introduced as a fusion protein to facilitate immunoprecipitation of the target protein. If the target protein does not have a well-characterized monoclonal antibody, a FLAG-tag can allow for highly-specific immunochemistry reactions for protein analysis studies [17, 18].

Fermentation conditions are optimized for highly-expressing recombinant protein *E.coli* clones. The fermentation is initiated when a single colony is inoculated into a growth media containing the selective-pressure antibiotic coded into the plasmid. This small volume culture is cultured using high agitation (>200 RPM) and at optimal growth temperature, 37.0°C. Upon reaching the mid-logarithmic phase, the culture is used to inoculate a larger volume of culture media to be used for protein expression. The inoculum restarts the growth phase to the lag phase or early-logarithmic phase [19, 20]. Induction of protein expression using IPTG is often initiated during the mid-logarithmic phase. The culture media still has sufficient nutrients to supply the cells during the induction phase [16]. Often, the induction phase is operated at a lower temperature. The lower temperature reduces the effect of proteolysis, and often facilitates better protein folding and the inhibition of formation of inclusion bodies. Inclusion bodies, although enhance the purification efficiency of the protein, are often expensive and unsuccessful in re-folding the target protein [21-23]. After the induction phase, the expressed protein is harvested by centrifugation. The *E. coli* cells are separated from the culture media in a cell pellet. The cells can be mechanically, chemically, or acoustically lysed, releasing the target protein [24]. The cell target

protein can be purified using filtration or chromatography. Chromatography can exploit the fusion proteins for the rapid purification of the protein [16]

E. coli strain DH5-α was designed for the efficient amplification of plasmid DNA. The wild-type K-12 had several genetic modifications, either deletion, mutation, or introduction. The strain has been designed for high transformation efficiency, plasmid stability, and recombination efficiency [25, 26]. *E. coli* has endonucleases in the periplasm that efficiently digest exogenous DNA. The mutation of the endA1 gene enhances plasmid transfer rates by decreasing endonuclease degradation. [27]. Traditional cloning utilizes antibiotic resistant genes on backbone plasmids and sticky or blunt ends of nucleotide sequences for ligation. The mutation of the recA1 gene reduces the efficiency of homologous recombination [25, 26]. The DH5-α strain is commonly used during plasmid construction, modification, and amplification.

E. coli strain BL21 derivatives were designed for efficient recombinant protein production. E.coli strains contain genes that express proteases that digest heterologous proteins [28]. The BL21 strain has deficient genes responsible for expression of outer membrane protease and cytoplasmic protease [29, 30]. The λDE3 lysogen contains a gene that expresses T7 RNA polymerase. This polymerase is induced using IPTG which facilitates the expression of proteins downstream of the T7 promoter [31, 32]. Rosetta BL21(DE3) is a strain from Thermo (Carlsbad, CA) that is traditionally used for recombinant expression of heterologous proteins. It also has the unique ability to express proteins with codons not traditionally used in *E. coli.* Recombinant plasmids are introduced by heat-shock and expressed utilizing the T7 promoter.

1.2.2 Fluorescent Proteins

Luminescence is the emission of light. An electronically excited state from any substance can emit luminescence. Luminescence, existing in two forms, fluorescence and phosphorescence, has become an important tool for life science applications. Phosphorescence is emitted when an electron in an excited orbital is paired with an electron of a ground state orbital of similar spin. Fluorescence is emitted with an electron in an excited orbital is paired with an electron of a ground state orbital of opposite spin. Relatively, fluorescence emission (10^{-8} seconds) occurs at a much more rapid rate than phosphorescence (10^{0}-10^{3} seconds). Therefore, a fluorophore has a lifetime of 10s whereas a phosphor has a lifetime of seconds to minutes [33]. Fluorophore's have a distinct advantage in that many kinetic reactions occur on the same order of a fluorophore lifetime allowing for the distinct measurement using sensitive instruments.

When a fluorophore absorbs light, two specific processes occur: excitation into a higher energy orbital, followed by the return to thermal equilibrium and emission of fluorescent light. The excitation of the fluorophore occurs relatively rapidly, on the order of 10^{-12} second or less, to a higher vibrational state. Since this occurs rapidly, it does not alter the nuclear geometry. When the fluorophore returns to its ground state and produces an emission spectrum, and the transition occurs at a relatively slower rate, on the order of 10^{-8} seconds. The fluorescence emission typically occurs at lower energy and longer wavelengths than the excitation, observed as the Stokes Shift [34, 35].

The Stokes Shift that occurs between the excited state and the emission state occurs due to decay of the fluorophore's ground energy state, solvent-dependent energy transfer, or changes in the molecular structure during excitation [36].

Fluorescent proteins can be defined by their excitation and emission wavelengths or their fluorescence lifetime and quantum yields. The brightest emitting fluorescent proteins emit the highest number of photos relative to the number absorbed. This ratio, the quantum yield, can be calculated using two rate constants: the emissive rate of the fluorophore and the rate at which the fluorophore decays [37, 38]. The same rate constants can be used to calculate the lifetime of the excited state of the fluorophore. The fluorescence lifetime of a fluorophore is measured by the duration of the excited state.

A fluorescent protein's intrinsic ability to produce an emission spectrum can be quenched. Removed by molecular interactions or trivial mechanisms, the quenching of fluorescent proteins is a useful tool in biochemical applications. The interaction of the excited fluorophore with an exogenous molecule often interrupts the process of returning to its ground state and emitting a fluorescent wavelength. Moreover, this interaction could cause a molecular change in the fluorophore preventing the excitation phase completely. Trivial and non-molecular mechanisms, such as buffers or masking agents that absorb emission [39-41].

Fluorophore's have been widely used as research tools in biotechnology, drug discovery, medical diagnostics, and forensics. Cellular mechanisms, binding kinetics, and protein folding can be detected using fluorescent proteins. The most widely utilized protein, Green Fluorescent Protein (GFP) is a 27 kDa protein that occurs

naturally in the jellyfish *Aequorea Victoria* as a major excitation peak at 395nm and a major emission peak at 509nm. GFP is widely used as a gene expression reporter and an easily detectible marker for extracellular and intracellular protein dynamics [42]. In 1961, O. Shimomura began catching jellyfish in nets and mechanically separating the jellyfish's umbrella. After isolating the bioluminescent Aequorin, a calcium-activated fluorophore, Tsien, observed an additional fluorophore, GFP [43, 44]. D.C. Prasher published the nucleotide sequence and the results of the molecular cloning of GFP in 1992, but could not express heterologous GFP that emits fluorescence [43, 45]. M. Chalfie received the non-fluorescent clone from Prasher and identified an inhibitory sequence at the C-terminal end of the cDNA. Removal of the inhibitory sequence created the promoter-controlled expression of GFP in *C.elegans* [43, 46]. R. Y. Tsien also obtained the clone from Prasher and performed the early development studies of the structure-function of GFP. Through random mutagenesis, and mutations identified from biophysical studies, Tsien created several GFP variants with enhanced fluorescence, as well as fluorescent proteins that emit yellow, cyan, and blue fluorescence [43, 47, 48].

GFP has been extensively utilized for studying recombinant protein expression and production. GFP remains a stable biomarker for intracellular fluorescence and protein identification [49]. The tertiary structure of GFP permits functional bioactivity of the fluorophore when expressed alongside a macromolecule. The N- and C-termini of GFP are distant from the fluorophore-inducing beta-barrel structure, creating an efficient linker to a fusion partner. When expressed alongside the fusion partner, the beta-barrel is unaltered and the fluorescence maintained [50].

GFP can be expressed alongside a target protein for the rapid detection and measurement of intra and intermolecular processes. For proteins that transport molecules through and outside of the cell, GFP can be used a fusion partner for the localization of the movement. Four color variants of fluorescent proteins can report the distinct four phases of the cell growth cycle: G1, S1, G2, and M. The cells, Fucci modified cells, can report yellow, green, and red during the latter three cell growth phases [51]. GFP has been fused to reporter and transport proteins to aid the study of how molecules move through cells [52], monitor heterologous expression in real-time [53], and quantitate amplification of recombinant baculovirus particles [54].

The recombinant expression of GFP must undergo a series of first-order modifications to emit fluorescence. This first-order reaction is rate-limited by oxygen concentration and wild-type GFP in A. *Victoria* requires 3.5h for maturation of the fluorescent signal after protein synthesis. When expressed on a recombinant plasmid, GFP persists intracellularly even after promoter activity has been halted [46, 55-57].

1.2.3 Inteins

Inteins function as protein ligation catalysts, facilitating the splicing of two translated peptide sequences into a functional protein. Over 35 different organisms have an intein-based ligation system integrated into their DNA, originating primarily from the Bacteria and Archaea family. These inteins can function as DNA and RNA polymerases, proteases, and enzymes [58]. Intein splicing requires three key components: An N-terminal fragment (N-extein), the intein, and a C-terminal fragment (C-extein). The entire process stems from translation of a single monocistronic mRNA that is comprised of each of these components. In Figure 1A.,

the target of interest is split into a N-extein and a C-extein that flank the functional intein. Based on thermodynamic stability from the N and C-terminal amino acid of the functional intein, the intein excises from the translated protein sequence. The excising of the intein forms a covalent bond between the extein fragments, creating a functional protein [59-65]. The nonspecific excised intein is discarded. The overall efficiency of intein expression and splicing is reduced as homing endonucleases are present in the full sequence [66]. Intein-based protein expression has been leveraged to enhance and streamline protein purification, post-translational modifications, and mini-intein structures that eliminate the homing endonuclease domain and facility cyclic protein expression.

Inteins are considered as "selfish" proteins. Genes present in single-celled bacteria and archaea containing inteins have been replaced with a non-intein based protein from a related organism with no difference in gene function [67]. Since the inteins are integrated in highly-conserved regions, it would be difficult for natural selection to delete and thus face extinction. It has been hypothesized that inteins enhance the selectivity, function, and possibility of deletion for highly-conserved protein motifs [68]. Moreover, if an intein in a highly-conserved region is not expressed, it has been shown the overall function and fitness of the host protein is modulated. Thus, protein expression could be modulated by inteins responding to their environment: pH, buffer, or thiol concentration [69-71]. This phenomenon has been observed in principal in nature, but not demonstrated. The cyanobacteria, *Synechocystis,* has the intein domain split in halves. The N-extein is located on the N-terminus of the DnaE gene, while the C-extein is located at the C-terminus. The

DnaeE intein functions as a post-translational modification *in vitro*. The specific activity of the intein for host protein control has not been demonstrated in its native organism, only replicated as a heterologous protein expression tool [58, 72, 73].

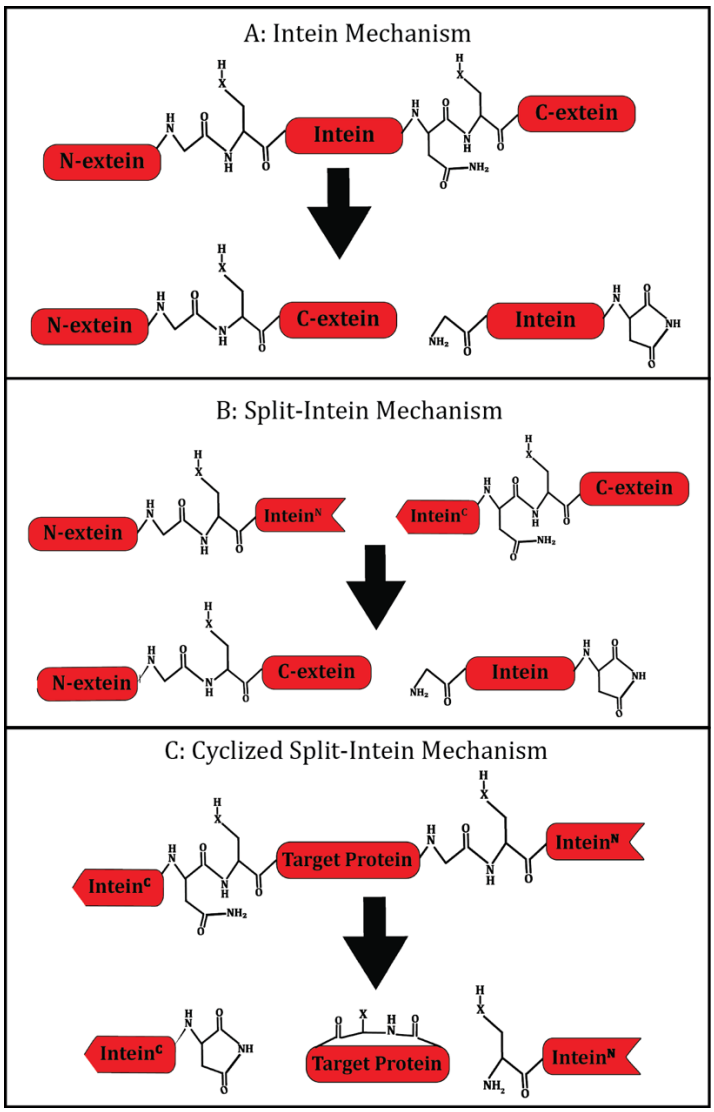


Figure 1. Three mechanism of intein ligation: traditional inteins, split-inteins, and cyclized split-inteins. The traditional intein mechanism is a single translated protein that results in the intein flanked by an N-terminal extein and a C-terminal extein. Upon ligation, the two extein fragments form a covalent bond and release the intein [source]. The split intein mechanism is initiated with two independently translated protein sequences. The intein fragment from the traditional intein mechanism was expressed in two parts. The first half was expressed on the C-terminal end of the N-extein, and the second half was expressed on the N-terminal side of the C-extein. The two halves of the intein form a covalent bond and the traditional intein mechanism proceeds. The cyclized split-intein mechanism is a single translated sequence. The intein halves from the split-intein mechanism are expressed as flanking regions to a target protein. The N-terminal half is expressed on the C-terminal side whereas the C-terminal half is expressed on the N-terminal side. Upon the completion of the processing, the target protein is excised as a cyclized protein and the two intein fragments are released as independent protein fragments.

1.2.3.1 Cyclic peptide synthesis using inteins

The modified intein based splicing mechanism, found in *Synechocystis,* expresses the initial translated protein sequence into two independent motifs: an N-terminal fragment containing the N-extein and half of the Intein, and a C-terminal fragment containing the C-extein and the remaining half of the Intein (*Ssp* DnaE). In this case, the two sequences are translated from two independent expression vectors, and purified separately. The reaction takes place from two purified, translated sequences. In Figure 1B, there are two ligation reactions in this mechanism: first the intein ligates itself forming an active intein-extein complex, then undergoes the similar intein splicing mechanism of Figure 1A producing a ligated protein and an excised intein [71, 74, 75]—Frame B of Figure 1. Finally, in Figure 1C, the split-intein based system developed by Benkovic and coworkers [76-79], creates a circularized target protein—*SICLOPPS Intein*.

Figure 2. The Key amino acids of the translated cyclized, split-intein mechanism (SICLOPPS). The C-terminal amino acid (IC36) of the C-intein is Asparagine, the N-terminal amino acid of the target protein is a cystine or serine (TP01), and the N-terminal

In this mechanism, the target protein is translated between the functional intein that has been split into two parts. The N-terminal intein fragment has been translocated to the C-terminus and the C-terminal intein fragment translocated to the N-terminus—both fragments are translated in reverse. The location of the functional

amino acids that flank the intein in Figure 1A have been inverted. In Frame C of Figure 1, the translated sequence has three segments: a N-terminal intein fragment (C_{intein}), the target protein, and a C-terminal intein fragment (N_{intein}.) The target protein is excised from the two-intein fragments as a circularized protein. The two intein fragments are discarded.

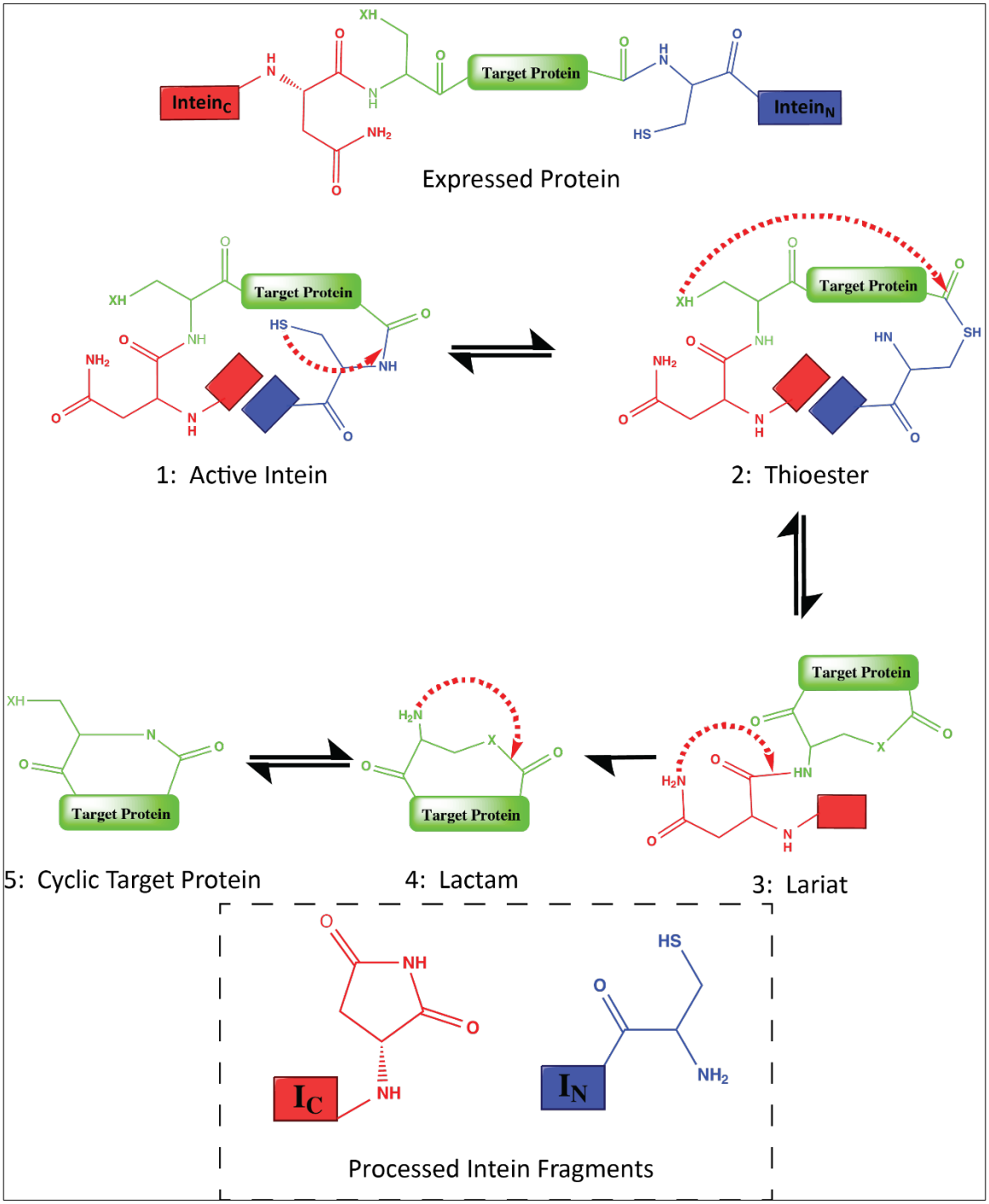


Figure 3. Split-intein based cyclization of a target protein (SICLOPPS). The four major steps that occur during a split-intein derived cyclization (SICLOPPS) of a target protein are: acyl shift, transesterification, asparagine cyclization, and an acyl shift. The key amino acids (Figure 2) that drive the process are the N-terminal amino acid of the Intein^C (Asparagine—IC36), the N-terminal amino acid of the target protein (Cystine or Serine—TP01), and the N-terminal amino acid of the Intein^N (Cystine—IN01)

1. Acyl shift of the N_{intein} to form a Thioester: *The HS side chain of the N-terminal amino acid of the N_{intein} (Cystine) exercises a nucleophilic attack on the C-terminal carboxy group of the target protein. The leaving group (NH) from the N-terminal amino acid of the N_{intein} is cleaved. The I_C-Target Protein-I_N motif is maintained.*
2. Transesterification of Target Protein and N_{intein} to form a Lariat: *The XH side chain of the N-terminal amino acid of the target protein (Cystine/Serine) exercises a nucleophilic attack on the C-terminal carboxy group of the target protein. The leaving group (SH) from the N-terminal amino acid of the N_{intein} is cleaved. This cleaves the N_{intein} from the I_C-Target Protein-I_N motif [73].*
3. C_{intein} Asparagine cyclization fragment releases a Lactone: The C-terminal residue of the *C_{intein} fragment is asparagine. The amine group of the asparagine exercises a nucelophilic attack on the N-terminal carboxyl group of the asparagine. This cleaves the C_{intein} from the target protein forming a lactone (a circular target protein) [80].*
4. Acyl shift in the lactone to form a functional cyclic target protein: *The H_2N side chain of the of the target protein (Lactone) exercises a nucleophilic attack on the C-terminal carboxy group of the target protein. The leaving group (XH) is cleaved forming a functional cyclic target protein [81].*

1.2.3.2 Cyclic peptides as therapeutic models

Cyclic peptides offer a promising template for therapeutic models. Having shown enhanced thermodynamic stability, and function, utilization of a cyclic peptide has shown promising clinical data. Anti-bacteria peptides have been cyclized: gramicidin has been chemically synthesized and expressed alongside the *Ssp* DnaE split-inteins [82], tyrocidine has been chemically synthesized and expressed alongside the *Nostoc punctiforme* PCC73102 (*Npu*) DnaE split inteins[83, 84], and cyclosprorin A has been expressed alongside both the *Npu* and *Ssp* split-inteins [84, 85].

1.2.4 Cell Penetrating Peptides

The outer membrane of a cell controls the transport of material inside the cell. Small molecules—salts, sugars, and amino acids, are transported by membrane channels and signaling proteins. Larger molecules—proteins, nucleic acids, and other large molecules, are transported by endocytosis or exocytosis. The outer membrane contains a fluid, phospholipid bilayer, stabilized by a polar outer surface of and a nonpolar interior. Hydrophobic proteins that prohibits the passage of peptides, proteins, nucleic acids, and other foreign substances from entering the cell. The selectivity of the outer membrane complicates the manner for which therapeutic or diagnostic macromolecules can penetrate the cell membrane. To combat the hydrophobic nature of the cell membrane, physical mechanisms have been used to deliver cargo: microinjection and electroporation. These mechanical methods often induce stress to the cell resulting in programmed cell death or changes in cell physiology. To prevent the mechanical stress that would be imposed by using a physical mechanism, biochemical methods have been employed: liposome or viral-

based vectors. In 1988, the HIV TAT transactivating factor was discovered and determined it could translocate across the cell membrane. This protein, 86 to 101 amino acids in length depending on the sub-type, has a short 11-amino acid motif that encodes the translocation factor of the protein [86]. Since the discovery, numerous natural and synthetic cell-penetrating peptides (CPPs) have been studied to facilitate the transfer of macromolecules inside the cell in a low cytotoxic and non-mechanical transduction pathway. Since its discovery, the TAT peptide has shown low cytotoxicity, low cell-specificity, and dose-dependent cell translocation efficiency with minimal restrictions of the type and size of the delivery cargo.

1.2.4.1 Cell Transduction Methods and Thermodynamics

Macromolecules can cross the plasma membrane of cells through energy-dependent or energy-independent transport. One method of energy-dependent transport of proteins, nucleic acids, or macromolecules is endocytosis [87, 88]. Energy is required to create a surface vesicle after the hydrophilic polypeptide or oligonucleotide has adsorbed to the plasma membrane proteins [89]. Inside the vesicle, the macromolecule is tagged for digestion, degradation, or release into the cytoplasm. Energy-independent transfer across the membrane is not receptor-specific, nor does the hydrophilic molecule adsorb to a membrane protein [90, 91]. This method of macromolecule transduction, direct penetration, leads to the non-cell specific nature of many cell-penetrating peptides [92].

Cell-penetrating peptides are utilized to deliver macromolecules to cells primarily through endocytosis. Nucleic acids, proteins, imaging agents can be covalently or

non-covalently linked as a "cargo" to a short peptide sequence that is non-selective and passive in targeting a delivering the cargo to cells [93, 94]. The cargo can be modified to increase selectivity to a specific class of cells to enhance cell transduction, mediate gene expression, or modulate protein expression and function.

The exact mechanism of CPP transduction has not been confirmed through multiple studies. It is likely that CPPs without cargo can transduce the cell through multiple pathways: direct penetration of the cell membrane or endocytosis [95-97]. When CPPs are coupled to cargo used to detect these pathways, there have been contradictory results claimed—most likely due to the experimental nature of the pathway study. The cellular uptake and transduction efficiency of CPPs and their cargo could change drastically when studied under physiological conditions where highly precise measurement tools, such as Nuclear Magnetic Resonance spectroscopy (NMR), cannot be employed [98, 99]. It is likely additional comprehensive studies are required to better understand the specific transduction methods of specific cell-penetrating peptide molecules for specific cell substrates.

1.2.4.2 Endocytosis

Endocytosis is a cellular process that facilitates the passage of solutes and solvents through the cell membrane. The four different methods of endocytosis, lipid-raft-mediated endocytosis, macropinocytosis, cholesterol-dependent clathrin-mediated endocytosis, or clathrin-independent endocytosis, are energy-dependent mechanisms for permeation of macromolecules across the cell membrane [91, 100-102]. After transport across the membrane through endocytosis, the CPP-cargo complex often ends up in intracellular vesicles and trapped in the cytoplasm. TAT

pathway energy independent. Arginine-rich CPPs, such as TAT, first encounter a series of negatively charged molecules at the cell surface: carbohydrates (mainly glycosaminoglycans), lipids, and membrane proteins. The positively charged CPPs rapidly bind to their negative molecules and facilitate a controlled mechanism for cellular uptake.

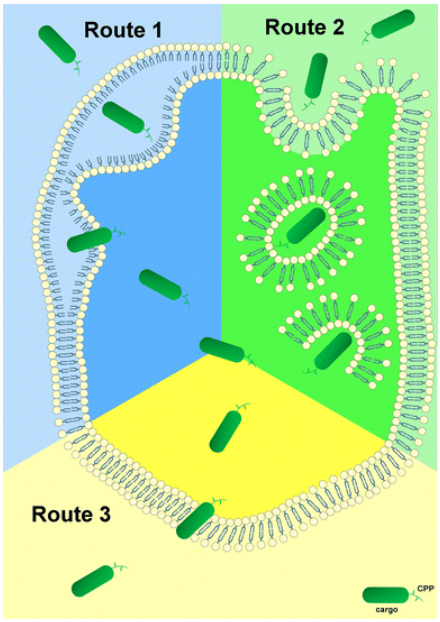


Figure 4. Three proposed methods for cell transduction of cell-penetrating peptides and their cargo: Caveolae/clathrin independent endocytosis, caveolae-mediated endocytosis, and direct penetration. Route 1 is the caveolae/clathrin independent endocytosis method. This method is energy dependent, but does not require a vesicle to transduce through the cell membrane. Route 2 is the proposed primary method for cell transduction through a caveolae-mediated endocytosis. The macromolecule is transduced across the cell membrane and into the cytoplasm using a hydrophilic vesicle. Most macromolecules that are transduced using vesicles will remain trapped inside the vesicle. This method initiates with the adhesion of the CPP-cargo to a receptor-specific cell membrane protein. Route 3 is the direct penetration method that is energy independent.
[Figure from Krautwald]

Endocytosis is temperature and energy dependent

Wedner et. al. published results confirming the TAT-peptide can transduce cells even at 4°C. At this temperature, all mechanisms that drive endocytosis are paralyzed [103].

Richard et. al. published results that showed lower efficiency and kinetics of TAT transduction at 4°C. Moreover, when cells were treated with Sodium Azide or deoxyglucose (agents that consume cellular ATP), the transduction efficiency and kinetics were greatly reduced. Several studies have confirmed the TAT peptide transduces into the cell via numerous methods: endocytosis and direct penetration [104].

1.2.4.3 Caveolae-mediated endocytosis requires the adsorption of the CPP-protein complex to the plasma membrane

Tyagi et. al. published results that confirmed TAT-GFP or TAT-GST cell transduction requires the adsorption to cell surface heparin sulfate proteins. wtCHO-K1 and wtA-745 cell lines, defective of surface heparin-sulfate could not effectively transduce the TAT-protein fusions [105-107].

Wedner et. al. published results where they compared three isoforms of the TAT-peptide sequence: 49RKKRRQRRR57): the native sequence order with native peptide isoforms, the native sequence order with only D-amino acids, and the inverse native sequence order with only D-amino acids (57RRRQRRKKR49). All three isoforms had similar transduction efficiency and kinetics which would refute the requirement of the TAT peptide to adsorb to the plasma membrane prior to the

transduction through the cell membrane. The retro-inverso sequence of the peptide (57RRRQRRKKR49) to its native state (49RKKRRQRRR57) [103].

1.2.4.4 Clathrin-dependent endocytosis requires sufficient potassium and the formation of hydrophobic clathrin vesicles

Richard et. al. treated cells with chlorpromazine (an inhibitor of clathrin vesicles) and transduced cells in the absence of potassium in the culture media. The inhibition of clathrin vesicles and absence of potassium had a significant effect on the transduction of TAT-GFP complexes [104].

Table 1. Confirming and refuting experimental results of endocytosis, caveolae-mediated, and clathrin-dependent endocytosis pathways. All experiments were performed using TAT-GFP complexes and showed experimental results either confirming or refuting the potential transduction pathway.

Transduction Method	Confirming	Refuting
Endocytosis	Inhibition at 4°C and requires ATP [107]	Efficient transduction at 4°C [108]
Caveolae-mediated endocytosis	Heparain adsorption is required [105]	Using all D-amino acids of the native sequence as well as the D-isoform in a reverse sequence functioned equally to the native form [109]
Clathrin-dependent endocytosis	Clathrin vesicles and potassium are required for efficient transduction [110]	Cells expressing anti-Clathrin antibodies could effectively transduce TAT [111]

Ter-Avetisyan G et al. used an ATCC cell line BHK21-tTA/anti-clathrin heavy chain that showed efficient transduction of TAT-GFP complexes. This cell line has an inducible anti-clathrin antibody plasmid. After inducing the cells, the

TAT-GFP complexes were co-incubate to show clathrin-independent endocytosis [111].

1.2.4.5 Endosome Escape

Endocytosis of macromolecules will often result in the compartmentalization of the molecules inside an endosome. The endosome is a low-pH environment that often facilitates proteolytic digestion by lysozyme. The digested macromolecules are then internalized to a specific cellular location or transported outside of the cell [112, 113]. For molecules that are released from the endosome, four proposed mechanisms have been identified:

1. Pore Formation: Proteins or peptides that are trapped in an endosome have the potential to bind to specific areas of the membranous wall. High-affinity binding could lead to disruption in the lipid bilayer, causing a localized pore. This localized pore is a channel for which molecules inside the endosome can freely escape [114, 115].
2. Ion Pump: Proteins or peptides that are localized within the endosome can become protonated in the low-pH environment. If the degree of protonation is excessive, the osmotic pressure within the endosome could reach a critical limit causing rupture in the membranous wall [116].
3. Membrane Fusion: Viral-based peptides or proteins often have fusogenic properties as their primary function. Under physiological conditions (pH~7.4), the fusogenic molecules remain inert and non-functional. However, the low-pH environment of an endosome could induce a fusogenic-effect

where the protein or peptide will fuse to a lipid bilayer molecule in the endosomal membrane. The fusogenic-effect could cause disruption and compromise the integrity of the endosomal membrane leading to leakage [117-120].

4. Synthetic Endosome-Escape Domain (sEED): Hydrophobic small molecules, when fused to proteins or peptides, enhance endosome escape by destabilizing the membrane lipid-bilayer. In a similar mechanism to membrane fusion, sEED molecules have a hydrophobic tail that embeds into the lipid-bilayer causing destabilization and pore formation. If a protein or peptide is fused to the sEED molecule, the pore formation allows for the release from the endosome [121-123].

For efficient delivery of active biomolecules, the endosome escape must be in a non-toxic mechanism. From the above described endosome escape mechanisms, the last two, membrane fusion and sEED appear to disrupt endosome membranes, free of deleterious pore formation or endosome rupture.

The two shortcomings of arginine-rich cell-penetrating peptides is the complete understanding of transduction mechanisms and the efficiency of endosome escape. As described earlier, conflicting publications have provided experimental results to describe the specific mechanism of action for cell transduction: energy dependent or independent methods. Secondly, upon transduction of the outer membrane, the cell-penetrating peptide and associated cargo are often held in an endosomal compartment rendering their function useless and inactive inside the cell. Qian, et al. have discovered that a cyclic cell penetrating peptide (cCPP), increases the efficiency of

endosome escape for arginine-rich peptides that have transduced the cell membrane [124]. Peptides were chemically synthesized as a linear fragment and then a complex reaction, treated with benzotriazole-1-yloxy-tris-pyrrolidino-phosphonium hexafluorophosphate for the covalent linkage of the n-terminal and c-terminal amino acid. The cCPP was labeled with a small fluorophore, naphthofluorescein (NF), to measure endosome escape efficiency and localized diffusion upon entry to the cytoplasm.

The cCPP binds to an outer membrane phospholipid for endocytosis transduction. Energy-dependent endocytosis was confirmed by observing a decrease in transduction at lower temperatures. At 4°C, endocytosis was unmeasurable and completely interrupted. Upon transduction, endosome escape was measured by observing the relative fluoresce intensity of NF over time. NF is a pH-sensitive fluorophore with a pKa of 7.8. Inside the endosome is an acidic environment (pH<6.0) which diminishes the fluorescent signal of NF. Upon endosome escape, the cCPP enters the cytoplasm where the pH is ~7.4. Entering the cytoplasm, the fluorescent signal of the NF-labeled peptide increases confirming release in the endosome.

Endosome escape mechanisms are potentially deleterious to cellular function from pore formation or endosome destruction [116]. Varkouhi, et. al. proposed a mechanism for endosome escape of peptides or proteins where the biological molecules are protonated in the acidic endosome causing an influx of hydrogen and chlorine. The influx of the ions causes osmotic pressure in the endosome and eventual bursting of the compartment. Quan, et al. have provided a conflicting study

where the cell substrate was labeled with an endocytosis marker—Rho-labeled dextran (dextran^{Rho}). Upon endocytosis transduction, the dextran^{Rho} molecule will remain entrapped in the endosome. If co-incubated with a cell-penetrating peptide that causes deleterious endosome escape, the dextran^{Rho} will co-locate in the cytoplasm as escape will be rapid from a porous compartment. Incubation of dextran^{Rho} alongside the NF-labeled cCPP resulted in endosome localization of the dextran^{Rho}-mediated fluorescence, while the NF-labeled cCPP fluorescence was throughout the cytoplasm.

1.2.4.6 Cell penetrating peptide structure-function

The first generation of cell-penetration peptides were truncated amino acid sequences of naturally occurring proteins or chimeric peptides. Naturally occurring active peptide sequences were the first CPPs discovered and often only a few amino acids of a large protein directed translocation. The natural, truncated CPPs could be combined to further enhance their transduction efficiency. Chimeric peptides were often constructed by combining a CPP with a signal peptide introducing cell or substrate selectivity [125] Transportan is a chimeric CPP that is 27 amino acids in length: 12 amino acids at the C-terminus from the signal peptide, Galanin, and 15 amino acids at the N-terminus from the toxic peptide, Mastoparan. Mastoparan, a natural peptide found in wasp venom, directs the translocation of the chimeric peptide. Galanin provides selectivity to neural cell lines [125]. The first class of peptides catalyzed a broad study to optimize the amino acid sequence through

synthetic manipulation to facilitate modified structure and function of the cell-penetrating peptides.

The second generation employed amino acid substitution or addition, or the covalent linking of side chains to the truncated naturally-occurring CPPs. The side-chain modification of CPPs enhanced transduction efficiency or stability, or enhanced the stability of the CPP-cargo complex. Adding hydrophilic groups to an existing CPP would enhance its transduction properties.

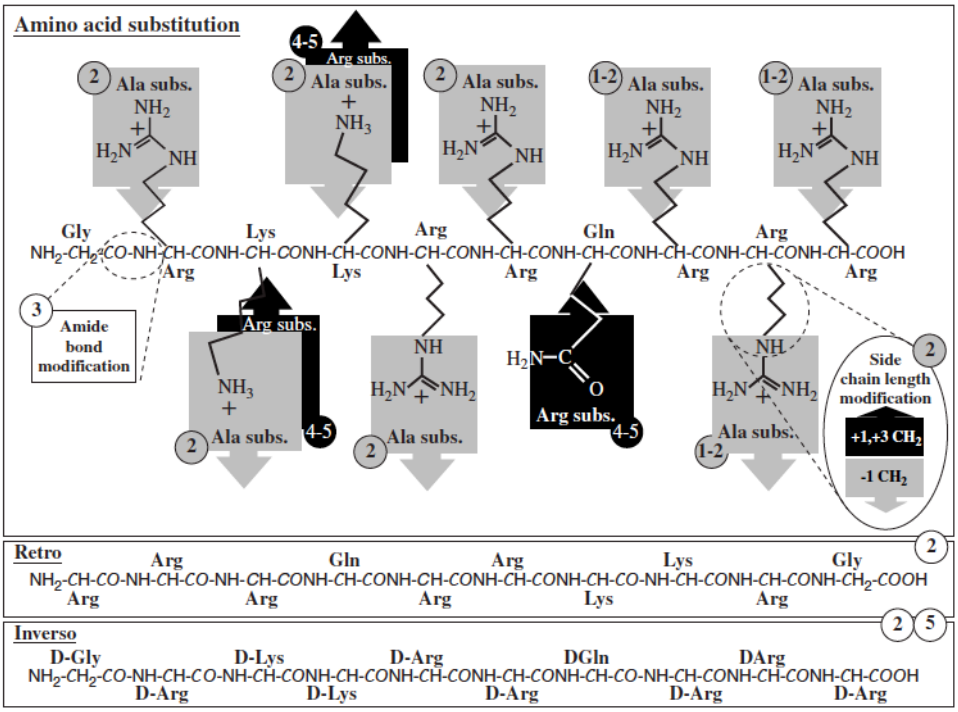


Figure 5. Second generation TAT-peptide (GRKKRRQRRR) derivatives: amino acid substitution, side chain modification, and peptide stereochemistry. [1]Vives *et al.* modified the arginine residues at positions 55,56, and 57 through deletions or substitutions with Alanine. For TAT-GFP complexes, these modifications enhance internal cell fluorescence indicating a greater transduction efficiency[126]. [2]Wender *et al.* also modified the arginine residues at 49,52,53,55,56,57 and the Lysine residues at position 50. Moreover, the CH2-side chain of Arginine 56 was extended. These changes also enhanced the internal cell fluorescence. The use of only D-amino acids, and the inverse amino acid sequence using only D-amino acids did not change the internal cell fluorescence[109]. [3]Umezawa *et al.* synthetically modified the amide bond between the C-terminal Glycine (48) and Arginine 49[127]. [4]Mitchell *et al.* substituted Lysine 50,51 and Glutamine 54 to Arginine and observed enhanced

fluorescence[128]. [5]Futaki used the amino acid substitutions from Mitchell *et al.* and used only D-amino acids to observe similar cell fluorescence[129]. Figure adopted from Vives *et al.* [130].

The third generation are completely synthetic peptide sequences developed by computer models or empirical scientific design and execution. These types of CPPs involve predictive amino acid sequences that are based on structure-activity studies of peptides: most CPPs are comprised of positively charged amino acids, have low hydrophobic interactions with membrane proteins, and are generally 15-25 amino acids in length. Arginine residues contribute to cell transduction potential of CPPs more than lysine groups [125]most likely due to the ability of the extra hydrogen bond donation from the side chain guanidine. Guanidine is a strong base and is rapidly protonated under physiological conditions [86]. Secondly, if the hydrophobic interactions of the cell membrane proteins and the CPP are too high, transduction efficiency is reduced [108, 131]. Finally, the last aspect of a CPP is the short length of the amino acid sequence, as the 13-amino acid TAT peptide is truncated from the 86-101 amino acid backbone [132].

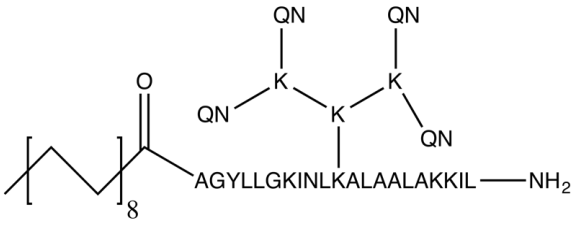
CPP Generation	Name	Origin	Sequence
1st	TAT (48-60)	Truncated human immunodeficiency virus type 1 (HIV-1)	GRKKRRQRRRPPQ
1st	Transportan	Galanin-Lys-mastroparan	AGYLLGKINLKALAALAKKIL
2nd	PepFect 6	Side chain modifications of Transportan	
3rd	Polyarginine	Empirical discovery	Rn, n= 8 or 9

Table 2. Three generations of cell-penetrating peptides. The first class are truncated amino acid sequences from naturally-occurring proteins, the second class are side-chain modifications to enhance the transduction properties of the first class CPPs, and the third class are completely synthetic amino acids sequences that exhibit the properties of naturally-occurring CPPs.

Modeling the kinetics of cell transduction by cell-penetrating peptides and their cargo can utilize a simplified non-reversible first-order rate process:

$$CPP_{bulk} \xrightarrow{k} CPP_{internalized} \quad \text{[Equation 1]}$$

CPP_{bulk} is assumed to be a uniform concentration of the peptide-macromolecule in the bulk phase. The bulk phase-cell membrane is assumed to be free of concentration gradients and macromolecules rapidly diffuse to and from the cell membrane without resistance. Moreover, the bulk concentration is assumed to be significantly higher than the maximum concentration of internalized CPP. To measure the first-order rate constant, samples from the bulk phase can be measured for CPP concentration over a duration of time where an equilibrium has been reached and the bulk concentration and the internalized concentration have reached an asymptotic pattern [96, 133].

Experimental conditions can also deduce the impact of structural changes (or amino acid modifications) of the CPP by altering the bulk concentration, temperature, bulk composition, cell type, or pH of the bulk phase. Different experimental conditions can alter the first order rate constant, k, which can be calculated through the first-order rate kinetics equation:

$$[CPP]_{internalized} = [CPP]_{equilibrium}(1 - e^{-kt}) \quad \text{[Equation 2]}$$

$$t_{0.5internal} = \frac{ln2}{k} \quad \text{[Equation 3]}$$

An additional parameter than can be useful in determining the effectiveness of a CPP-macromolecule or different environmental conditions is the half-time of internalization [$t_{0.5internal}$] [134, 135]. This can be calculated from the rate constant, k, in **Equation 2**. Since the kinetics of internalization are quite rapid, precise and time-sensitive measurements need to be employed to accurately measure the concentration of the CPP-macromolecule either in the bulk phase or internalized.

1.2.5 HIV-transcription activating factor

The HIV-transcription activating factor, or TAT, is a protein responsible for the replication of HIV. Comprising of 86-102 amino acids (depending on the specific viral serotype), three distinct structure-function regions are encoded: the transactivation region, a DNA-binding region, and a nuclear import region. The nuclear import region has the specific ability to transport macromolecules across the cell membrane. It also can deliver calcium-independent binding of proteins to the cell membrane. In its native state, the Tat protein is routinely secreted and transduced back into the cell by endocytosis or direct penetration. After re-internalization, the

protein localizes in the cytoplasm and nucleus [136]. Some studies have shown TAT transduction is time and temperature dependent leading to endocytosis [137]. However, other studies have shown temperature independent transport leading to direct penetration [138].

1.2.5.1 Bioactivity and pre-clinical models

Cell-penetrating peptides are generally believed to have low cytotoxicity; however, a new class of peptides have shown the dual ability to penetrate cell membranes of tumor cells and induce apoptosis. The tumor suppressing protein, p14ARF, has a 22-amino acid motif on the n-terminal region to has shown pro-apoptotic activity. Moreover, the CPP transduces the cell membrane through endocytosis.

Pre-clinical models:

1. Peritoneal carcinomatosis and peritoneal lymphoma in animal models were treated with a TAT-fused peptide, RI-TAT-p53C', that provided lifetime extension as well as a disease-free model [139].
2. Bladder cancer tumors expressing the mutant form of protein, p53, were treated with HA2-fused proteins, T24 and J82, that increased the survival time of animal models [140].
3. Human glioma-initiating cells were effectively transduced with the flock house virus cell-penetrating peptide (FHV) fused to an apoptosis inducing protein, dPasFHV-p53C0. The study concluded the fusion of auto-phagosomes with lysosomes was prevented after transduction of the CPP-protein fusion [141].

1.2.5.2 Clinical studies using cell penetrating peptides for therapeutic delivery

Capstone Therapeutics: AZX100 is a 23-amino acid that rapidly transduces human fibroblast cells and enhances the expression of phosphor-caveolin (Y14). Capstone executed a Phase IIa clinical trial to evaluate the efficacy of AZX100 in preventing scarring after arthroscopic shoulder surgery. The clinical trial was a double-blind, randomized study tested intradermal administration of the peptide or saline at day 9 and/or 21 after surgery. Scar color, size, pain, and morphology (after surgical removal of small tissue samples) were evaluated to determine the effectiveness of AZX100 in comparison to saline. Unfortunately, the product was discontinued as a clinical model after the effectiveness of AZX100 could not be distinguished from the placebo [142].

ReVance Therapeutics: A proprietary peptide sequence inserted at the N-terminal and C-terminal regions of the 150kDa onabotulinumtoxinA protein (botox). From transdermal application (RT001), the flanking peptides permit transduction through multiple layers of epidermis cells as the peptide functions as a cell-penetrating peptide. When delivered subcutaneously (RT002), the flanking peptides rapidly transduce the protein into the epidermis cells and thus control the diffusion of the botox protein to unwanted tissue. ReVance executed a Phase IIIa clinical trial to evaluate the effectiveness of their peptide-protein complex to reduce the appearance of fine lines around the eyes—glabellar lines. The clinical trial was a double-blind, randomized study that unfortunately did not show effectiveness of the peptide-botox complex superior to botox alone [143]. ReVance currently has an active Phase II clinical trial to study the effectiveness of RT002 in preventing cervical dystonia

[144], and an active Phase II clinical trial to study the effectiveness in RT001 in preventing axillary hyperhidrosis [145].

Amgen: Etelcalcetide is an 8-amino acid synthetic peptide that has side chain modifications to enhance the function of the cell-penetrating peptide. A single L-Cysteine is bound through a disulfide bridge to 7 amino acid peptide backbone. The seven amino acids in the backbone chain are all D-isomers, whereas the single Cysteine is an L-isomer [146]. The peptide was studied in a phase 3 clinical trial to treat secondary hyperparathyroidism (sHPT) in adult patients with chronic kidney disease that are currently on dialysis. The peptide activates calcium-sensing receptors which in turn decrease excess secretion of the parathyroid hormone (PTH). Excess PTH in the blood stream leads to an excess concentration of calcium and phosphorus in the bones. The European Medicines Agency (EMA) provided Amgen a positive opinion on behalf of the Committee for Medicinal Products for Human Use (CHMP) in September 2016 [147]. (FDA) rejected the peptide as an approved product in August of 2016 after an extensive Phase III clinical trial[148]. Amgen has submitted a new drug application to the European Medical Agency.

Company	Molecule	Target	Mechanism of Action	Clinical Status
Capstone Therapeutics	AZX100: 23 Amino Acid Peptide	Human Termal Keloid Fibroblasts: Trocar sites of arthroscopic shoulder surgery sites	Cell penetration by endocytosis--lipid rafts[149]	Phase IIa Clinical Trials Terminated
Revance Therapeutics	RT001: peptide-botulin-ium toxin	Treatment of glabbelar lines (crow's feet)	Cell penetration by endocytosis--lipid rafts[150]	Phase IIIa Clinical Trials Terminated
	RT001: peptide-botulin-ium toxin	Axillary Hyperhidrosis	Cell penetration by endocytosis--lipid rafts[150]	Ongoing Phase II clinical studies
	RT002: peptide-botulin-ium toxin	Cervical dystonia	Cell penetration by endocytosis--lipid rafts[150]	Ongoing Phase IIa clinical studies
Amgen	Etelcalc-etide	Calcium-sensing receptor activation	Direct Penetration [146]	Awaiting EMA Approval/ Received NDA Rejection by FDA
KAI Therapeutics	KAI-9803: TAT-PCK(d) Inhibitor	Myocardial ischemia reduction	Cell penetration by endocytosis--lipid rafts[151]	Phase IIa Clinical Trials Terminated

Table 3. Active and terminated clinical studies using cell-penetrating peptides. All clinical studies passed the phase I clinical trial confirming the low cytotoxic mechanism that cell-penetrating peptides transduce into the cell. However, several clinical studies were halted due to efficacy.

1.2.6 Protein Stability and Thermodynamics

Thermal stability of a protein can be theoretically predicted by calculating conformational entropy. The conformational entropy can be estimated based on the limited number of accessible conformations possible by a protein—linear and circular motifs [152, 153]. Using the Gibbs free energy equation and conformational probability (**Equation 4** and **Equation 5**), we can estimate the reduction in conformational entropy;

$$G(p,T) = U + pV - TS \qquad \text{[Equation 4]}$$

$$W_{linear} \approx 6^{2N} \qquad \text{[Equation 5]}$$

$$W_{circular} \approx \sum_{N_x=0}^{N} \sum_{N_y=0}^{N-N_x} \frac{(2N)!}{(N_x!)^2 (N_y!)^2 (N_z!)^2} \qquad \text{[Equation 6]}$$

$$\Delta G_{circular} \approx -RT * ln \frac{W_{circular}}{W_{linear}} \qquad \text{[Equation 7]}$$

Empirically, we can measure the thermal stability of a protein through its enthalpy of transition (**Equation 7**). Using differential scanning calorimetry (DSC), the target protein can be heated beyond a temperature that denatures and unfolds the protein[154-156]. This transition temperature determines the relative thermal stability of a circular motif.

$$\Delta H_{unfolding} = \int_{T_1}^{T_2} C_p dT \qquad \text{[Equation 8]}$$

The transition temperature of the unfolding state can empirically confirm the increased thermal stability of a circular protein in comparison to its linear counterpart.

1.2.7 Proteases

Proteases are naturally occurring enzymes that function to digest native or non-native proteins. In bacteria, proteases can degrade expressed recombinant proteins

decreasing the overall yield. The specific activity of proteases can be divided into their region of protein digestion—endopeptidases digest internal amino acids, whereas exopeptidases digest terminal amino acids [157]. Carboxypeptidase-A (CpA) is an exopeptidase that digests the C-terminal amino acid of a protein. If a protein is expressed as a circular motif, it lacks a C-terminal amino acid, and would be protected from CpA [158].

1.2.7.1 Protease degradation of internalized CPP-protein complexes

Internalized CPPs through endocytosis are often trapped within the endosome or if endosome-escape is facilitated, the macromolecule is susceptible to protease degradation. For CPP-protein complexes that target nuclear localization, protease degradation often causes the accumulation of transduced molecules in the cytoplasm [159].

1.2.8 Present Study

This present study examined the potential of an engineered fluorometric reporter system's cell transduction potential. This novel reporter system combined the background theory and relevant literature research of three isochronic post-translational features: (1) a cyclical protein, (2) a cell-penetrating peptide, and (3) a fluorescent protein.

Chapter 2: A fluorescent, cyclized, protein-peptide fusion of an active subunit of ß-hCG exhibited low *in vitro* cytotoxicity of a PC-3, HTB-123, and KSY-1 clonogenic assay

2.1 Abstract

Maternin, a 14-amino acid subunit of ßhCG, exhibited *in vivo* efficacy of tumor reduction and low toxicity on patients with AIDS-related Kaposi's sarcoma in a phase I/II clinical trial. The active peptide was synthesized and showed comparable colony reduction rates to taxol when treated against PC-3 (prostate), HTB-123 (breast), and KSY-1 (epithelial) cell lines. A peptide, in a complex environment such as the human body, can experience thermal, enzymatic, and chemical degradation. Moreover, the potential for the peptide to fail to reach the target is also a significant hurdle. In this study, Maternin was expressed as an eGFP-fusion protein, utilizing the split-intein expression system. The purified fusion protein was co-incubated with PC-3, HTB-123, and KSY-1 cells at 1.0µg/mL, 0.5µg/mL, and 0.1µg/mL. An eGFP, split-intein fusion, absent of Maternin, and PBS were used as a negative control. Synthetic, cyclized Maternin (mono-disulfide bridge) produced by Aapptec (Louisville, KY) and Taxol, a widely utilized, anti-carcinoma agent was used as a positive control [160-162]. Both control groups were tested at 1.0µg/mL, 0.5µg/mL, and 0.1µg/mL. Survival fraction, to PBS control for 1.0µg/mL, 0.5µg/mL, and 0.1µg/mL respectively: Taxol, 24.7 (±1.81), 16.7 (±2.24), 9.32 (±1.30); Synthetic Maternin, 28.9 (±1.39), 15.6 (±0.98), 9.89 (±1.39), eGFP-Maternin, 87.5 (±2.24), 88.0 (±2.00), 87.6 (±2.66), and eGFP, 87.1 (±3.65), 86.1 (±2.00), 86.9 (±2.60%).

eGFP-Maternin did not exhibit the high cytotoxicity as taxol or synthetic, cyclized Maternin, however, the high survival rates of the cell lines denote the low cytotoxicity of the cyclized eGFP backbone.

2.2 Introduction

KSY-1 is a cell line derived from the pleural fluid of a human patient with AIDS-related Kaposi's sarcoma. This cell line induces tumors, angiogenesis, and metastases when injected into immunocompromised mice (scid). However, when injected into pregnant scid mice, tumor proliferation is inhibited. This led to the discovery of an active sub-unit of the pregnancy hormone, β-hCG—Maternin. Maternin, a 14-amino acid molecule, exhibited efficacy in tumor reduction *in vitro* and *in vivo* [162].

During the discovery phase of Maternin, several preparations of ß-hCG were isolated. These preparations were all small peptide molecules (< 35 amino acids) and exhibited varying effectiveness in tumor suppression. The active preparations also exhibited a positive dose-response correlation. The parent molecule, ß-hCG, did not show *in vitro* efficacy. Figure 6 illustrates the dose-response of Maternin in the phase I/II clinical trial of AIDS-related Kaposi's sarcoma patients. Figure 7 illustrates the *in vitro* efficacy of tumor cell inhibition of several Maternin-related sequences, all subunits of ß-hCG.

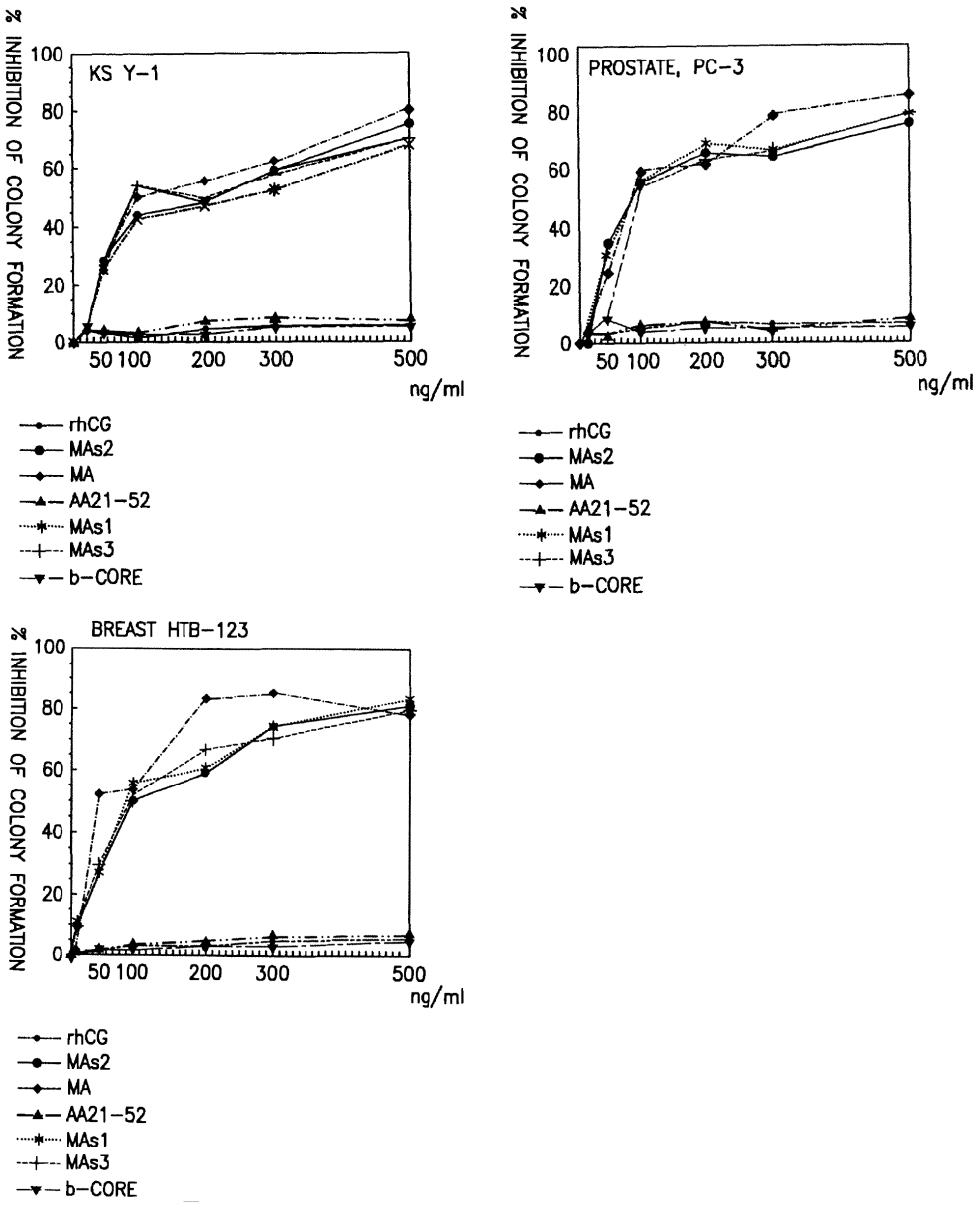


Figure 6. In vitro efficacy of colony forming unit inhibition of KSY-1, PC-3, and HTB-123 using preparations of ß-hCG, from Gill, et.al. (1996)[162]**.** The parental molecule, ß-hCG, tested as a natural and recombinant-expressed molecule, exhibited minimal colony formation inhibition. The five preparations of Maternin, MAs2 (synthetic Maternin, alternative sequence 2), MA (synthetic Maternin, sequence 2), AA21-52 (sub-unit of ß-hCG, amino acid sequence 21-52), MAs1 (synthetic Maternin, alernative sequence 1), and MAs3 (synthetic Maternin, alternative sequence 3) all exhibited high % inhibition of cancer cell line inhibition.

Clinical studies for Maternin were terminated for unknown reasons. It can be speculated that a lack of efficacy or dose-dependency could not be verified in a controlled, clinical-study. A low efficacy in clinical studies could be due to low activity of the molecule, or low bioavailability. A cyclized, peptide-protein fusion could enhance the stability of the peptide and increase bioavailability. In this study, Maternin was expressed as a fusion partner to eGFP, using the intein-based system for a cyclized, peptide-protein fusion. *In vitro* efficacy was studied using a clonogenic assay. Three carcinoma cell lines, used in previous efficacy studies for Maternin, were tested alongside a cyclized, synthetic Maternin and taxol.

VARIABLE	hCG-DOSE GROUP			
	250 IU	500 IU	1000 IU	2000 IU
Responses to therapy				
No. of patients	6	6	6	6
No. with responses*	1	3	4	5
No. of lesions treated with hCG	12	12	12	12
Overall no. of responses†	1	5	5	10
Lesions with biopsy-confirmed complete responses	1	1	2‡	5
No. of lesions treated with diluent	6	6	6	6
Overall no. of responses†	1	0	1	4
Complete responses	1	0	0	4
Size of hCG-treated lesions				
Responding lesions				
No.	1	5	5	10
Median size (mm^2)	66	91	98	53
Size range (mm^2)	—	25–180	49–324	35–112
Nonresponding lesions				
No.	11	7	7	2
Median size (mm^2)	62	84	144	245
Size range (mm^2)	10–1053	50–132	75–400	49–440
P value	0.76	0.78	0.19	0.52

*P = 0.03 for the comparison of dose cohorts by the Cochran–Armitage test for trend.

†These numbers include both complete and partial responses.

‡Both lesions were in the same patient.

Figure 7. Carcinoma lesion size response after treatement with an active sub-unit of ß-hCG. Table from Gill, et al. (1996)[162]. A dose-response can be denoted from the treatement of Maternin on AIDS-related Kaposi's sarcoma patients.

Peptide therapeutics create significant challenges for clinical studies and eventual commercialization. The primary route of administration for peptide drugs is oral delivery. Since peptides are rapidly degraded when entering the stomach by gastric acid and peptidases in the blood, bioavailability is s significant hurdle [163]. Cyclizing peptides has enhanced bioavailability in recent studies [164]. However, inclusion of a fluorometric reporter to a peptide-protein fusion could be a functional tool for pre-clinical discovery and assay development.

2.3 Materials and Methods

To generate a purified protein, *E. coli* codon optimized genes (Appendix 1) were prepared and proteins were expressed in competent cells. The proteins were purified using a two-column platform and then dialyzed into PBS for clonogenic assays. The clonogenic assay is an effective analytical tool for measuring the capacity of a compound to inhibit tumor proliferation in carcinoma cell lines.

2.3.1 Protein Expression

Rosetta (DE3) Competent cells (Novagen/Millipore, Billerica, MA) were transformed with purified plasmids, as described in Appendix A. A single colony was inoculated into LB-media (BD, 30g/L) and incubated overnight at 37°C and 225 rpm (1'' stroke length). The overnight culture was centrifuged at 5000rcf for 10 minutes. The pellet was inoculated at 1% (v/v) into LMR media (Appendix G.1), and incubated at 37.0°C and 225 rpm. At A_{600} ~ 1.0, the culture was rapidly chilled to 16.0°C and induced with 0.5µM IPTG. The induced culture was incubated for 16 hours at 16.0°C and 225rpm and then centrifuged at 5000rcf for 10 minutes. The pellet was lysed in 4mL buffer/g pellet, of IMAC Load/Lyse buffer (Appendix G.2) using a microfluidizer, set at 30,000psi using 3 passes. The lysate was centrifuged at 30,000rcf for 30 minutes. The supernatant was recovered and further purified.

2.3.2 IMAC Purification

Fast flow sepharose (Amersham, Waltham, MA) was loaded onto a gravity column and washed with H20 for 10 column volumes (CV). 5CV of 100mM $NiSO_4$ charged the column as IMAC-Ni^{2+}, and then washed with 20CV of IMAC Load/Lyse buffer.

Lysate (From 3.3.1) was gravity fed onto the column and flow-through was discarded. The column was washed with 10CV of IMAC wash buffer and discarded. The column was eluted using 1CV additions of IMAC elution buffer and further analyzed for target protein concentration using fluorescence. Fractions without target protein were discarded.

2.3.3 Concentration/Dialysis of Protein Solutions (target protein >30kDa)

Dialysis of elution fractions was performed using a 10kDa centrifuge column (Millipore, Billerica, MA). Elution fractions were first concentrated to 10% of the filter capacity by centrifuging at 5,000rcf. 90% of the filter capacity was replaced with 1xPBS, pH 7.4. The filter was centrifuged at 5,000rcf until 10% of the filter capacity was reached. The concentration and dialysis was repeated a total of 3 times. The resulting retentate fraction was recovered and stored at 2-8°C.

2.3.4 Size Exclusion Purification of eGFP-variants

A Superdex-75 10/300GL SEC column (GE, Fairfield, CT) was washed with 2CV of 1xPBS, pH7.4 at 0.5mL/min. 250µL of protein solution was loaded onto the column at 0.5mL/min and further washed with 1xPBS, pH7.4. Flow-through and wash fractions were collected in 250µL fractions. Flow-through was monitored using 280nm and 475nm to monitor total protein and fluorescence, respectively.

2.3.5 Clonogenic Assay

PC-3, HTB-123, and KSY-1 (cell lines from Institute of Human Virology, University of Maryland-Baltimore, Baltimore, MD) were cultured in RPMI (Invitrogen, MD), containing 15% fetal calf serum and incubated at 37.0°C and 5% CO_2. Cells were harvested using 0.4% trypsin and magnetically sorted for cells expressing $CD133^+$ (cancer stem cell marker) and $CD44^+$ (cell-surface glycoprotein) using a LS Column (Miltenyi Biotec, San Diego, CA). 200 viable cells with positive $CD133^+/CD44^+$ markers were plated in a 24-well plate and incubated overnight. The media was aspirated and replenished with fresh media containing the target compound. The plate was incubated for 7-days in a humidified environment. The cells were stained with 0.5% Gentian Violet in methanol and scored as dead if the colony formation was < 25 cells.

2.4 Experimental Design

Clonogenic cell survival assay (clonogenic assay) can determine the effectiveness of a compound's inhibition of cancer cell survival and proliferation. A clonogenic assay's metric is the inability of a single cell to form a colony of > 25 progeny cells. Cells unable to form a colony are considered dead, in contrary to the metric the trypan-blue exclusion method. A typical trypan-blue assay measures the porosity of a cell membrane. A non-viable cell will have a porous membrane permitting trypan blue stain penetration [165, 166]. However, cell can also retain its ability to synthesize DNA and proteins, and even execute one or two mitoses, but unable to form a colony (>25 cells). This cell would not be considered viable under a

trypan-blue assay, but this is considered a dead cell using a clonogenic assay [167-169]. Therefore, a compound could be effective in inhibiting colony formation, yet considered viable. For clonogenic assays measuring a compounds effect on tumor suppression, a clonogenic assay is a more effective metric for effectiveness.

The clonogenic assay in this experiment included two control arms. The positive control arm included a cyclical synthetic Maternin and taxol. The cyclical synthetic Maternin was produced by Aaptec. The cyclical peptide was synthesized using an S-S bond between the N-terminal and C-terminal amino acids. Taxol, a common clinical anti-carcinoma agent was the second positive control group. The negative control arm included PBS (1x, pH 7.4) and (C)eGFP (absent of Maternin). After the 7-day incubation with the compounds the stained plates were analyzed for plating efficiency and colony survival.

To verify the identity of the carcinoma cell lines, the cells were positively selected for two key cell markers: $CD133^+$ and $CD44^+$. $CD133^+$ is a key marker of cancer stem cells. Expression of $CD133^+$ is indicative of a progenitor cell capable of cell division and tumor-initiation capacity [170, 171]. $CD44^+$ is a multi-functional cell surface molecule, involved in cell proliferation, division, and signaling [172] Cancer cells expressing $CD44^+$ indicate the capability of pathogenic activities [173]. A clonogenic assay using cells with positive $CD133^+$ and $CD44^+$ markers, employs an orthogonal selection method for a carcinoma cell line capable of tumor formation.

The compounds tested in a clonogenic assay must be relatively pure and in a physiologically inert matrix. Protein impurities, sourced from the host-cell, that have non-specific binding capacity to purification resins can be minimized by utilizing a

two-column purification strategy. In this experiment, the primary capture column is a Ni^{2+}-charged IMAC column. Proteins containing a polyhistidine-tag will readily bind to the column. The secondary, polishing column, is a size exclusion column (SEC). The SEC column will further purify the target protein from impurities that co-eluted from the IMAC column. This is an essential step for intein-based proteins as the artifacts from the protein ligation (unprocessed inteins, N_{intein} fragment, and C_{intein} fragment) will co-elute with the target, cyclized protein on the Ni^{2+}-charged IMAC column. The SEC column mobile phase is PBS, a physiologically-inert buffer with low cytotoxicity. This allows for the elution fraction of the SEC column to be immediately utilized in the clonogenic assay.

It was hypothesized the (C)Maternin-eGFP compound would function to suppress carcinoma cell proliferation. The clonogenic assay would function to compare the (C)Maternin-eGFP compound to Taxol and cyclized, synthetic Maternin, two compounds that have shown *in vitro* efficacy in tumor suppression using a clonogenic assay. Previous clonogenic assays, testing cyclized synthetic Maternin and taxol have shown 60% colony inhibition (40% survival) at 0.1µg/mL and >80% colony inhibition (<20% survival) at 0.5µg/mL [174]. A cyclized, eGFP-Maternin that inhibited colony formation in carcinoma would provide a therapeutic model to enhance the study and treatment models of cancer cells.

2.5 Results and Discussion

This experiment was executed in two phases: the production and purification of (C)eGFP and (C)Maternin-eGFP recombinant proteins and the execution of a clonogenic assay to measure the effectiveness of the recombinant proteins of inhibiting cancer cell proliferation.

2.5.1 Protein Production and Purification

Recombinant proteins, (C)eGFP and (C)Maternin-eGFP, were expressed as described in 3.3.1 and purified using an IMAC-Ni^{2+} column as described in 3.3.2. The elution fraction from the column was concentrated and dialyzed into PBS as described in 3.3.3 and then purified using a SEC column as described in 3.3.4. Elution fractions in Figure 8 that correspond to the protein gel lane #6 were analyzed for protein concentration using a Bradford Assay (Biorad 5000001) and adjusted to 100µg/mL using PBS. Two additional concentrations were prepared, 50µg/mL, 10µg/mL by diluting the 100µg/mL with PBS.

The post-translational modification of the split-intein system produces several artifact proteins. As observed in Figure 8, a linear eGFP protein can be observed in the elution fraction of the reduced and non-reduced alongside the cyclized eGFP. Since the cyclized eGFP has a reduced conformational entropy as described in Equation 7, the amount of exposed charge groups is less. Since the amount of charged groups on a protein is directly correlated to the residence time in a SDS-PAGE gel, the cyclized protein will migrate faster than the linear protein. In Figure 8, the IMAC column elution fraction is presented as a native (non-denatured, Lane 6)

and a denatured (Lane 5). In both lanes, two proteins can be observed at approximately 27kDa. Both proteins emit fluorescence. The IMAC column provides an enriched fraction of the cyclized GFP protein.

A SEC column enhances the purification of the cyclized eGFP elution fraction from the IMAC column. Since all protein artifacts from the split-intein system will have eGFP, it is necessary to use a non-affinity based purification method to separate the GFP fractions. In Figure 9, the SEC column can enhance a specific fraction for the cyclized eGFP fraction as observed in the corresponding protein gel. The molar ratio of the linear and cyclized protein appears to be uniform in the IMAC elution fraction, however, in SEC fraction 6, there is significantly more cyclized eGFP than linear eGFP. The SEC column is an effective purification tool for purifying an enhanced cyclized eGFP protein solution.

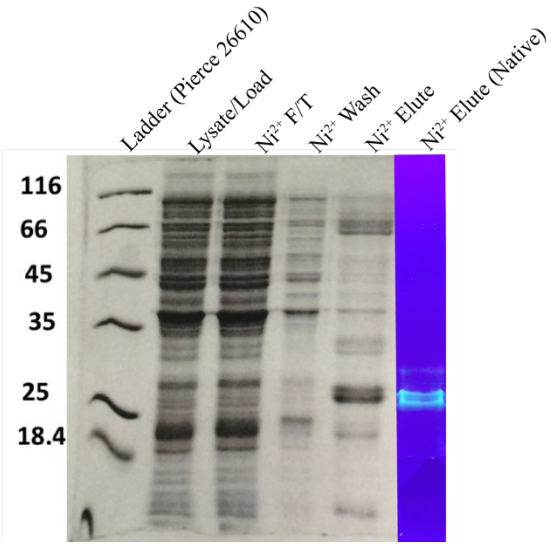


Figure 8. IMAC-Ni^{2+} Purification of (C)eGFP Lysate. The lysate of a (C)eGFP expressed *E. coli* culture was purified on a IMAC-Ni^{2+} and eluted to enhance the cyclized eGFP protein. Lane 5, the column elute, would be furthur purified on a SEC column. The protein gel is a 10-20% Tris-Glycine gradient gel. The sample was mixed equally with 2x Laemmli Buffer (+ßME). Lane 6 is the column elution sample under non-reducing conditions. The sample buffer did not contain ßME and the running buffer of the gel did not contain SDS.

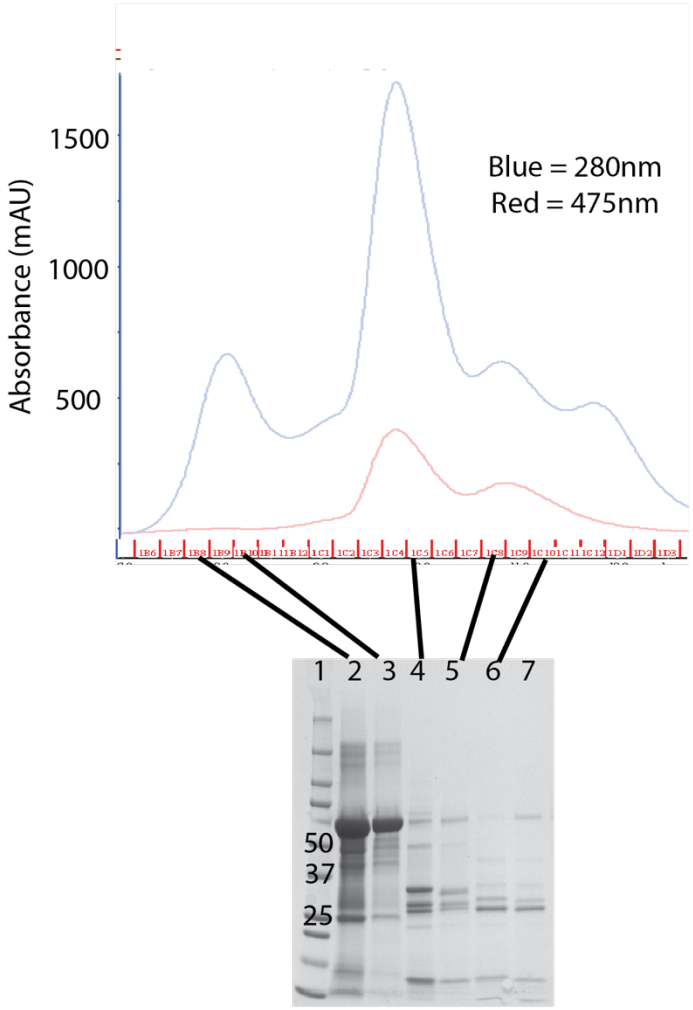


Figure 9. Size-Exclusion purification of (C)eGFP IMAC elution fraction samples. The fractions in the SEC column were widely used in this document. The fraction loaded onto Lane 3 is the unpurified intein fraction. The fraction loaded onto Lane 6 of the gel is the purified cyclic protein sample. The fraction loaded onto Lane 7 is a diluted Lane 3 fraction spike into the Lane 6 fraction. The protein gel is a 4-15% Tris-Glycine gradient gel. The samples are mixed equally with 2x Laemmli buffer (with ßME).

Lane	Sample Description
1	Protein Ladder (Biorad 1610324)
2	SEC Fraction #1
3	SEC Fraction #2
4	SEC Fraction #3
5	SEC Fraction #4
6	SEC Fraction #5—Enhanced cyclized eGFP fraction
7	Diluted SEC Fraction #1 spiked into SEC Fraction #4

Table 4. SEC fraction description of (C)eGFP IMAC elution. The IMAC elution sample was purified on a SEC column. Proteins were purified using an SEC column, smaller proteins migrate faster through the column. Fraction 2 and 3 in the table contain an enhanced C_{intein}-eGFP fraction which is unwanted. Fraction 4 is the desired fraction for a enhanced cyclized eGFP sample.

2.5.2 Clonogenic Assay

The clonogenic assay tested the effectiveness of several compounds on cancer cell proliferation. The experiment's two positive controls: Synthetic (cyclical) Maternin, and Taxol were tested at three different concentrations: 1.0µg/mL, 0.5µg/mL, and 0.1µg/mL. The two positive control arms inhibited colony formation for all three cell lines: KSY-1, PC-3, and HTB-123 under a dose-dependent response. This confirms the assay as reproducible to the published results of Maternin [162].

$$Plating\ effeciency\ (PE) = \frac{total\ colonies\ counted\ \left(>25\frac{cell}{colony}\right)\ [Day\ 7]}{Colonies\ Plated\ \ [Day\ 0]} \qquad \text{Equation 9}$$

$$Survival\ Fraction\ (SF) = \frac{PE_{compound}}{PE_{PBS}} \qquad \text{Equation 10}$$

The purified (C)eGFP did not inhibit carcinoma cell proliferation in the clonogenic assay. It was expected the (C)eGFP would function as a negative control in the assay and produce a survival fraction close to the test control of PBS. The primary function of the (C)eGFP control was to determine the relative cytotoxicity of the Intein backbone for peptide-delivery of the Maternin compound. If the (C)eGFP had induced cytotoxicity during the clonogenic assay, then it would be difficult to ascertain the primary mechanism of action of the carcinoma growth inhibition.

The purified (C)Maternin-eGFP did not inhibit carcinoma cell proliferation in the clonogenic assay. The purified protein had low cytotoxicity, comparable to the (C)eGFP survival fractions. Therefore, based on the positive control of the synthetic Maternin and the negative control of the (C)eGFP, it can be concluded the cyclized eGFP is not cytotoxic to carcinoma cell lines.

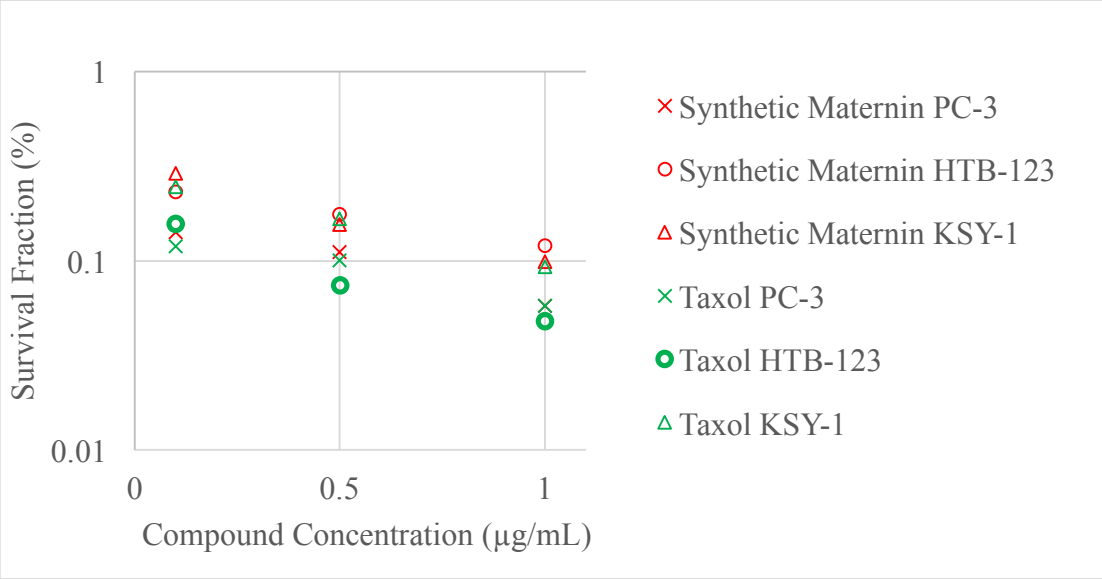


Figure 10. Compound dose-response semi-log plot of Synthetic Maternin and Taxol after 7-days incubation with PC-3, HTB-123, and KSY-1 cell lines. Three concentrations of Maternin and taxol were incubated with PC-3, KTB-123, and KSY-1 for 7 days and the survival fraction was determined.

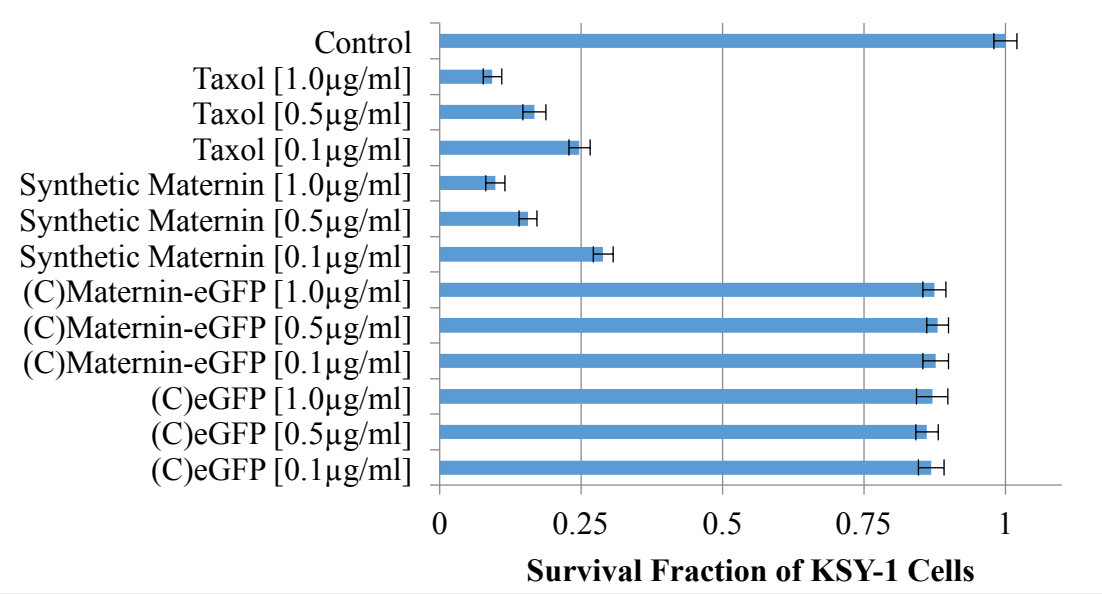


Figure 11. Survival Fraction of KSY-1 Cells after 7-day incubation with 1.0µg/mL, 0.5µg/mL, and 0.1µg/mL of PBS (Control), Taxol, Synthetic Maternin, (C)Maternin-eGFP, and (C)eGFP. KSY-1 cells were incubated with compounds for 7 days and then examined for colony proliferation. Taxol and Synthetic Maternin inhibit colony profileration where as the (C)Maternin-eGFP and (C)eGFP appear to exhibit low cytotoxicity on the cancer cell lines. The results shown are are the average of three replicates with one standard deviation.

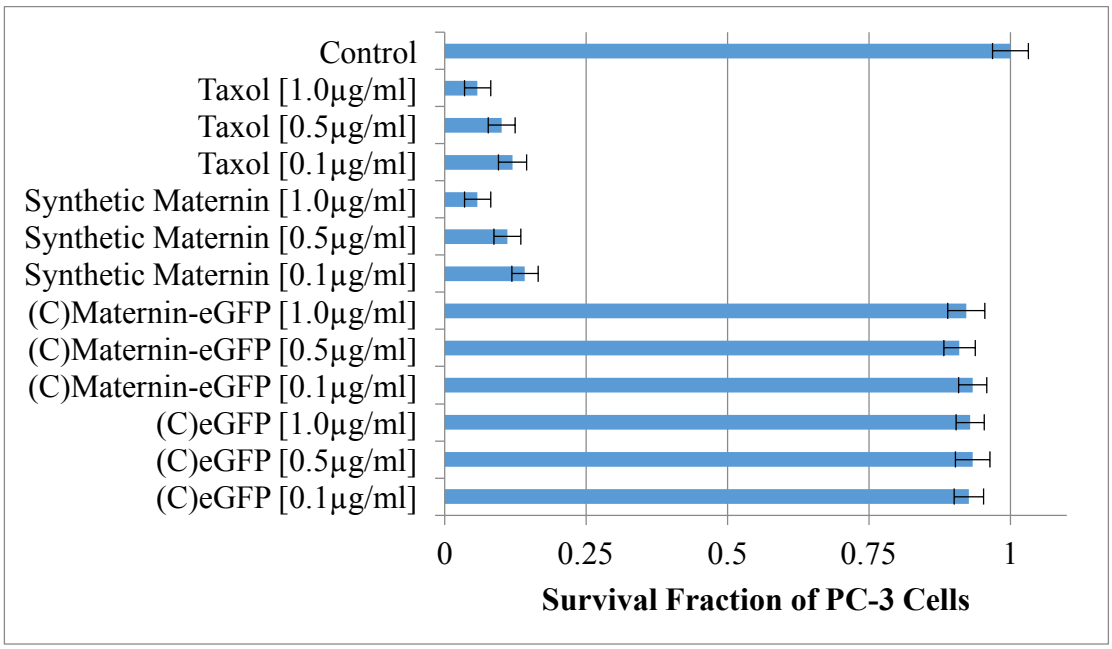


Figure 12. Colony Forming Units of PC-3 Cells after 7-day incubation with 1.0µg/mL, 0.5µg/mL, and 0.1µg/mL of PBS (Control), Taxol, Synthetic Maternin, (C)Maternin-eGFP, and (C)eGFP. PC-3 cells were incubated with compounds for 7 days and then examined for colony proliferation. Taxol and Synthetic Maternin inhibit colony profileration where as the (C)Maternin-eGFP and (C)eGFP appear to exhibit low cytotoxicity on the cancer cell lines. The results shown are are the average of three replicates with one standard deviation.

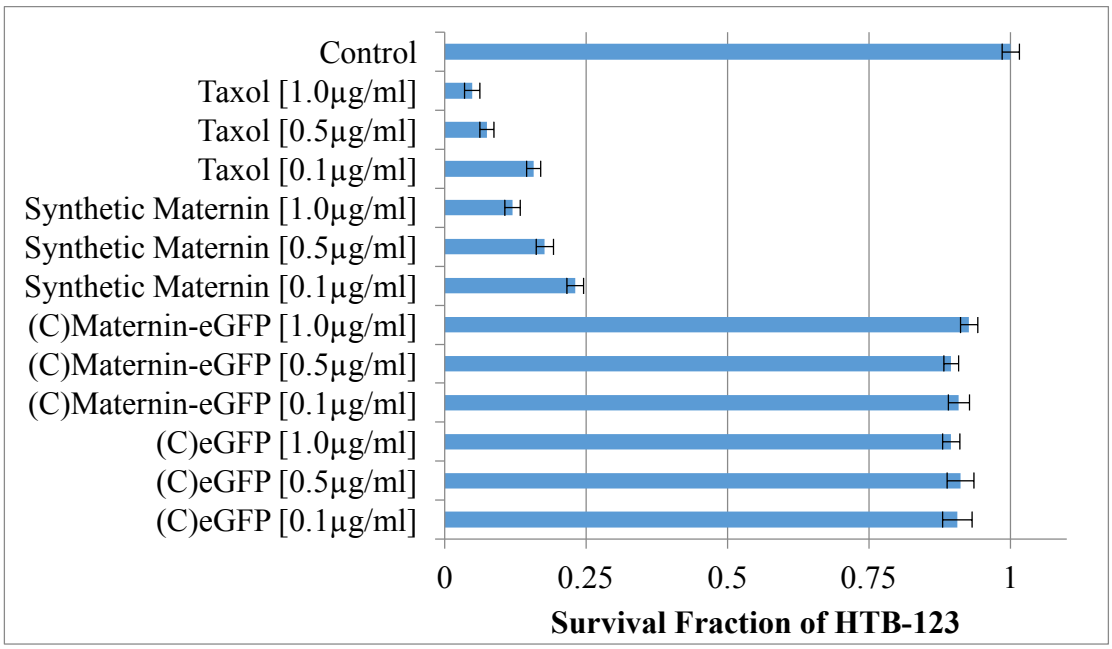


Figure 13. Colony Forming Units of HTB-123 Cells after 7-day incubation with 1.0µg/mL, 0.5µg/mL, and 0.1µg/mL of PBS (Control), Taxol, Synthetic Maternin, (C)Maternin-eGFP, and (C)eGFP. HTB-123 cells were incubated with compounds for 7 days and examined for colony proliferation. Taxol and Synthetic Maternin inhibit colony profileration where as the (C)Maternin-eGFP and (C)eGFP appear to exhibit low cytotoxicity on the cancer cell lines. The results shown are are the average of three replicates with one standard deviation.

2.6 Conclusions

The purified (C)Maternin-eGFP did not inhibit colony formation of carcinoma cell lines in a clonogenic assay. It was expected the cyclized Maternin-eGFP would function similarly to the cyclized synthetic compound. However, two conclusions can be deduced from the experimental results: first, the (C)eGFP backbone has low cytotoxicity to carcinoma cell lines, and second, the Maternin compound requires intracellular delivery for efficient function.

The *in vitro* test of a therapeutic protein is an essential gateway to clinical studies. Execution of an *in vitro* test with the properly designed set of positive and negative controls allows for the most essential conclusions to be drawn from the experiment. In this experiment, the positive controls functioned to inhibit colony formation, and the negative controls did not inhibit colony formation. One of the controls, the (C)eGFP was a successful control. The delivery mechanism of the proposed therapeutic protein was determined to have low cytotoxicity. Essential in determining proper mechanism of action, a scientist must rule out all extraneous conditions and variables to conclude their target protein functions properly. For this experiment, the delivery mechanism of Maternin—cyclized and as a fusion partner to eGFP, did not induce cytotoxicity. Therefore, the delivery mechanism can still function as a fluorometric tool for future clonogenic assay studies.

Maternin has low effectiveness when presented to the extracellular domain of carcinoma cell lines. When presented as a stand-alone peptide, Maternin can freely penetrate the cell membrane of a carcinoma cell. The stand-alone peptide has displayed the ability to inhibit colony formation in the clonogenic assay. However,

when tagged to a larger fusion-protein, the peptide is non-functional. Since eGFP is too large to freely penetrate a cell membrane, a Maternin-eGFP fusion protein will only interact with a carcinoma cell's extracellular domain. For a Maternin-eGFP fusion protein to inhibit colony formation, it must penetrate the cell membrane.

If transduced across the cell membrane, a Maternin-eGFP fusion could offer enhanced potency of the molecule. Peptides are rigid structures that typically do not fold to form tertiary structures. A fusion of Maternin and a larger protein could create artificial tertiary structures in the small peptide. It is possible the new tertiary structures could provide enhanced function.

Chapter 3: Thermal, chemical, and enzymatic stability of a cyclical, cell-penetrating fluorometric reporter system

3.1 Abstract

Forced degradation is a technique employed by biopharmaceutical companies during drug development. Stability of the target protein is an essential characteristic to support clinical modeling and lifecycle predictions for commercialization [175]. The stress conditions tested for a protein stability program can be defined by regulatory guidelines [176] or by the individual company sponsoring a protein therapeutic. The protein stability program is introduced early in the drug development cycle to ensure comparability of efficacy and safety testing in research, manufacturing, and clinical environments [177]. The first stage in drug development is the candidate selection phase. During this phase, a broad stability program is executed on lead candidates to help define studies that will enhance the formulation matrix of the protein. This formulation matrix will provide enhanced stability for the protein when exposed to thermal, chemical, and enzymatic conditions that induce protein degradation [178-180]. Since degradation patterns under physiological conditions could have year-long time scales and half-life degradation patterns, enhanced or extreme conditions are employed to rapidly determine the relative stability of a protein. Although a protein may never experience temperatures exceeding refrigeration conditions (2-8°C), a stability test may incubate the protein at 37.0°C to understand the thermal degradation patterns under accelerated conditions [181]. A second stress-test a stability program could employ would introduce the protein to harsh chemical conditions, such as Guanidine HCl, that function with

chaotropic effects. A broad understanding of the long-term effects of GuHCl could better define a protein stability matrix [182].

Early drug development candidate screening stability studies require assays utilizing reporter systems with short-time scales. In this study, a fluorometric reporter system can analyze accelerated stability studies, under forced degradation conditions, using real-time analysis. A cyclical, cell-penetrating peptide-protein fusion is exposed to four conditions that typically induce protein degradation: *in vitro* conditions during protein expression, chaotropic agent protein degradation, thermal stability under physiological conditions, and exopeptidase digestion. The cyclical protein, (C)TATeGFP, has a eGFP-fusion that functions as the fluorometric reporter. For *in vitro* stability, the (C)TATeGFP protein exhibits enhanced half-life stability during protein expression in relation to the wild-type TATeGFP and eGFP; (C)TATeGFP exhibited enhanced equilibrium time during protein expression in relation to the wild type TATeGFP and eGFP. For GuHCl stability, at 5.0M the (C)TATeGFP had enhanced half-life stability in relation to the wild type TATeGFP and eGFP, at 3.0M, the (C)TATeGFP had enhanced half-life stability in relation to the wild type TATeGFP and eGFP. The (C)TATeGFP offered minimal half-life stability at 1.5M GuHCl in relation to the wild type TATeGFP and eGFP. For thermal stability at 37.0°C, the (C)TATeGFP did not enhance half-life stability in relation to the wild type TATeGFP and eGFP. Finally, the cyclical protein had no observed digestion from carboxypeptidase A. The cyclical, peptide-protein fusion is a fluorometric reporter that exhibits enhanced stability under forced degradation conditions.

3.2 Introduction

This experiment was designed to measure the stability of a cyclized fluorometric reporter to enzymatic, thermal, and chemical degradation. eGFP has been extensively studied as a fluorometric reporter for *in vitro* and *in vivo* assays. Since fluorescence is a non-invasive, on-line measurement, the fluorometric reporter can serve as a correlation to native state folding of eGFP. The initial reading of eGFP fluorometric reporter assays would be considered its most native state. A condition that would affect the stability of the native state or induce cleavage, would theoretically decrease the fluorescence. In the assay, this would be read as a decrease in fluorescence, and measured in real-time under non-invasive conditions.

It is hypothesized, the covalent bond between the N-terminal and C-terminal amino acid in the cyclized eGFP protein will form a more constrained secondary structure in the native protein. A more constrained structure will have less flexibility and decreases the opportunity for susceptible hydrophobic regions from exposure to denaturing conditions. Moreover, the constrained structure will experience a reduced conformational entropy as discussed in Section 1.6. The covalent bond will also remove the opportunistic activity of exopeptidases that target N-terminal and C-terminal amino acids. This study will present a cyclical fluorometric reporter and its linear, wild-type counterpart to forced degradation conditions to observe stability enhancement of the new system.

To monitor the relative stability of eGFP variants during protein expression, cells are monitored for fluorescence in real-time, while maintaining optimal growth conditions. Induced cells can be agitated and incubated in a plate reader capable of

on-line fluorescence to report maturating of a synthesized protein. Since several factors contribute to the relative fluorescence of a culture: promoter strength, cell metabolism and growth kinetics, temperature stability of the native-folded eGFP, and protease activity of the cells. The cell density of the culture can be monitored by A_{600} and native-folded eGFP can be monitored by fluorescence, relative fluorescent units [RFU] (480ex/516em filter.) By combining these two metrics, fluorescence can be analyzed as a function of optical density, $RFU*A_{600}^{-1}$, and the specific fluorescence of individual cells can be measured. Leveau and Lindow published results that analyze the rate of change of the individual cell fluorescence ($\frac{dRFU}{dOD_{600}}$) as the slope of a linear plot of RFU as a function of A_{600}. Their results conclude that a mathematical relationship can be defined that estimates the relative degradation of eGFP during protein expression.

The Gibbs free energy equation, $\boldsymbol{G}\ p,T = U + pV - TS$ [Equation 4], is a function of temperature. Each protein's native state is most favored, when the difference of the denatured and native state free energies is most positive. Since enthalpy and entropy are temperature-dependent functions, incubation of a protein at a temperature above its native state will cause unfolding and eGFP fluorescence degradation. The cyclical proteins are exposed to forced degradation at 37.0°C for a prolonged duration under *in vitro* conditions of protein expression and absent of cells. Halter, et. al. performed a live cell imaging analysis of fibroblasts transfected with eGFP. Protein expression was halted using cyclohexamide. The non-cyclical, eGFP could be modeled by first-order decay kinetics with a fluorescence half-life of 2.8h [183]. Under non-*in vitro* conditions at 37.0°C, the fluorescence half-life has

reported results of 26 hours [184]. This experiment will serve the measure the half-life fluorescence of the cyclical eGFP protein at 37.0°C during protein expression and forced thermal degradation studies.

A chaotropic agent is effective a destabilizing the hydrogen bonding that facilitates a protein's native folding state. Urea and Guanidinium HCl (GuHCl) are two common chemicals that are effective in disrupting the hydrophobic interactions of a protein. Protein stability can be measured by incubation of a fluorescent reporter system with GuHCl. Using real-time monitoring, the kinetic rate of unfolding, modeled by first-order decay kinetics can be measured. Chirico, *et al.*, studied the unfolding of eGFP in various GuHCl concentrations and calculated the first-order rate constant at 5.0M (0.065 min^{-1}), 3.5M (0.0039 min^{-1}), and 1.5M (0.0075 min^{-1}). Moreover, his experiment determined it would take between 50-100 minutes to fully denature eGFP at 2.5M GuHCl [185]. Whereas, Jung *et al.*, expressed cyclized eGFP in a PI-P*fu*I, naturally split-intein, and determined the fluorescence half-life to be ~24 hours at 7M GuHCl with a rate constant of 0.008 min^{-1} at 2.5M GuHCl [186].

Digestion of C-terminal amino acids are a technique often utilized for amino acid sequencing of proteins. Three carboxypeptidases, A, B, and Y, can be utilized based on their specificity for a certain amino acid. Carboxypeptidase-A (**a**romatic/**a**liphatic) have higher activity for aromatic or branched amino acids. Zn^{2+} is a required co-factor and the digestion of arginine (Arg, R), lysine (Lys, K), and proline (Pro, P) is slow and sometimes completely inhibited [187, 188]. Carboxypeptidase-B has specific activity for the digestion of Arg/Lys [189], while carboxypeptidase-Y has universal activity [190]. By analyzing the C-terminal amino

acid sequence of a protein, a specific exopeptidase can be employed to better interpret the degradation patters of enzymatic digestion. The degradation patterns can be measured using mass spectrometry and the amino acid sequence can be predicted. The sequential cleavage activity of Carboxypeptidase-A, coupled with the slow rate of digestion of Arg, Lys, Pro amino acids, create a robust experiment to study the C-terminal digestion of a linear and cyclic protein.

3.3 Materials and Methods

3.3.1 GuHCl Degradation of SEC Column Elution Fractions

Expressed proteins, purified by SEC column chromatography, as described in **2.3.4**, were analyzed for protein concentration using a Bradford Assay and adjusted to 10μM. 10μL of the protein solution was added to a 96-well clear-bottom, black plate. Using the robotic dispensing pumps on the Biotek Neo2 plate reader, 5.0M, 3.0M, or 1.5M GuHCl was added to the protein solution using a volumetric ratio of 6M GuHCl and water, for a total reaction solution volume of 200μL. The eGFP fluorescence was immediately measured using a bottom-read, 485 excitation/516 emission filter.

3.3.2. Carboxypeptidase-A digestion of SEC Column Elution Fractions

The SEC column mobile phase was modified to 25mM Tris, 500mM NaCl, pH 7.5 as described in Section **2.3.4** for optimal activity of the enzyme. Protein solution was mixed 1:100 (w/w) with Carboxypeptidase-A with 1mM $ZnCl_2$ and incubated at 20°C. After 24 hours, a fraction was quenched with 1mM PMSF and stored at 4°C. After 72 hours, a fraction was quenched with 1mM PMSF and stored at 4°C.

3.3.3 MALDI-TOF Analysis of Protein Samples

Samples were desalted using a C-18 ZipTip (EMD Millipore) and eluted using 30% ACN/ultrapure water. 2µL of 10g/L Sinapinic Acid was overlaid on the sample target and vacuum dried. 2µL desalted sample was overlaid on the dried Sinapinic Acid. 2µL of 10g/L Sinapinic Acid was overlaid on the dried sample and vacuum dried. Samples were analyzed using a Bruker Pariflex MALDI-TOF in the linear positive mode at a laser frequency of 60Hz within a mass range of 15 000 – 75 000 kDa and 1000 shots. The spectrophotometer had a PIE delay of 100ns using an Ion source voltage of 20kV/18.3kV, lens voltage 9V, and a linear detection voltage of 2.698 kV, and a sample rate of 1 ns.

3.4 Experimental Design

This study contains two sets of experiments, designed to measure the stability of cyclized (C)TATeGFP under protein denaturing conditions, and to determine the transduction efficiency of the cyclized protein. The first set of experiments were performed to interrogate a protein's stability under chemical, thermal, and enzymatic conditions that would typically induced protein degradation or unfolding. The second experiment measured the relative stability of the cyclized protein during protein expression.

Protein stability during production and purification is critical. Optimal conditions for protein stability are empirically determined and measured against overall process yields. During a clinical or commercial manufacturing process, there are set-point ranges that are determined that allow for operators a timeframe to

execute a critical processing step without the potential for protein loss or yield. Operating temperature of vessels [191], induction and harvest time [192], and pH titration targets require operating ranges [193]. The vessel of a bioreactor will experience temperature fluctuations from temperature gradients existing between the vessel jacket and the culture media. A key process development study for a clinical manufacturing process would be to operate a bioreactor process at varying temperature ranges to determine the optimal temperature range. During empirical studies of protein yield over induction time, a maximum will be observed. However, it is not feasible to ensure the manufacturing process will accurately initiate and complete the induction during this time. Therefore, process development studies will determine an optimal timeframe for induction time to maximize protein yield.

3.4.1 Single cell fluorescence stability during protein expression

Protein expression can be monitored by online fluorescence monitoring using a plate reader capable of maintaining agitation and temperature set-points. Three eGFP constructs were utilized during this experiment: eGFP (**Appendix A.3**), TATeGFP (**Appendix A.8**) and (C)TATeGFP (**Appendix A.9**) were induced with four half-log varying concentrations of IPTG: 0.3, 0.1, 0.03, and 0.01mM in a 96-well plate (n=5). The plate was incubated at 37.0°C and agitated at 205 centimeters per minute (cpm) in a Biotek Neo2 plate reader. Every 10 minutes, fluorescence was measured using a eGFP—filter (485nm excitation, 516nm emission) and optical density (A_{600}). The resulting curve was analyzed to determine the time-dependent stability of eGFP fluorescence. Cells entering the stationary phase of the growth curve still have protein transcription and translation capacity [194, 195]. However,

cells entering the death phase, develop porous membranes that permit the extracellular mobility of expressed proteins—often into buffer matrices that could induce protein instability [196]. Moreover, internal proteases are often deployed by *E. coli* when the growth environment becomes nutrient limited [197, 198]. The fluorescence will be continually monitored as protein degradation is observed, confirmed by a depressed fluorescent signal.

Two competing rate equations model the relative fluorescence of the observed culture, the rate at which eGFP is synthesized and folded into a native state [F] and the rate at which the native state is degraded [F_d]. When Equation 11 is at steady state, the rate of synthesis of native-folded eGFP is in equilibrium with the rate at which native-folded eGFP is degraded to a non-fluorescent structure. These two rate equations can be combined to find the rate of change of fluorescence in Equation 12.

$$\frac{dF}{dt} = k_n F - k_d F_d \qquad \text{Equation 11}$$

k_n ~ native folding state constant for eGFP, 0.45 h [199],
k_d ~rate of native folding degradation

$$\frac{dA_{600}}{dt} = \mu A_{600} \qquad \text{Equation 12}$$

By combining Equation 11 and Equation 12, fluorescence is a function of optical density, and can be analyzed as the relative fluorescence of an *E. coli* cell that has synthesized native-folded eGFP, the rate of change of this ratio can be calculated:

$$\frac{dF}{dA_{600}} = \frac{k_n F - k_d F_d}{\mu} \qquad \text{Equation 13}$$

When the rate at which eGFP is synthesized, and folded into a native state is in equilibrium with the rate at which the native state is degraded, the single cell

fluorescence is at equilibrium, F_{ss}. Therefore, a plot of F as a function of A_{600} will produce a straight line with slope of $\frac{dF}{dA_{600}} = F_{ss}$

Since different heterologous proteins will be synthesized at a specific rate, a comparative metric can be introduced to correlate the specific activity of a cell in synthesizing a native-state eGFP. This is a very simplified model, only modeling the specific fluorescence activity of the cells and the optical density. From this simplified model, it can be correlated to the rate at which the promoter functions to process gene expression from the plasmid sequence. Assume the rate of protein degradation approaches zero (F_d=0), eGFP synthesis (both native and unfolded) is the difference in promoter activity (P) and the rate at which unfolded eGFP matures from the unfolded state (k_nN) and the specific productivity of the cells at synthesizing eGFP (μN):

$$\frac{dN}{dt} = P - k_n N - \mu N \qquad \text{Equation 14}$$

We combine Equation 14 and Equation 11 at steady state for rate of [native-folded] protein synthesis:

$$P = F_{ss}\mu * (1 + \frac{\mu}{k_n}) \quad \text{Equation 15}$$

The rate of protein synthesis can be modulated by the IPTG concentration and modeled by the Hill Equation [200]:

$$P = P_{max} \frac{[IPTG]^h}{[IPTG]^h + K^h} \qquad \text{Equation 16,}$$

$$K \sim [IPTG]_{P=0.5Pmax}$$

In this dissertation, the primary model for protein degradation is first-order exponential decay. **Equation 2** can be modeled for fluorescence degradation in

Equation 17. Leveau, *et al.*, have published one of the only attempts to fully model fluorescence synthesis and degradation using one rate expression, using Michaelis-Menten kinetics. This experiment will show the complexity of fluorescence emission during protein expression and the proposed, simplified model can be improved. Data collected was modeled using first-order kinetics, and to compare data described in [200], a Michaelis-Menten kinetic model of decay will also be calculated to determine the rate of fluorescence degradation [Equation 18]:

$$\frac{dF_d}{dt} = k_d F_d \quad \text{Equation 17}$$

$$\frac{dF_d}{dt} = \frac{F_{dmax}F}{N+F+K_{Fd}} \quad \text{Equation 18,}$$

$$K_{Fd} \sim (N+F)_{Fd=Fdmax}$$

In this experiment, four different concentrations of IPTG were used to express three different eGFP variants: eGFP, TATeGFP and (C)TATeGFP. The relative fluorescence and optical density was measured in real-time and the output was used to calculate the degradation rate of the two linear proteins in comparison to the cyclized protein.

3.4.2 Forced degradation studies using SEC purified proteins

Purified proteins will experience thermal, enzymatic, and chemical degradation conditions during a manufacturing process. The three purified eGFP constructs were interrogated by three conditions that could induce protein degradation. First, proteins were incubated in a chaotropic agent. The destabilization of hydrogen bonding, resulting in the exposure of hydrophobic regions of a protein,

manipulate biophysical properties as the protein changes is conformational shape [201]. The purified proteins were diluted 10-fold (10μL protein to 90μL chaotropic agent) in Guanidinium chloride (GuHCl) over three different final concentrations: 5.0M, 3.0M, and 1.0M. Since the fluorescence depression is rapid, the Biotek plate reader has the capacity for a robotic pump to automatically dispense the GuHCl into a 96-well plate containing the purified protein. The delay in the dispense of GuHCl and the first read of fluorescence (485nm excitation, 516nm emission) is 0.3s. The fluorescence was every 0.2s thereafter.

3.4.3 Exopeptidase digestion of cyclized eGFP

Exopeptidases that digest the C-terminal amino acid of a protein can be monitored using a MALDI-TOF mass spectrometer. However, for cyclized proteins lacking a C-terminal amino acid, a carboxypeptidase would be inactive. After purification of the (C)TATeGFP protein using size exclusion chromatography **(3.3.4),** fractions with an enhanced unprocessed, full translated/transcribed intein sequence will be spiked into the fraction containing an enhanced cyclized TATeGFP protein solution. The gel filtration will be purified at pH 6.5 in 50mM Tris, 150mM NaCl, optimal conditions for Carboxypeptidase-A. After 24, and 72-hours the sample will be analyzed using a MALDI-TOF mass spectrophotometer to observe any molecular weight dispersions. In this study, CpA was utilized based on its specific activity for the six, C-terminal amino acids of the unprocessed intein fragment (six histidine residues) and the two, C-terminal amino acids of the translated EcoR*I* restriction enzyme flanking the C-terminal eGFP and N-terminal N_{intein}, eGFP (E,F). The amino

acid at the N-1 position of these two sequences is proline which should be slowly digested or completely inhibited by CpA. CpA has a molecular weight relatively close to the eGFP-C_{intein} fragment and therefore the full, unprocessed intein was the reference linear protein, susceptible to exopeptidase digestion. Proline is the N-6 amino acid of the unprocessed intein and the N-3 amino acid of eGFP (EcoR*I* is the flanking restriction enzyme of eGFP and N_{intein}). The digestion of proline would be slow or completely inhibited allowing for the molecular weight difference to be observed using MALDI. Using CpY would generate a spectrum of degraded products requiring complex analysis due to its universal amino acid enzymatic activity. The hypothesis of this experiment was the full, unprocessed intein would be digested through the C-terminal 6His tag and the digestion activity would be halted. For (C)eGFP, digestion activity would be completely inhibited. To measure the accuracy of the predicted proteins on the MALDI-TOF spectrum the mass error was calculated:

$$\Delta ppm = 10^6 * \frac{|Mass_{average} - Mass_{observed}|}{Mass_{average}} \quad \text{Equation 19}$$

The average molecular weight of the proteins was calculated by the *NIST Mass and Fragment Calculator*. From an input of amino acid sequence, the program takes the average mass incorporating three m/z ion charge states (1^+, 2^+, 3^+) [202].

3.5 Results and Discussion

This experiment used a series of fluorometric reporters to confirm intein-based protein expression models offer enhanced protein stability. The four reporters studied in this experiment: (C)TATeGFP, (C)eGFP, TATeGFP, and eGFP were

exposed to conditions that typically induce protein degradation, monitored by on-line fluorescence.

3.5.1. *In vitro* fluorescence stability

Induction phase fluorescence emission by eGFP, TATeGFP, and (C)TATeGFP by online and real-time measurement provided empirical models of enhanced *in vitro* stability of cyclized proteins. From an industrial biotechnology perspective, the harvest time is a critical decision. Executing a harvest early results in low product yield, whereas a late harvest could yield proteins with a sub-population that are degraded. *In vitro* models in this experiment were conducted using a plate reader capable of maintaining fermentation conditions. Induced cultures could be agitated, incubated, and fluorescence monitored in real-time.

Online fluorescence monitoring of eGFP expression over time produces a parabolic curve. The three different eGFP variants, expressed using five different IPTG concentrations exhibited differences in the relative fluorescence. In **Error! eference source not found.**, the fluorescence curves indicate a maximum with a relatively stable period of time before and after. This period of stability at the maximum is protein folding equilibrium as the rate of protein expression levels begin to decrease alongside an increase in protein degradation rate. The maximum fluorescence measured for all curves, F_{max} from **Equation 21**, was measured on TATeGFP [0.1mM IPTG] at 1010 minutes, post induction. There are two sub-populations for the fluorescence curves: 0.3mM and 0.1mM IPTG appear to have similar and relatively high expression rates based on fluorescence. Whereas 0.03mM,

0.01mM, and the un-induced culture have lower, but relatively similar expression rates based on fluorescence. It is typical to observe background expression of proteins under a T7 promoter [203, 204]. Expression of the proteins described in this work do not appear toxic to the host *E. coli* strain, therefore the background expression was not investigated. To normalize fluorescence of the three-different native-folded proteins, the relative fluorescence was calculated. This allowed for slight differences of induction optical density and the specific colony chosen for protein expression. The specific activity of the cells synthesizing native-folded eGFP will also be discussed later.

The duration of equilibrium and the rate of protein degradation are two key factors in protein stability during the induction phase. A period of equilibrium will have a zero slope. A protein degradation curve can be modeled by linear or exponential decay models. To accurately calculate the relative fluorescence stability of the constructs used in these experiments, the instantaneous rate of fluorescence change was calculated upon observance of the maximum fluorescence:

$$\Delta Relative\ Fluorescence = \frac{F_{t_2} - F_{t_1}}{t_2 - t_1} \qquad \text{Equation 20}$$

$$\text{Relative Fluorescence} = \mathrm{F}_t = \frac{F_{485ex,516em}}{\mathrm{F}_{\mathrm{max}_{485\mathrm{ex}/516\mathrm{em}}}} \qquad \text{Equation 21}$$

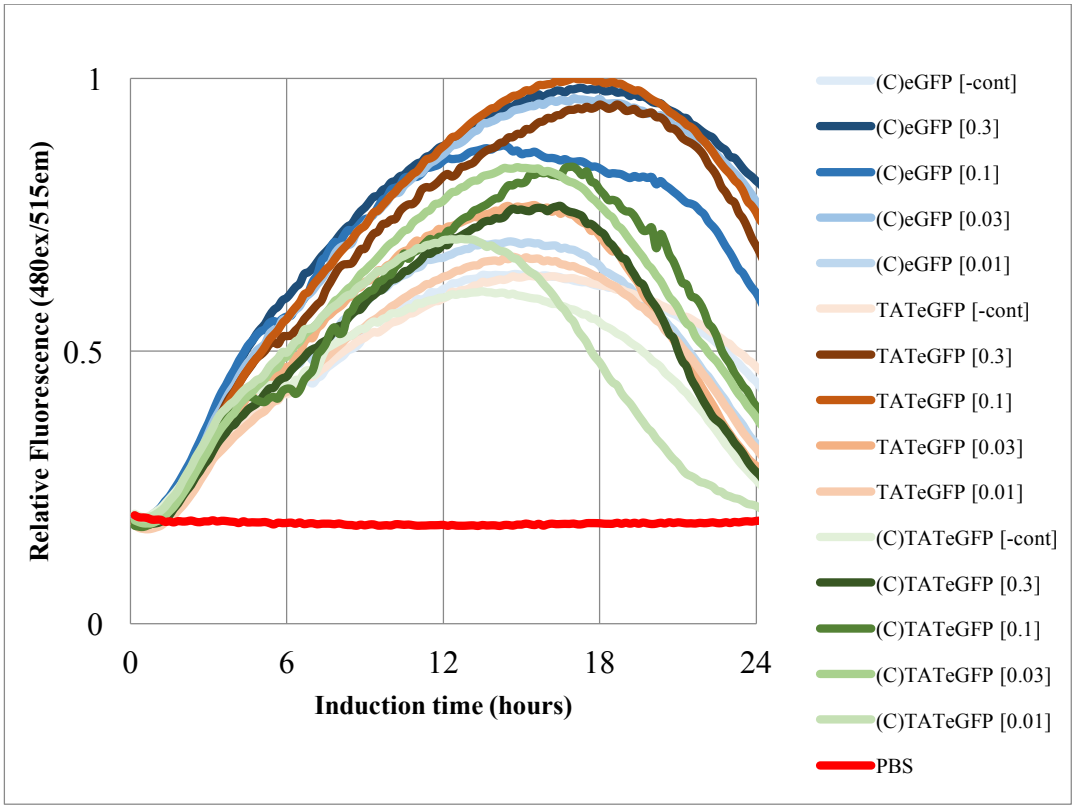


Figure 14. Relative Fluorescence during induction of eGFP, TATeGFP, and (C)TATeGFP using 2-log variance of IPTG over time (hours.) Fluorescence (arbitrary units) was measured over time (hours) for different eGFP protein constructs. Five independent inductions were performed for each protein, with different mM concentrations of IPTG, as listed in the legend with brackets.

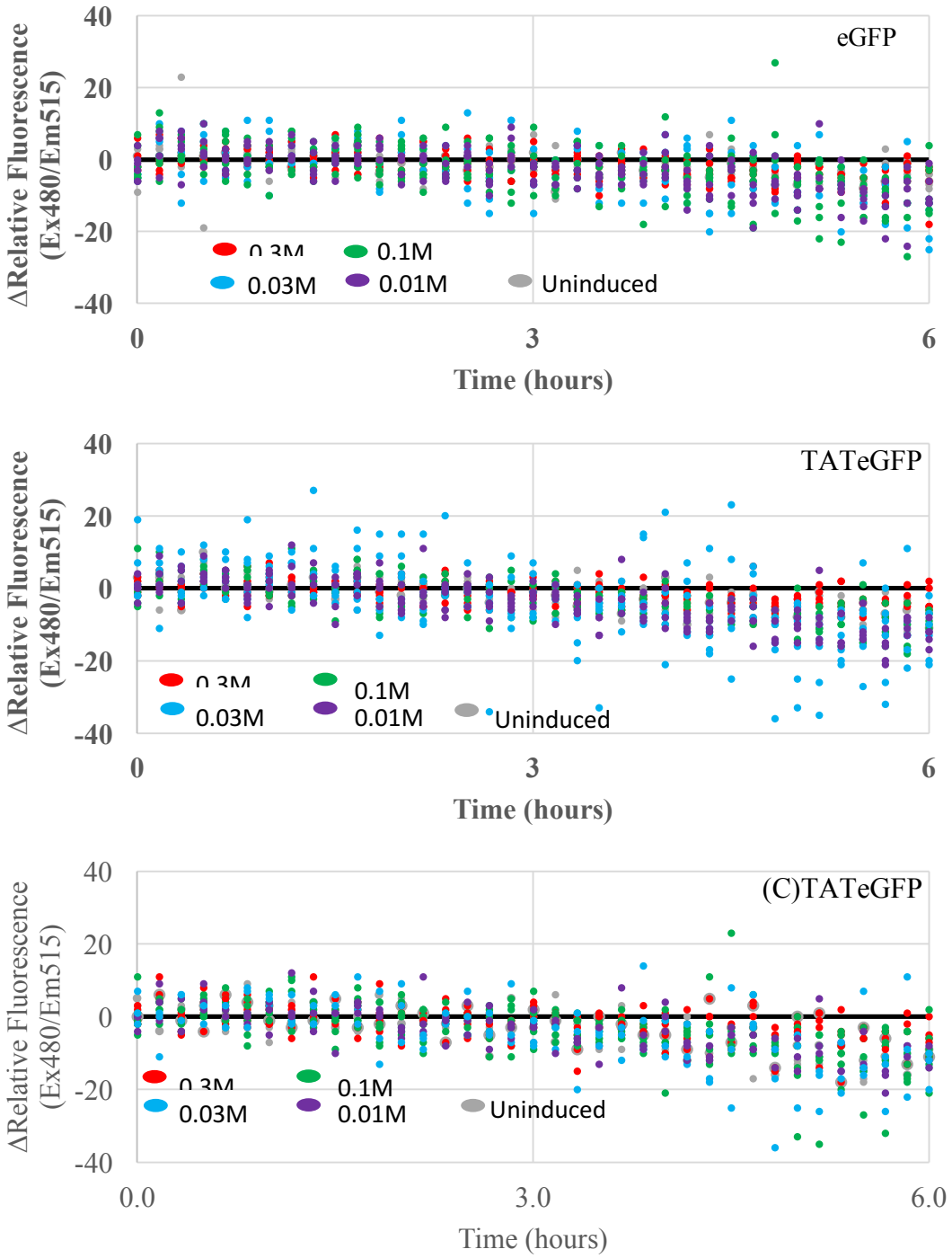


Figure 15: Equilibrium folding and onset of *in vitro* fluorescence degradation during protein expression of eGFP, TATeGFP, and (C)TATeGFP. The instantaneous rate of fluorescence was calculated for eGFP, TATeGFP and (C)TATeGFP over time after a maximum was observed in Figure 14. Five concentrations of IPTG were used to induce the three GFP variants.

3.5.1.1 Fluorescence Equilibrium of protein expression and protein degradation

To further quantitate the duration of fluorescence equilibrium *in vitro* during the induction phase, a quadratic modeling can be utilized. Performing regression analysis on five hours before and after the maxima of fluorescence, a quadratic equation can model the fluorescence curve, Equation 22, and when the derivative of this curve is at a minimum, the fluorescence is at equilibrium, Equation 23:

$$F_t = At^2 + Bt + C \quad \text{[Equation 22]}$$

$$F_t'(|2At + B|) \leq 0.0001 \quad \text{[Equation 23]}$$

$$SS_{res} = \sum_{i=t}(F_i - f_i)^2 \quad \text{[Equation 24]}$$

$$R^2 \equiv 1 - \frac{\sum_{i=t}(F_i - f_i)^2}{\sum_{i=t}(F_i - \overline{F})^2} \quad \text{[Equation 25]}$$

The equilibrium curves for each condition were modeled using regression analysis. The constants in **Equation 22** were calculated by minimizing the residual sum of squares as in **Equation 25**. **Table 5** also references the coefficient of variation to confirm the regression analysis. Then, taking the derivative of the second-order polynomial and finding the minimum, the boundary of equilibrium could be calculated from **Equation 23**. In **Figure 16**, there is an enhancement of the equilibrium time in the (C)TATeGFP curves. The calculated equilibrium time for eGFP and TATeGFP are similar, yet lower than the cyclized TATeGFP expressed using the intein system. Therefore, the cyclized protein was at equilibrium for a longer duration.

[mM] IPTG	A	B	C	R^2
eGFP [0.3]	-9.48E-07	9.69E-04	7.30E-01	0.991
eGFP [0.1]	-8.64E-07	8.61E-04	6.57E-01	0.967
eGFP [0.03]	-1.03E-06	1.08E-03	6.81E-01	0.992
eGFP [0.01]	-7.45E-07	7.43E-04	4.58E-01	0.982
eGFP [-cont]	-1.14E-06	1.25E-03	5.94E-01	0.992
TATeGFP [0.3]	-1.23E-06	1.31E-03	6.46E-01	0.936
TATeGFP [0.1]	-1.31E-06	1.39E-03	3.88E-01	0.986
TATeGFP [0.03]	-1.00E-06	1.05E-03	3.92E-01	0.968
TATeGFP [0.01]	-6.64E-07	6.75E-04	4.63E-01	0.988
TATeGFP [-cont]	-1.45E-06	1.45E-03	3.90E-01	0.976
(C)TATeGFP [0.3]	-1.45E-06	1.52E-03	4.13E-01	0.941
(C)TATeGFP [0.1]	-1.47E-06	1.54E-03	4.19E-01	0.957
(C)TATeGFP [0.03]	-1.38E-06	1.46E-03	3.01E-01	0.977
(C)TATeGFP [0.01]	-7.74E-07	8.38E-04	3.78E-01	0.964
(C)TATeGFP [-cont]	2.47E-08	-3.59E-05	1.94E-01	0.976
PBS	2.47E-08	-3.59E-05	1.94E-01	0.868

Table 5: Quadratic approximation constants of fluorescence emission of eGFP, TATeGFP, and (C)TATeGFP within 5 hours of fluorescence maximum. Four different concentrations of IPTG [mM] were executed alongside a [-cont] that was not induced with IPTG. The quadratic constants from **Equation 22** were analyzed for variance using the coeficient of variation,**Equation 24.**

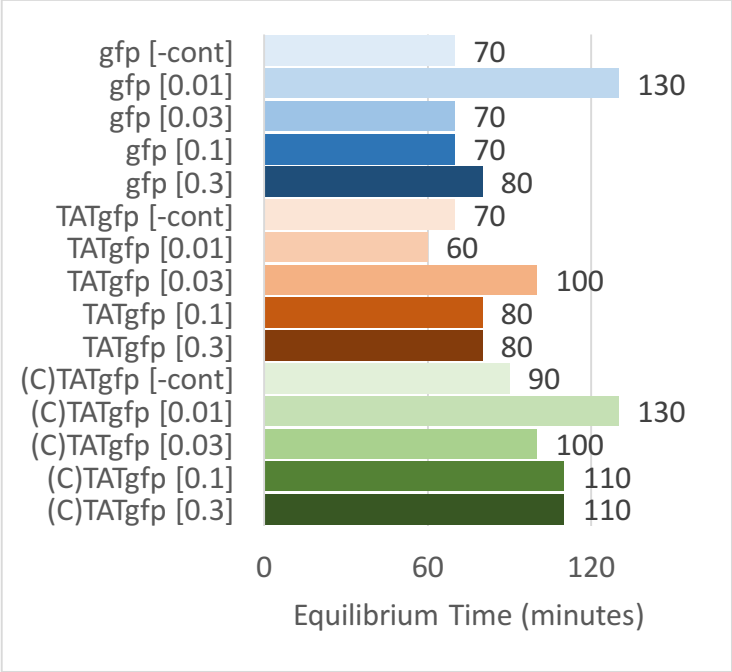


Figure 16: *In vitro* fluorescence equilibrium of eGFP, TATeGFP, and (C)TATeGFP. The fluorescence equilibrium time, from **Equation 23,** can quantify the relative stability of the expressed protein, *in vitro,* during protein expression.

3.5.1.2 Modeling protein *in vitro* protein degradation using Michaelis-Menten kinetics

The calculated specific growth rate of each *E. coli* variant harboring a different genetic construct should be relatively equivalent. This hypothesis is confirmed in **Table 6** as all specific growth rates are within two standard deviations. Then the RFU as a function of A_{600} begins to show relative differences of each eGFP variant. The rate of native-folded fluorescence is modeled in **Figure 17**. Since eGFP is the fluorometric reporter, it is expected that using eGFP as a fusion protein could alter the amount of time the protein is synthesized and then folds into the native state. This is readily observed by the Hill Equation plot in **Figure 18.** Expression of the eGFP protein alongside the TAT peptide and the intein-based expression system reduces the rate of synthesis of native-folded eGFP.

From **Figure 17**, three distinct phases can be observed during protein expression: first, there is a rapid increase of fluorescence as the cells remain in an active state. After the fluorescence reaches a maximum, the rate of fluorescence degradation is slow. The optical density remains unchanged whereas the fluorescence decreases linearly. When the cells enter the death phase, marked by a decrease in optical density, the fluorescence rate rapidly increases. The cells internal proteases become active and digest the native-folded eGFP at a faster rate than the cells synthesize native-folded eGFP. The rate of native-folded fluorescence has approached zero.

The kinetic rate of fluorescence degradation did not model Michaelis-Menten theory. The original theory proposed a single rate expression for fluorescence

degradation, k_dF. The data supporting the promoter activity were in-line with previous published results in [200]: the Hill constant (h) from Table 6 was 0.9, 1.8, and 1.9 for eGFP, TATeGFP, and (C)TATeGFP respectively. The published results from [200] were between 1.5-2.0. However, as observed in Figure 19, a multi-factor expression would be required to fit the curve. The data presented in this section does not yield a single curve. The slope of Figure 17 does not yield a single-straight line, but two distinct phases.

The specific activity of the cells in synthesizing native-folded GFP and the onset of fluorescence degradation should be modeled independently. The originally proposed model from Leveau, *et al.*, does not account for the additional two phases that are observed in Figure 17. These two phases, existing after the cells are harboring a positive specific growth rate, do not follow a single-rate expression model, such as Michaelis-Menten kinetics. The Hill equation is generally applied to model the rate at which a gene is regulated by transcription factors. However, the data presented in this section supports that protein degradation occurs when the specific growth rate of the cells is negative, indicating decay. Thus, the specific growth rate and the degradation rate should be modeled independently.

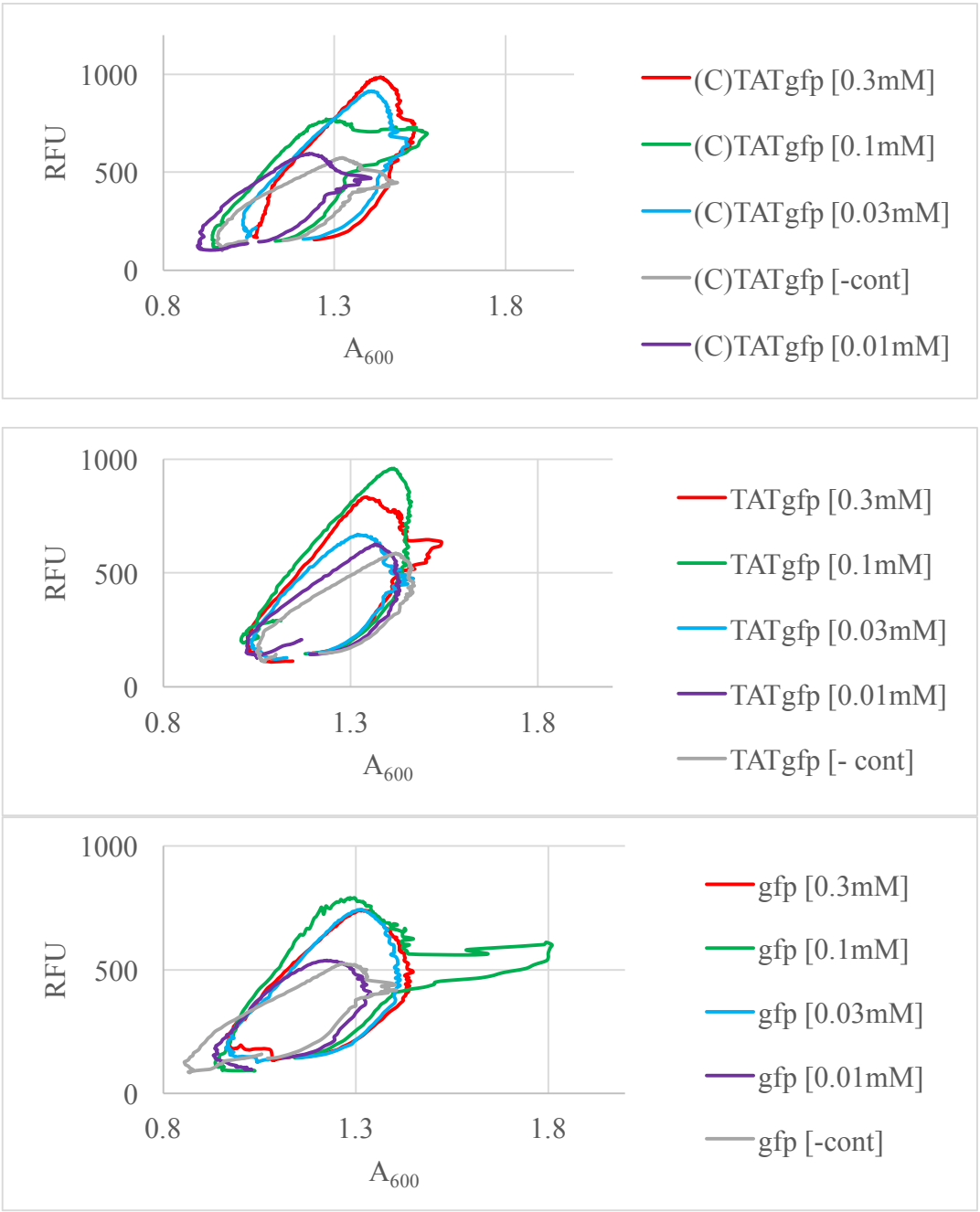


Figure 17 Relative Fluorescent Units as a function of optical density of eGFP variants after induction. Four concentrations of IPTG were used to induce expression of (C)TATeGFP, TATeGFP, and eGFP proteins. In a plot of RFU as a function of A_{600}, three distinct phases can be observed: the first is an accumulation of fluorescence as the cells are actively transcribing protein and it folds into a native-state, the second is when the cells enter a stationary phase and the activity of the promoter decreases, and the third phase is the cells enter the death phase and the rate of protein degradation reaches a maximum.

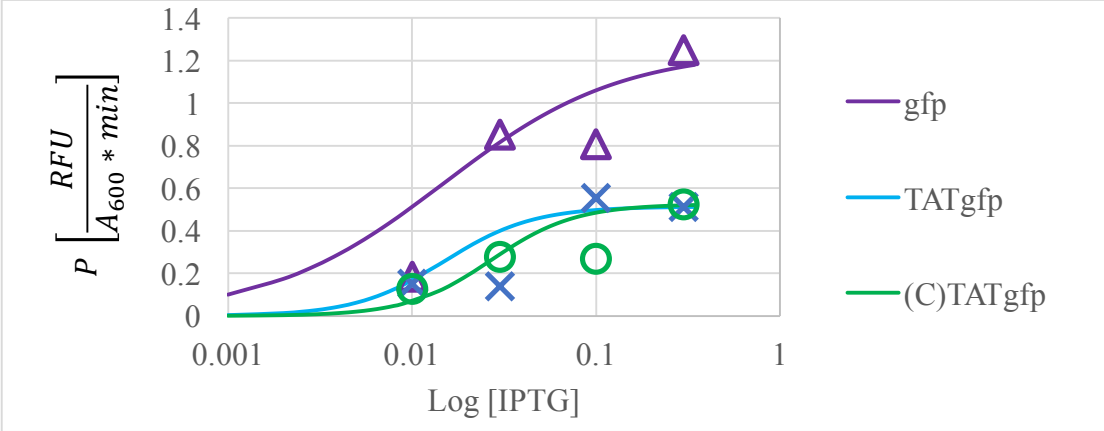


Figure 18. Modulation of single cell fluorescent protein synthesis as a function of IPTG concentration. The rate of protein synthesis, modeled by a fluorometric reporter, was modulated by the IPTG concentration.

	[IPTG]	A_{600}(t) Slope		F(A_{600}) Slope	P	P_{max}	K	h
		μ x10^3	(2**σ**) x10^3	F_{ss} σ				
		min^{-1}		$\frac{RFU}{A_{600}}$	$\frac{RFU}{A_{600} * min}$		Log [IPTG]	
eGFP	0.3	1.25	0.44	1578	1.98	1.245	0.015	0.910
	0.1	1.11	0.16	1381	1.53			
	0.03	1.25	0.49	1267	1.58			
	0.01	1.13	0.38	810	0.91			
	0	1.17	0.42	626	0.73			
TATeGFP	0.3	0.91	0.36	1297	1.18	0.512	0.015	1.800
	0.1	1.07	0.13	1152	1.22			
	0.03	0.84	0.20	965	0.81			
	0.01	1.02	0.35	812	0.82			
	0	0.99	0.28	682	0.67			
(C)TATeGFP	0.3	1.02	0.22	997	1.01	0.522	0.027	1.900
	0.1	0.92	0.32	826	0.76			
	0.03	1.01	0.43	759	0.77			
	0.01	0.93	0.41	667	0.61			
	0	0.87	0.41	563	0.49			

Table 6 Protein synthesis metrics as a function of RFU, A_{600}, and IPTG concentration. Fluorescence and optical density curves were examined to calculate the single cell fluorescence of eGFP, TATeGFP, and (C)TATeGFP. Using a 2-log range of IPTG concentrations, the relative protein synthesis activity of a cell is indicative of the rate of native-folding of eGFP isoforms.

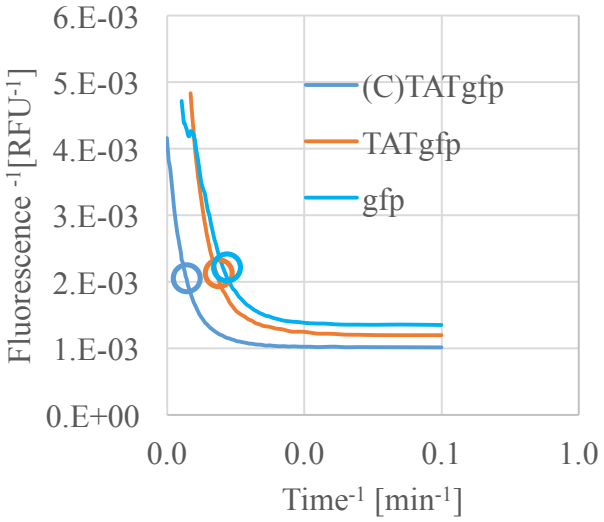


Figure 19 Michaelis-Menten Kinetic Model of [0.3mM IPTG] eGFP, TATeGFP, and (C)TATeGFP during fluorescence degradation. After the maximum, the degradation pattern is complex and Michaelis-Menten kinetic model does not converge for the data. It is expected the plot would produce a straight line. The experimental half-life fluorescence is the enlarged point on the respective curve.

3.5.1.3 Modeling *in vitro* protein degradation using first order decay kinetics

Upon fluorescence degradation, mathematical modeling can extrapolate an estimated fluorescence half-life for the eGFP, TATeGFP, and (C)TATeGFP constructs during protein expression. The half-life of fluorescence during protein expression can determine the relative stability of the protein when exposed to possible chemical, thermal, and enzymatic degradation conditions. These conditions can be induced when the protein is expressed intracellularly, or upon the cell death phase, and the creation of a porous membrane, extracellular degradation. Upon observation, the degradation of fluorescence emission does not follow a standard first-order kinetic decay. A semi-log plot vs. time of the fluorescence after the maxima does not form a straight line as a first-order reaction should, as observed in **Figure 20**. The three eGFP variants, expressed at 0.1mM IPTG exhibit slow decay and then a rapid decrease. Therefore, the modified kinetic decay model for fluorescence incorporates a parabolic function for the decay constant, $At^2 + Bt + C$.

$$F_t = F_{max} e^{-k_d t} \qquad \text{Equation 26,}$$

$$-k_d t \sim At^2 + Bt + C$$

$$\tau = \frac{\ln 2}{k_d t} \qquad \text{Equation 27}$$

The degradation rate can be modeled by exponential regression and decay kinetics. The residual sum of squares in **Equation 23** was minimized while modifying the constants from **Equation 26**. The half-life can then be calculated from

Equation 27. The number of empirical half-life calculations were limited based on the duration of the experiment.

All four IPTG concentrations of the (C)TATeGFP model extended fluorescence half-life when compared to compared to the TATeGFP and eGFP models. However, there also appears to be increased half-life fluorescence as the IPTG concentration increases. When comparing the half-life of the (C)TATeGFP models, the half-life is at a maximum using 0.3mM IPTG and a minimum using the lowest concentration of 0.01mM. For the non-cyclized protein models, there is no correlation between half-life fluorescence and IPTG concentration.

The relative fluorescence of the (C)TATeGFP constructs were relatively lower than that wild type eGFP and TATeGFP constructs. This is expected as the intein processing is a kinetic reaction and some of the intein processing occurs *ex vivo*. The data correlates relatively well with previous published results of eGFP fluorescence degradation during protein expression, 2.6 hours for half-life. The average half-life of eGFP over the four IPTG concentrations was 3.56 ± 0.26 hours.

The parabolic rate law proposed in this experiment is an empirical model with no fundamental underlying mechanism. This law was proposed to allow for a single exponential function to fit the complex degradation patterns of native-eGFP fluorescence during protein induction. The Michaelis-Menten kinetic fitting supported a new hypothesis that fluorescence degradation is complex. Modeling the degradation using a single-parameter expression or first-order kinetics does not account for the complex factors that contribute to the fluorescence degradation.

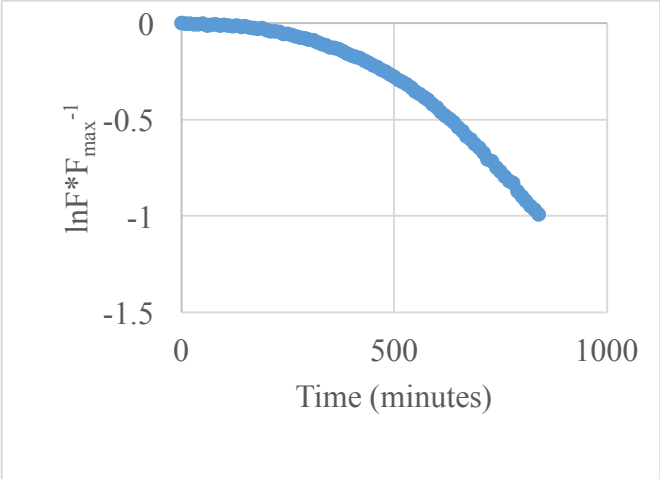


Figure 20. Semi-logarithmic plot of Fluorescence as a function of time for (C)TATeGFP [0.3mM] after Fluorescence Maximum. After the maximum fluorescence, the decay in fluorescence does not follow a first-order kinetic model with a single-constant, exponential decay model. This model suggests the decay constant is a parabolic function, $At^2 + Bt + C$.

	0.3mM IPTG		
	A	**B**	**C**
eGFP	-5.6E-07	-1.5E-03	1.3E-01
TATeGFP	-3.4E-06	1.7E-04	-1.2E-02
(C)TATeGFP	-1.7E-06	2.6E-04	-1.2E-02
	0.1mM IPTG		
	A	**B**	**C**
eGFP	-6.3E-06	7.6E-04	-2.6E-02
TATeGFP	-2.1E-06	6.5E-05	2.0E-02
(C)TATeGFP	-2.7E-06	7.9E-04	-7.2E-02
	0.01mM IPTG		
	A	**B**	**C**
eGFP	-3.1E-06	-1.7E-05	1.1E-02
TATeGFP	-1.3E-06	-1.5E-03	1.1E-01
(C)TATeGFP	-1.8E-06	-1.1E-04	4.2E-02
	0.03mM IPTG		
	A	**B**	**C**
eGFP	-3.9E-06	9.4E-05	-1.4E-03
TATeGFP	-3.2E-06	1.6E-04	-1.2E-03
(C)TATeGFP	-3.3E-06	5.0E-04	-1.8E-02

Table 7: Exponential coeficients of eGFP, TATeGFP, and (C)TATeGFP by regression analysis, during *in vitro* fluorescence degradation. The regression coeficients were determined to estimate the fluorescence half-life of the constructs using four different concentrations of IPTG during induction: 0.01, 0.03, 0.1, and 0.3mM. The results shown are are the average of five replicates with one standard deviation. All models had >98% coefficient of determination.

	0.3mM IPTG		0.1mM IPTG	
	Experimental ± σ	Regression ± σ	Experimental ± σ	Regression ± σ
eGFP	225 ± 51.8	232 ± 4.51	195 ± 36.0	232 ± 4.51
TATeGFP	235 ± 30.8	237 ± 5.24	295 ± 20.6	298 ± 4.51
(C)TATeGFP	355 ± 29.9	360 ± 4.54	325 ± 25.4	322 ± 3.87
	0.03mM IPTG		**0.01mM IPTG**	
	Experimental ± σ	Regression ± σ	Experimental ± σ	Regression ± σ
eGFP	230 ± 49.1	235 ± 4.24	210 ± 33.5	217 ± 4.11
TATeGFP	200 ± 41.4	203 ± 3.06	245 ± 57.3	247 ± 2.59
(C)TATeGFP	305 ± 81.5	304 ± 3.65	275 ± 53.3	267 ± 2.59

Table 8. *In vitro* half-life fluorescence (minutes) of eGFP, TATeGFP, and (C)TATeGFP, induced at half-log IPTG concentrations (0.01, 0.03, 0.1, and 0.3mM), based on exponential regression kinetics. After induction, eGFP was monitored by online fluorescence (485ex, 516em) every 10 minutes. After culmination of the stationary phase, the fluorescence reached a maximum for all conditions. Therafter an expontential decrease in fluorescence was modeled using regression analysis. For the experimental data, the results shown are are the average of five replicates with one standard deviation. For the regression data, the half-life error is the standard deviation of the regression analysis.

3.5.2 Fluorescence half-life under induced chemical degradation conditions

Exposing a protein to a chaotropic agent results in protein unfolding and a rapid, forced degradation study. The intein-fusion proteins in this experiment, (C)TATeGFP and (C)eGFP, exhibit enhanced fluorescent stability under chaotropic agents. From **Figure 21**, GuHCl induced rapid decay of fluorescence at 5.0M, whereas the intein-fusions required the duration of the experiment (20 minutes) for fluorescent decay. A decrease in GuHCl resulted in an increase fluorescent stability from all four models. The (C)eGFP models retained >50% fluorescence during the experiment, whereas the linear models (eGFP and TATeGFP) had complete decay within 15 minutes. All conditions modeled equal stability under the lowest GuHCl concentration of 1.0M.

The regression modeling of the decay kinetics correlates to experimental data. For conditions where a half-life can be experimentally determined, the regression modeling half-life calculation compares. In **Equation 26**, three variables are determined by regression analysis: the initial fluorescence [F_i], first-order rate constant [k_d], and an offset factor [C]. The offset factor is included in the case a chaotropic concentration does not completely denature the protein. When assessing the model validity, the half-life, calculated from the first-order rate-constant extrapolates to similar experimental half-life empirical values for all 5.0M GuHCl conditions, and the two linear eGFP motifs for 3.0M GuHCl. The initial fluorescence, from regression analysis should model the normalized initial fluorescence and approach unity. All regression models in this experiment have

initial fluorescence approaching unity. Since all models used a relative fluorescence where $F_t = F*F_i^{-1}$, a regression model should have an $F_i \sim 1.0$.

The chemical degradation of eGFP variants in this experiment demonstrate the enhanced stability of an intein-based, cyclized protein. Two (C)eGFP proteins were tested in this experiment alongside their wild-type counterpart, and in the 3.0M and 5.0M conditions, the intein-based proteins exhibited enhanced fluorescence half-life. Since the intein-based proteins have a covalent bond that forms between the N-terminal and C-terminal amino acids, the backbone appears to be more resistant to destabilization of hydrogen bonding and maturation of the proper folding conformation for protein functionality. Since eGFP fluorescence is an excellent reporter for backbone stability and maturation of the proper folding, it can be deduced that a longer fluorescence half-life, under chaotropic conditions, results from protein cyclization.

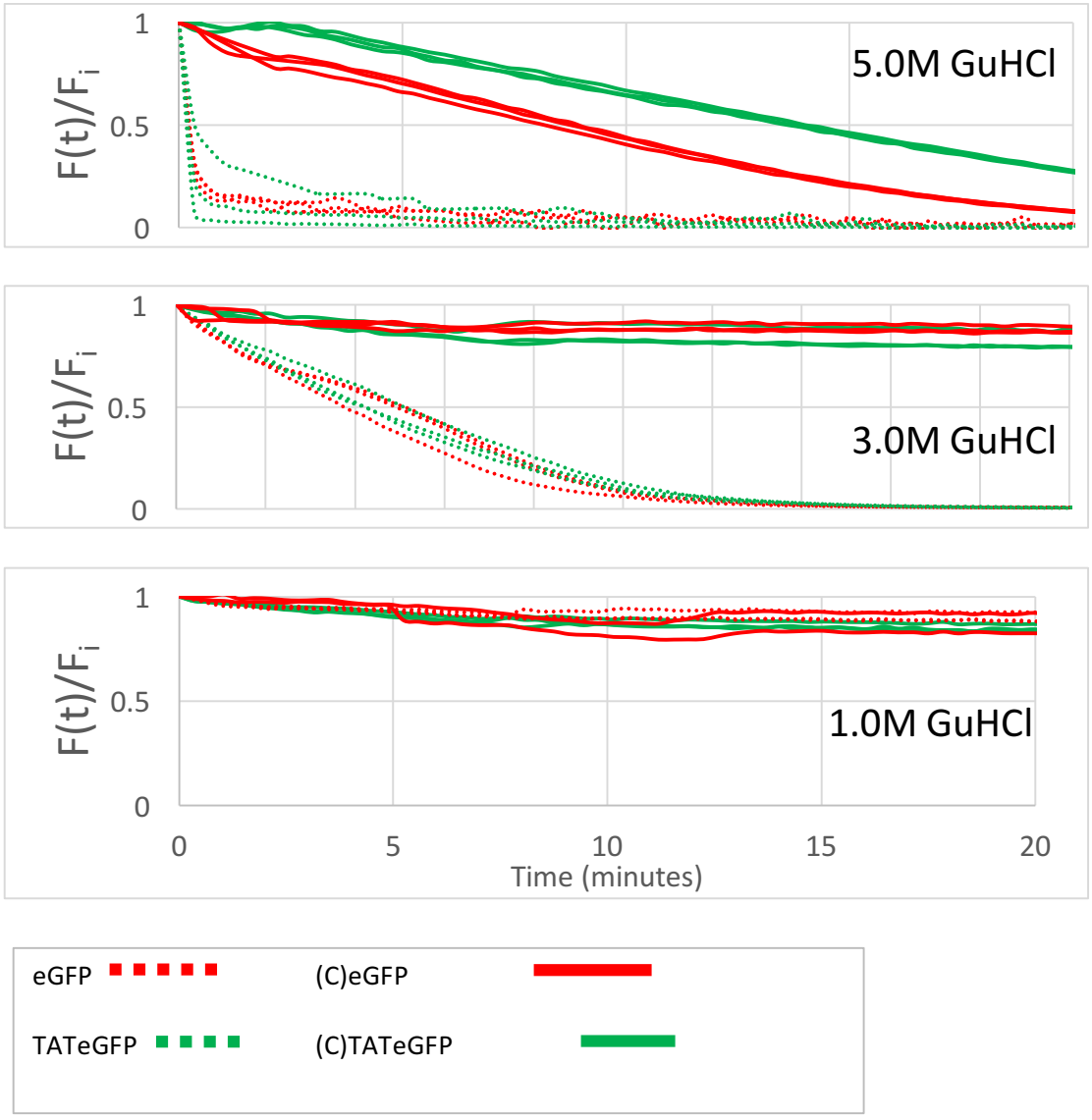


Figure 21. Forced chemical degradation from GuHCl on purified gfp, (C)eGFP, TATeGFP, and (C)TATeGFP from a chaotropic solution. Purified gfp variants were exposed to chaotropic conditions and the fluorescence was monitored in real-time. Unfolding of proteins could be monitored by the fluorometric reporter as a function of time.

	5.0M GuHCl		3.0M GuHCl	
	$F_i \pm \sigma$	**k_d [min^{-1}]**	**$F_i \pm \sigma$**	**k_d [min^{-1}]**
eGFP	0.951 ± 3.72E-02	2.98	1.06 ± 3.85E-02	0.207
TATeGFP	0.958 ± 3.18E-02	3.10	1.08 ± 3.42E-02	0.200
(C)eGFP	0.958 ± 5.72E-02	0.105	0.92 ± 1.46E-02	2.52E-03
(C)TATeGFP	0.958 ± 7.45E-02	7.00E-02	0.91 ± 2.14E-02	5.34E-03

	1.5M GuHCl	
	$F_i \pm \sigma$	**k_d [min^{-1}]**
eGFP	0.940 ± 2.67E-02	5.88E-03
TATeGFP	0.945 ± 6.09E-03	5.38E-03
(C)eGFP	0.946 ± 1.13E-02	2.77E-03
(C)TATeGFP	0.936 ± 1.74E-02	4.81E-03

Figure 22. Exponential coeficients of eGFP, TATeGFP, (C)eGFP, and (C)TATeGFP by regression analysis, from forced chemical degredation by GuHCl. The regression coeficients [F_i, k_d] were calculated from Equation 25 and Equation 27. The results shown are are the average of three replicates with one standard deviation.

eGFP Variant	5.0M GuHCl		3.0M GuHCl	
	Experimental [min]	**Regression [min ± σ]**	**Experimental [min]**	**Regression [min ± σ]**
eGFP	<0.35	0.232 ± 0.03	4.20-4.55	3.35 ± 0.03
TATeGFP	<0.35	0.223 ± 0.03	4.20-4.55	3.47 ± 0.04
(C)eGFP	8.40-8.75	6.63 ± 0.52	N/A	274 ± 2.11
(C)TATeGFP	13.65-14.00	9.90 ± 0.74	N/A	129 ± 1.46
	1.5M GuHCl			
	Experimental [min]	**Regression [min ± σ]**		
eGFP	N/A	117 ± 2.68		
TATeGFP	N/A	128 ± 6.07		
(C)eGFP	N/A	249 ± 1.13		
(C)TATeGFP	N/A	144 ± 1.74		

Figure 23. Fluorescence half-life (minutes) of eGFP variants under induced chemical degradation. Four variants of eGFP were exposed to high concentrations of GuHCl to destabilitize hydrogen bonding and induce a depression in fluorescence, regression half-life fluorescence listed in minutes. The results shown are the average of three replicates with one standard deviation.

3.5.4. Fluorescence half-life under induced thermal degradation conditions

Thermal degradation studies are often executed above normal operating conditions of a protein. For a protein therapeutic, the majority of the residence time is occupied in a frozen vial, awaiting administration to a patient. Therefore, a stability study at 37.0°C will be an accelerated stability study. In this study, the accelerated temperature stability study did not reveal a significant enhancement of the (C)TATeGFP as shown in Table 10. The experimental setup for this study was not executed long enough to determine an experimental half-life as the protein solution experienced evaporation of the PBS matrix over the 16-hour duration of the experiment. Using the regression analysis methods described in **3.5.1.3**, the regression fluorescence half-life was calculated. Wild-type eGFP exhibited the highest fluorescence half-life. TATeGFP, alongside (C)eGFP and (C)TATeGFP had regression-derived fluorescent half-life within 1-standard deviation.

Reported fluorescent half-life of wild-type eGFP is a broad range, from 24-26 hours at 37.0°C [184, 205], to 2.8 hours *in vitro* expression using fibroblast cells [183], 4.5-6.1. It is possible the thermal stability enhancement of (C)TATeGFP could occur at higher temperatures or could be empirically observed during a longer-duration experiment. Theoretically, enhanced stabilization of the hydrogen bonding in eGFP that is created using intein-based proteins should provide enhanced thermal stability. This theoretical prediction was confirmed when measuring fluorescence half-life during *in vitro* protein expression and protein degradation using GuHCl.

eGFP Variant	37.0°C	
	F_i	**k_d [1/min]**
eGFP	0.995	$7.2x10^{-4}$
TATeGFP	0.981	$8.1x10^{-4}$
(C)eGFP	1.01	$7.7x10^{-4}$
(C)TATeGFP	1.00	$7.8x10^{-4}$

Table 9. Exponential decay kinetic constants of induction of thermal degradation of fluorescent eGFP variants.

eGFP Variant	37.0°C	
	Exp	**Regression ± σ**
eGFP	N/A	996 ± 14.1
TATeGFP	N/A	854 ± 38.2
(C)eGFP	N/A	896 ± 27.0
(C)TATeGFP	N/A	893 ± 22.1

Table 10. Fluorescence half-life (minutes) of eGFP variants under induced thermal degradation. Four variants of eGFP were exposed to 37.0°C for 16 hours to induce degradation from temperature induced fluorescence degradation. The regression analysis (listed in minutes) does not appear to show any significant enhancement from the cyclized Intein proteins. The results shown are are the average of three replicates with one standard deviation.

3.5.5 Enzymatic Degradation of (C)eGFP

One of the unique features of the intein-based, cyclized protein expression system, is the synthesized protein's covalent bond between the N-terminal and C-terminal amino acid. In this study, a carboxypeptidase was used to measure the activity of C-terminal amino acid digestion. The hypothesis was that a linear protein would be subjected to C-terminal amino acid digestion, whereas a cyclized protein would be protected. In this experiment, after 24 and 72-hours, the cyclized-eGFP protein had no observable degradation. Whereas, the linear, unprocessed intein had multiple degraded artifacts.

Using MALDI-TOF analysis, the digested solution was analyzed for proteins and a calculated molecular weight could be observed. After 24 hours, the observed peak at the theoretical molecular weight of eGFP remains in-tact. However, several new peaks are observed for the linear, un-processed intein fragment. Suggesting exopeptidase digestion, this confirms the hypothesis of slow CpA digestion. After 72 hours, the peak at the theoretical molecular weight of eGFP remains in-tact. Only two predominant peaks are observed for the linear, un-processed intein fragment. This experiment confirmed our hypothesis as degradation could be observed after 24 and 72-hours of the unprocessed intein, whereas the (C)eGFP retained its theoretical average mass. Several degraded isoforms of the unprocessed intein were observed.

Using the amino acid sequence of the two proteins, the measured mass of the protein fragments can be predicted. The unprocessed intein and (C)eGFP C-terminal amino acid sequences (Figure 24) have several amino acids that are slowly digested by CpA—K, R, P. Using the NIST molecular weight calculator, digested proteins are

compared against the measured molecular weight of the MALDI in Table 11. It should be noted the measured and theoretical molecular weight of (C)eGFP for all three time points remains constant and accurate. Moreover, as a control for the assay, the measured and theoretical molecular weight of CpA are consistent. This experiment supports the hypothesis that a cyclized protein, absent of a C-terminal amino acid, will not be subjected to exopeptidase digestion.

Unprocessed Intein: LDLITLENIKQTEEALDNHRLDFGTIKHHHHHH-$_{COOH}$

(C)eGFP: DPNEKRDHMVLLEFVTAAGITLGMDELYKEF-$_{COOH}$

Figure 24 C-terminal amino acids of the (linear) unprocessed intein fragment and the transcribed (C)eGFP fragment. Three enzymes have slow kinetic rates of digestion by CpA (K, P, R) that are listed in red.

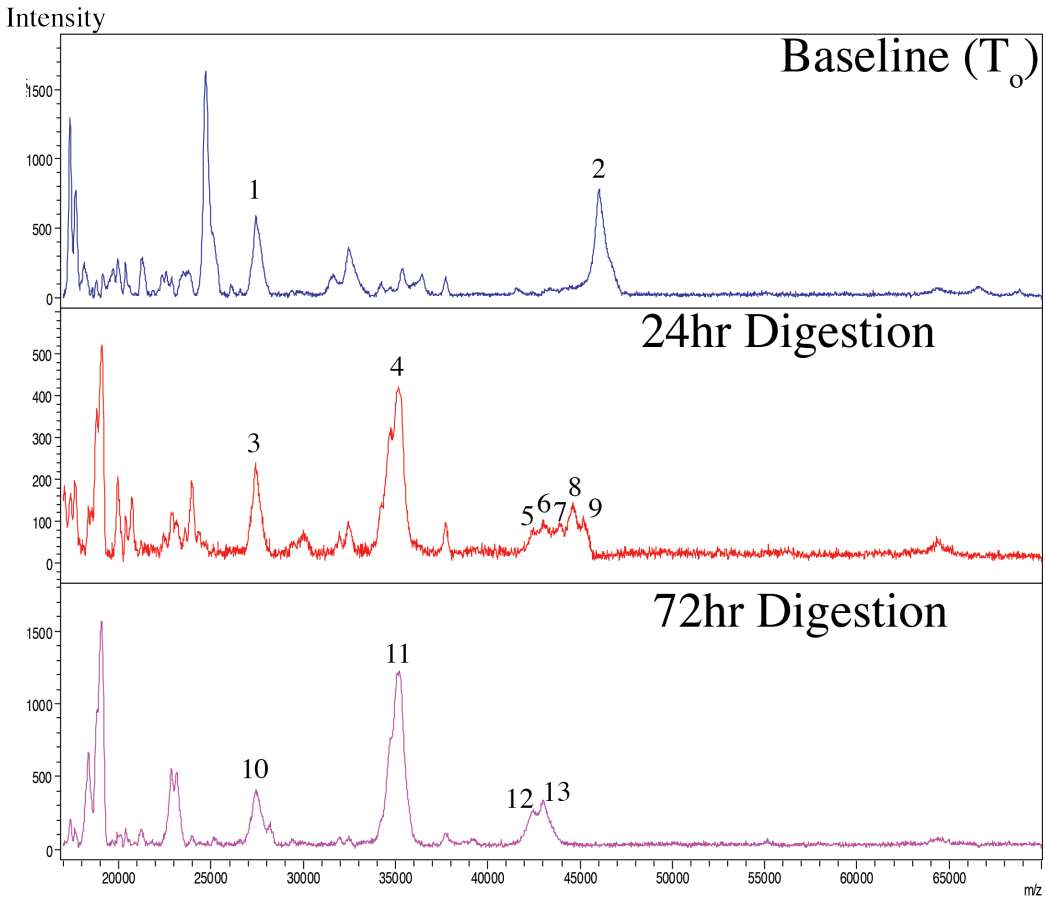


Figure 25. Digestion analysis of cyclic and linear proteins using MALDI-TOF. A protein solution containg a (C)eGFP and the unprocessed intein fragment were subjected to 24 and 72-hour digestion of Carboxypeptidase-A. After incubation, the reaction was quenched with PMSF and analyzed using a MALDI-TOF spectrum analysis. The molecular weight of the observed molecular weight peaks is identified in Table 11.

Peak	Theoretical Protein	Measured Mass (m/z)	Theoretical Average Mass (m/z)	C-terminal Amino Acid [Position]	Δppm
1	(C)eGFP	27423.04	27415.02	F [N-0]	292.54
2	Unprocessed Intein	46057.31	46,071.03	H [N-0]	297.80
3	(C)eGFP	27414.33	27415.02	F [N-0]	25.19
4	CpA	34815.46	34823.45		229.44
5	Unprocessed Intein	42990.97	42,955.69	I [N-29]	821.31
6	Unprocessed Intein	43001.86	42,955.69	E [N-25]	1074.83
7	Unprocessed Intein	43924.70	43,920.68	L [N-17]	91.53
8	Unprocessed Intein	44619.77	44,601.57	D [N-11]	408.06
9	Unprocessed Intein	45207.13	45,209.28	K [N-5]	47.56
10	(C)eGFP	27418.40	27415.02	F [N-0]	123.29
11	CpA	34835.78	34823.45		354.07
12	Unprocessed Intein	42452.05	42,485.12	I [N-29]	778.39
13	Unprocessed Intein	42990.97	42,955.69	E [N-25]	821.31

Table 11. Measured and Theoretical Mass Tabulation of MALDI-TOF resolved proteins after CpA digestion. After 24 and 72 hours of digestion, proteins were subjected to molecular weight analysis using mass spectrometry. The table calculates the relative error (Δppm) of the theoretical protein identified from the MALDI-TOF spectrum analysis.

3.6. Conclusions

The practicality of this experiment for transfer into an industrial biotechnology manufacturing environment for clinical or commercial products is significant. Without an online and real-time measurement for protein stability, a process must rely on a robust process to determine harvest timing of the process. As discussed earlier, this harvest time is a determined range that allows for a period of time when the harvest can be initiated. Enhancing the stability of a protein during expression allows for a longer duration of a harvest time. This could alleviate termination of a batch for missing a harvest time.

The accelerated stability study executed in this experiment provided empirical evidence that (C)TATeGFP has enhanced fluorescent stability under protein degradation conditions. As a fluorometric reporter, the (C)TATeGFP functioned to increase the fluorescent half-life when tested against wild-type TATeGFP and eGFP. The direct application of this study would be overall long-term recovery of (C)TATeGFP protein solutions under physiological conditions, storage temperatures, and protein expression conditions. Expressing a lead protein candidate as a cyclized intein would provide a clinical or commercial manufacturing company developing therapeutic proteins desired enhanced stability.

The hypothesis for enhanced thermal and chemical stability of intein-fusion proteins, originates from stronger hydrogen bonding in the protein backbone. If a protein has flexible regions and rigid regions, an intein-based expression system

could provide enhanced stability in either region. The hydrogen bonding in a protein backbone provides protection from hydrophobic regions in the protein. Essential to protein structure-function, these hydrophobic regions are critical to a proteins overall efficacy. When exposed to high temperatures or chaotropic agents, the rigid structures maintain protection of the hydrophobic regions for an enhanced period of time.

The covalent bond between the N-terminal amino acid and C-terminal amino acid in an intein-based expression system provides increased rigidity and protection for these hydrophobic regions. However, it does present limitations of the target cyclized protein. eGFP was selected for this study based on the close proximity of the C-terminal and N-terminal amino acids under native-folding conditions. For more complex proteins, such as antibodies, or proteins that do not have neighboring amino acids, the intein-based system might express a protein that folds into an inactive secondary structure.

Another significant enhancement of the intein-based protein expression system is a broader manufacturing range for harvest time. Using the fluorometric reporter system, the (C)TATeGFP exhibited a longer equilibrium time when the protein expression rate began to decline and the protein degradation rate increased. Moreover, when protein degradation was observed *in vitro*, during protein expression, the fluorescence half-life of the (C)TATeGFP was enhanced. For a clinical or commercial manufacturing process, a longer equilibrium or slower protein degradation rate is a more robust process. This will allow for a longer acceptable harvest range that will maximize protein expression yields during upstream manufacturing.

A novel kinetic rate law for fluorescence degradation was proposed. This rate law allows for a single kinetic expression to model the complex degradation of native-folded eGFP during protein expression. The parabolic rate constant models a time-dependent rate constant. As the protein induction phase progresses, the specific growth rate decreases as the rate of native-state fluorescence degradation increases. A second contributing factor, incubation temperature, should enhance the degradation capacity of the natural proteases. This empirical rate law served to model a complex degradation pattern and to confirm the enhanced stability a cyclized protein conveys.

Chapter 4: Transduction kinetics and efficiency of a cyclical, cell-penetrating peptide-protein fusion.

4.1 Abstract

CPP-mediated uptake can be rapidly measured by quenching extracellular fluorescence with trypan blue. Trypan blue will penetrate cells with compromised membranes that could transduce proteins unmediated. In this study, eGFP is expressed alongside the transactivator of transcription (TAT) peptide derived from HIV-1 that can function as a cell-penetrating peptide. The TAT-eGFP fusion protein is also expressed alongside an artificial split-intein system. A cyclical protein, expressed as a intein-based fusion protein, exhibits enhanced protein stability against thermal, chemical, and enzymatic degradation. The (C)TATeGFP exhibits enhanced *in vitro* stability during protein expression, resistance against exopeptidase digestion, and resistance against chaotropic agents. A cell-penetrating peptide-protein fusion can transduce a cell membrane to delivery an intracellular protein with low cytotoxic effects. A cyclical, cell-penetrating, peptide-protein fusion is a novel tool for fluorometric reporting studies and a backbone for the delivery of intracellular therapeutic proteins with enhanced thermal, enzymatic, and chemical stability.

This study used an *in vitro* assay to measure enhanced transduction of a cyclical, cell-penetrating, peptide-protein fusion using eGFP and mCherry as fluorometric reporters. HEK293 cells were plated on culture plates and incubated alongside 5μm and 2.5μm of fluorometric reporters: eGFP and mCherry, expressed as fusion proteins with the TAT peptide and the intein-based system. Extracellular fluorescence was quenched using trypan blue and the intracellular fluorescence was

measured using a plate reader. Trypan blue also functioned as a utility to confirm the low cytotoxicity of the compounds tested. Inclusion of the TAT-peptide in the (C)eGFP increased intracellular fluorescence by 75.4 ± 5.14 (5μm) and 74.6 ± 9.72% (2.5μm) for the cyclized (C)TATeGFP fusions in comparison to the 43.8 ± 6.37% (5μm) and 71.9 ± 9.79% (2.5μm) of the TATeGFP fusion. Inclusion of the TAT-peptide in the mCherry Intein increased intracellular fluorescence by 93.5 ± 18.2 (5μm) and 79.1 ± 15.3% (2.5μm) for the cyclized (C)TATmCherry fusions in comparison to the 133 ± 12.4% (5μm) and 108 ± 21.2% (2.5μm) of the TATmCherry fusion.

4.2 Introduction

Cell-penetrating peptides have the ability to transport molecules through the cell membrane, while retaining the function of the transported molecule. Discussed earlier, it is hypothesized that Maternin requires intracellular delivery for inhibition of tumor proliferation in carcinoma cell lines. As a fusion protein to eGFP, Maternin did not exhibit efficacy when presented to the outer membrane of KSY-1, HTB-123, and PC-3 cancer cell lines. A TAT peptide-eGFP fusion protein could be a functional reporter for intracellular delivery modeling and analysis. Essential for a functional delivery system, the transported molecule's intracellular function can be modeled by a fluorometric reporter. A TAT-fluorometric reporter fusion protein is an effective tool to measure transduction capacity. In this experiment, the effectiveness of a cyclized TAT-fluorometric reporter is tested. The TAT peptide has been covalently attached to wild-type eGFP and then cyclized, while maintaining eGFP fluorescence

in the cytoplasm of live cells [206]. However, this model leaves the transduced eGFP protein susceptible to thermal, enzymatic, and chemical degradation. It has been previously discussed, a cyclized TATeGFP fusion offers enhance thermal, chemical, and enzymatic stability to the wild-type counterpart. In this experiment, the first *Ssp* DnaE split-intein derived, fully-cyclized TATeGFP fusion, is tested for transduction capacity and intracellular fluorometric reporting.

A transduction assay should measure the potential of compounds to traverse the outer membrane of live cells. Using a fluorescent protein as a fluorometric reporter, the intracellular fluorescence of a cell is an effective metric from transduction capacity. However, the extracellular fluorescence of the surrounding media must not contribute to the signal. In this experiment, trypan blue is used to mute the extracellular fluorescence. Trypan blue is a multi-functional tool in the transduction assay as it will also quench intracellular fluorescence of dead cells that could permit transduction of all exogenous proteins. Endogenous auto-fluorescence is expected to contribute to the quenched fluorescent signal. The purified fluorometric reporter proteins will be quenched in PBS to determine the fluorescence quenching efficiency of trypan blue.

The second phase of the experiment is to measure the *in vitro* transduction capacity of the TAT-fluorometric reporter system. Actively-cultured cells, co-incubated with the proteins would be incubated for several days and the cell lysate fraction would be probed for fluorescent proteins using western blot analysis. It is common to include a ubiquitous protein, such as actin, to normalize the total lysate protein fraction concentration.

Previous experiments have tested the transduction kinetics and capacity of TATeGFP fusions. There has been a widely reported timescale for the maximum fluorescence measured of intracellular delivery of TATeGFP: from 15 seconds [207], to 30 minutes [208], to 24-hours [209]. Secondly, there has been a widely reported time scale for the rate of transduction, expressed as a fluorescent half-life, or the amount of time required for half of the maximum fluorescence to be measured: from <1 minute [210], to 13.2 minutes for fluorescence half-life [211]. Finally, for *in vitro* transduction capacity of TATeGFP proteins in a cell culture environment, the detection of eGFP in the lysate fraction has been identified on an anti-GFP western blot after 48 hours [212] and 72 hours [213]. We speculate several factors contribute to the broad range of kinetics: purity of the purified protein, unique ORF features, expression plasmid and promoter, and execution of the fluorometric assay.

This experiment is design to measure the rate of transduction of a novel cell-penetrating peptide-protein fusion, and long-term stability *in vitro*.

4.3 Materials and Methods

Purified protein samples were expressed and purified as described in 2.3.1, 2.3.2, 2.3.3, 2.3.4.

4.3.1 Cellular Transduction Assay

HEK293 cells (Pavlaikis, NCI-Frederick) were cultured in DMEM+2mM L-Glutamine and 10% FBS, incubated at 37.0°C and 5% CO_2. The cells were plated onto a 24-well plate at 5E5 vc/mL in 1.0mL and incubated overnight. The media was aspirated and the cells were washed three times in PBS, pH 7.4. A 1.0mL solution

containing the target compound was overlaid onto the cells and incubated at 37.0°C and 5% CO_2. Extracellular fluorescence was quenched by the addition of 100μL 0.4% trypan blue and fluorescence was measured using a 485ex/516em filter for eGFP and 588ex/609em for mCherry. Quenching of extracellular fluorescence was confirmed by incubating the compounds in PBS and adding 10% (v/v) of 0.4% trypan blue and measuring fluorescence.

4.3.2 Western blot analysis of HEK293 lysates

HEK293-F cells (Invitrogen, Irvine, CA) were cultured in serum-free, 293-Freestyle media in 125mL shake flasks, agitated at 0.55m/s at 37.0°C and 5% CO_2. 5μm of (C)TATeGFP and 5μm of (C)eGFP were added to the shake flask. 24, 48, and 72 hours, post-addition, a 5.0mL sample was removed from the culture and centrifuged at 200rcf for 10 minutes. The cell pellet was washed three times with PBS, pH = 7.4. The pellet was lysed with 0.1% TritonX-100 and denatured with 10U/mL Benzonase (EMD-Millipore, Billerica, MA). The lysate was centrifuged at 15,000rcf for 30 minutes. The resulting supernatant was mixed equally with 2x Laemmeli Buffer + βME. The proteins were separated using a 4-15% Tris-HCl gradient, reducing gel and then transferred to a nitrocellulose membrane. The blot was probed with 1/1000 dilution anti-GFP and 1/2000 anti-β-actin (Thermo, Rockford, Il) primary antibodies, conjugated with HRP.

4.4 Experimental Design

This experiment was designed to measure the transduction kinetics and efficiency of cyclical TAT-fusion proteins. A fluorometric reporter system was utilized by expressing the TAT-fusions with eGFP and mCherry. The cyclic, fluorometric reporter proteins were produced using an intein-based expression system. Trypan-blue was used to quench extracellular fluorescence of the reporter proteins and quench the intracellular fluorescence of non-viable cells with porous membranes. If the TAT-fusion protein could transduce its fluorometric reporter across the cell membrane, then intracellular fluorescence could be measured. Expressing the TAT-fusion protein with the intein-based system could measure the effect of a cyclized protein on transduction efficiency. We hypothesized inclusion of the TAT-peptide would increase transduction efficiency of a fluorometric reporter, expression a fluorometric reporter-TAT fusion using the intein-based system could transduce cells, and the fluorometric reporters would retain structure/function and fluoresce inside the cell. It is hypothesized that cells will emit auto-fluorescence and the fluorescent quenching efficiency of trypan blue will not be complete. Therefore, proteins were also incubated in PBS to determine the fluorescent quenching efficiency of trypan blue.

Forced degradation of the cyclized proteins were extensively discussed in Chapter 3. In this experiment, *in vitro* degradation could be measured by probing the lysate fraction of cells for the fluorometric reporter. Cells that were trypsinized and lysed could be probed for eGFP using a primary antibody. As a control to normalize for total protein fraction, a housekeeping protein, actin was employed. It was

hypothesized that the lysate fraction of cells incubated with a TAT-fusion protein could be uniquely identified by an anti-GFP antibody. Moreover, as the incubation time of the TAT-fusion protein increased, degradation of the internalized protein could be observed by the relative signal of the lysate fraction using the anti-GFP antibody.

4.5 Results and Discussion

4.5.1 Cellular Transduction Assay

The transduction assay used a fluorometric reporter system to measure cell-penetration of a target compound. In this experiment, eight compounds were tested: two fluorometric reporters (eGFP and mCherry), with and without the TAT peptide as a fusion protein, and with and without the intein-based expression system for production of cyclized proteins. Overall, proteins expressed with TAT as a fusion protein had higher transduction efficiency than proteins expressed without the cell-penetrating peptide. Expressing the fluorometric reporter with the TAT fusion in the intein-based system did not affect the transduction capacity of the protein. Finally, the two fluorometric reporters had relatively similar transduction capacity when expressed as a fusion protein with the TAT peptide.

Two concentrations of eGFP fluorometric reporter proteins were tested for transduction efficiency using adherent HEK293 cells. The relative fluorescence (Relative Fluorescence= $F_t = \frac{F_{485ex,516em}}{Fmax_{485ex/516em}}$ Equation 21) was utilized

to normalize the fluorescence signal and compare intra-assay samples. From Figure 28, increasing the concentration of the protein two-fold did not increase the transduction efficiency two-fold. Both transduction efficiency measurements were within one standard deviation of both measurements. For the (C)TATeGFP, the percent increase of 5µm and 2.5µm for the four time points (15, 30, 60, and 120 minutes) were 11.5, 14.1, 4.36, and 8.24% respectively. Whereas for TATeGFP, the percent increase of 5µm and 2.5µm for the four time points (15, 30, 60, and 120 minutes) were 1.39, (1.68), 6.23, and 13.01% respectively. There does not appear to be a dose-dependent effect on transduction efficiency. The saturation of transduced protein could be reached at 2.5µm

The kinetics of transduction were rapid and the maximum fluorescence was measured within 30 minutes of incubation. The sum of squares was minimized to calculate the F_{max}, and first order exponential rate-constant as described in **Equation 27** and **Equation 28**. The half-life ($\tau_{0.5}$) calculated from the regression models was < 6.5 minutes for all four TAT-fusion proteins (**Table 11**). The transduction experiment was repeated, measuring intracellular fluorescence of (C)TATeGFP and (C)eGFP every 2 minutes during incubation. The ($\tau_{0.5}$) calculated was more rapid, 1.77 minutes for the (C)TATeGFP (**Table 11**).

mCherry was used as a fluorometric reporter to determine if the transduction kinetics of the TAT cell-penetrating peptide are affected by the fusion partner. mCherry (28.8kDa) is of similar size to eGFP (26.7kDa) and forming a beta-barrel structure when under native folding conditions. The TATmCherry fusions both exhibited transduction capacity in HEK293 cells. The kinetic rate of transduction

was rapid for the cyclized and linear TATmCherry proteins. From Figure 30, the (C)TATmCherry and TATmCherry reached maximum fluorescence within 15 minutes. The intracellular fluorescence remained constant over the next 2 hours. The background fluorescence, from the (C)mCherry and mCherry incubated cells, was consistent for both proteins and both concentrations. At 120 minutes, the fluorescence of the TAT-peptide proteins decreased and was within one standard deviation of the wild-type proteins. The mCherry protein could have minimal cytotoxicity to the cells and this is expressed by the fluorometric assay after 2 hours.

Two factors could contribute to the background fluorescence of the fluorometric assay. From Figure 31, the fluorescence of the eGFP proteins incubated in PBS has a relative fluorescence of approximately 0.2. Whereas, the mCherry proteins have a nearly zero fluorescent signal. From Figure 28 and Figure 30, the wild type eGFP and mCherry proteins, respectively (without the TAT peptide) have a relative fluorescent signal of 0.25. Therefore, it can be assumed the quenching efficiency of mCherry is higher than that of eGFP.

The maximum fluorescence measured of trypan blue quenched HEK293 cells incubated with TAT-fusion proteins is higher than that without. The maximum relative fluorescence was determined for all eight proteins incubated alongside the adherent HEK293 cells and trypan blue. From Figure 32, the positive control compounds, linear TATeGFP and linear TATmCherry, both exhibited enhanced intracellular fluorescence in comparison to the non-TAT counterpart. The two proteins, (C)TATeGFP and (C)TATmcherry, exhibited comparable intracellular fluorescence to their linear counterparts. All proteins with the TAT peptide had an

enhanced intracellular fluorescence greater than one standard deviation of the non-TAT counterpart.

The results of this experiment are comparable to previously published results of testing TATeGFP proteins. The previous results utilized a non-cyclized TATeGFP and measured the amount of time required for fluorescence half-life between < 1 minute to 13.2 minutes. In this experiment, (C)TATeGFP and TATeGFP reached maximum fluorescence in less than 6 minutes. Likewise, the (C)TATmCherry and TATmCherry reached maximum fluorescence in less than 7 minutes. Based on this study, the TAT peptide does not show a difference in transduction efficiency or kinetics based on the fusion partner.

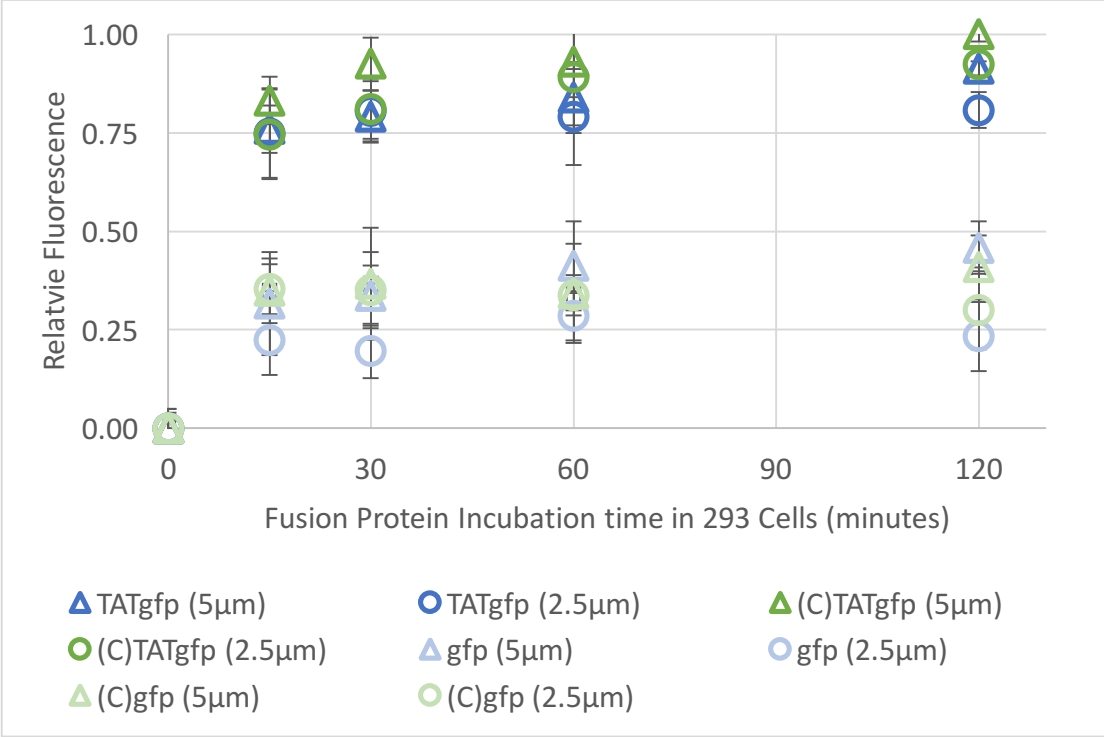


Figure 26. Transduction of HEK293 cells with 5µm and 2.5µm eGFP-fusion proteins. Cyclized (intein-based) and linear (wild-type) eGFP were used to measure the transduction capacity of the TAT-peptide. 5µm and 2.5µm of each protein (n=3) was incubated with adherent 293 cells for 2 hours. Extracellular fluorescence was quenched with 10% total volume of 0.4% trypan blue and fluorescence was measured using 485nm excitation/516nm emission filter, and the average of reads was calculated. The results shown are are the average of three replicates with one standard deviation.

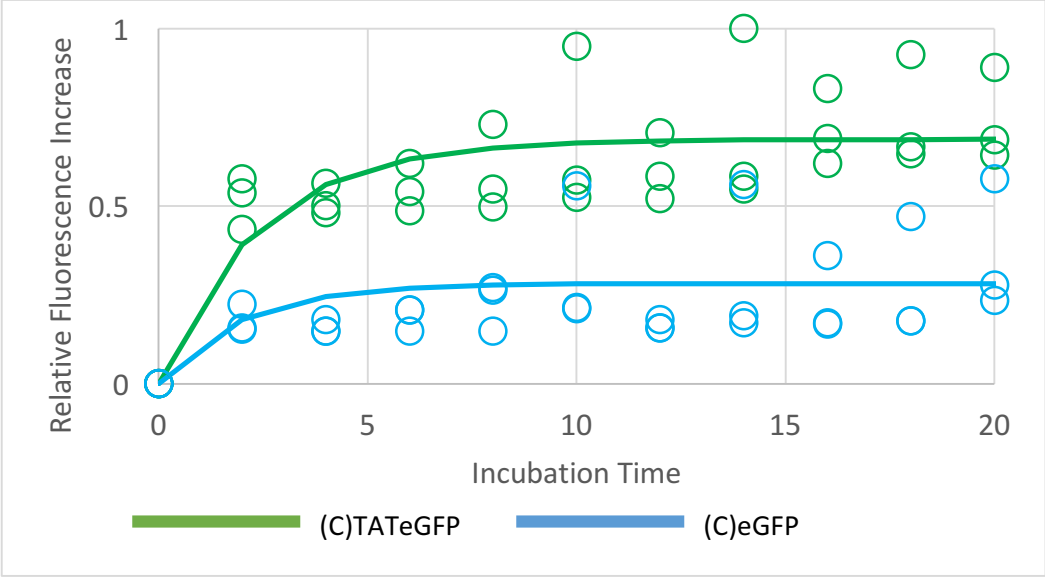


Figure 27. Short duration transduction analysis of (C)TATeGFP and (C)eGFP in 293 cells. 5µm of each fusion protein was incubated alongside HEK293 cells. Over 2-minute incriments, 10% total volume of 0.4% Trypan blue was added to quench extracellular fluorescence and intracellular fluorescence was measured for eGFP (485ex/516em).

	$k\left[\frac{1}{\text{min}}\right]$	F_{max} Regression (±σ)	F_{max} Experimental (±σ)
(C)TATeGFP	0.39	0.69 ± 0.17	0.78 ± 0.14
(C)eGFP	0.25	0.28 ± 0.16	0.36 ± 0.19

Table 12. First-order kinetic analysis of short duration transduction capacity of (C)TATeGFP and (C)eGFP fusion proteins in 293 cells. The results shown are are the average of four replicates with one standard deviation.

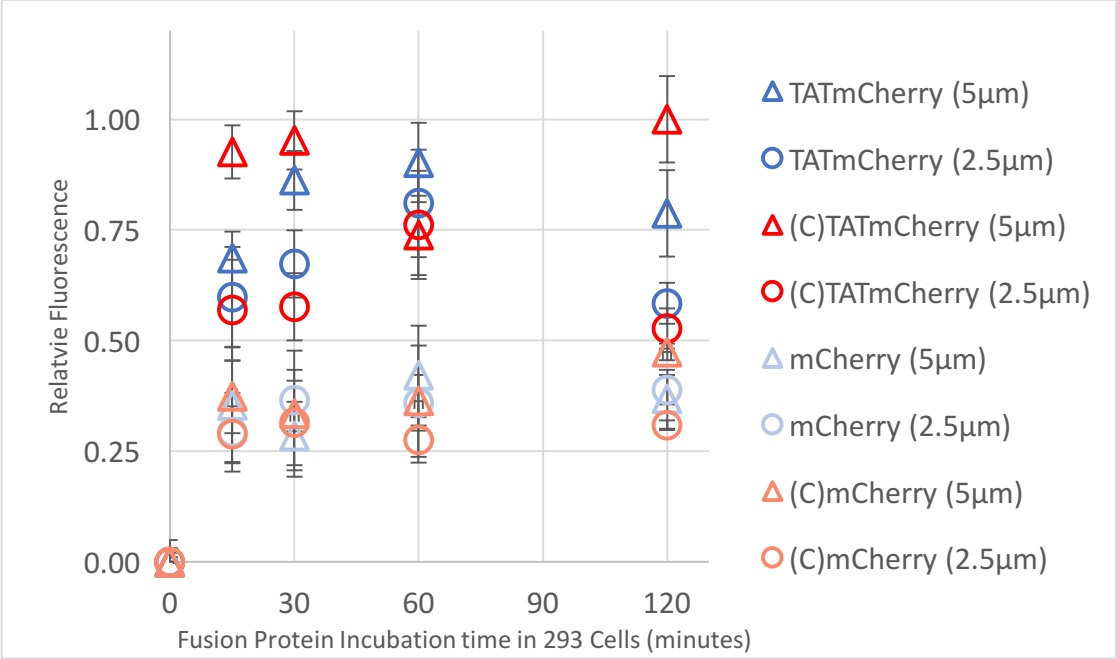


Figure 28. Transduction of HEK293 cells with 5µm and 2.5µm mCherry-fusion proteins. Cyclized (intein-based) and linear (wild-type) mCherry were used to measure the transduction capacity of the TAT-peptide. 5µm and 2.5µm of each protein was incubated with adherent 293 cells for 2 hours. Extracellular fluorescence was quenched with 10% total volume of 0.4% trypan blue and fluorescence was measured using 588nm excitation/509nm fluorescence. The results shown are are the average of three replicates with one standard deviation.

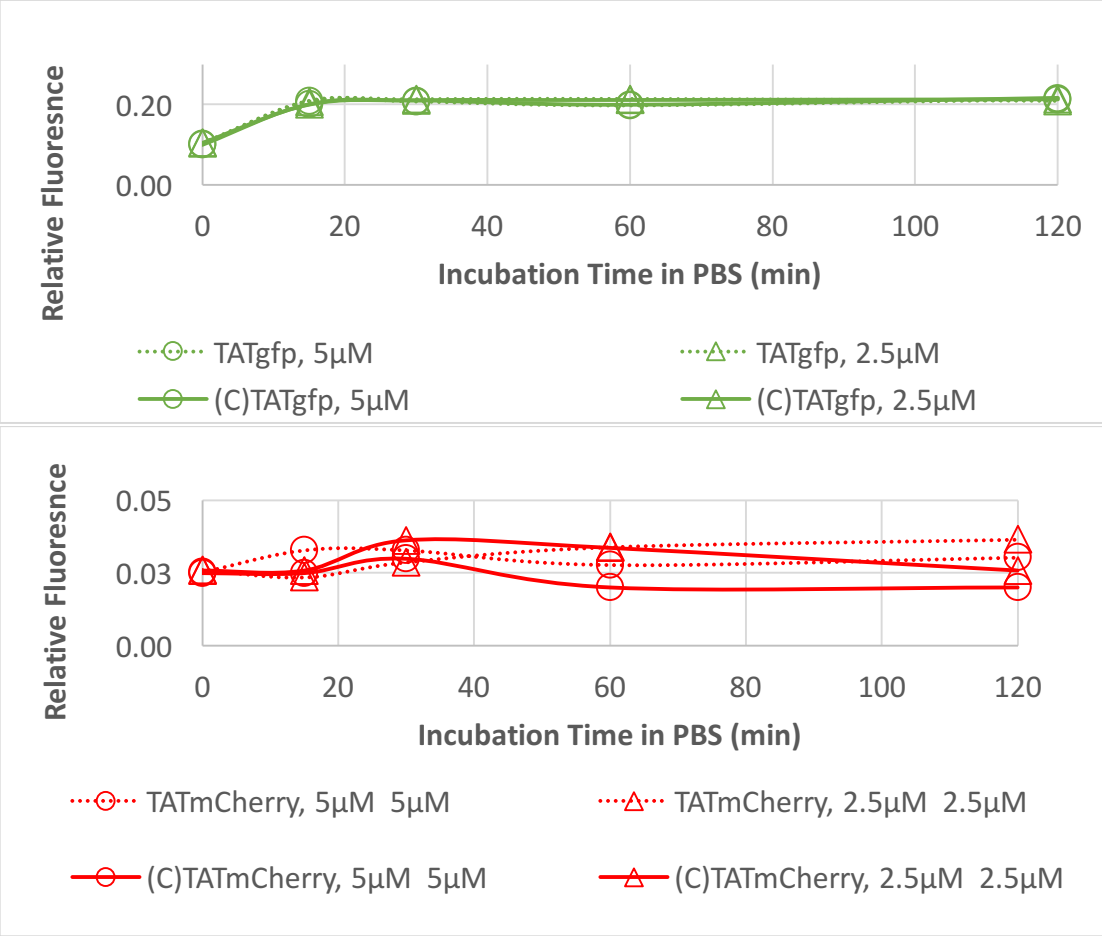


Figure 29. Relative Fluorescence of TAT-fusion proteins incubated in PBS. This figure measures the capacity of Trypan blue to quench extracellular fluorescence. In this figure, the fluorescence was quenched with 10% (v/v) of 0.4% Trypan blue of fluorometric reporter protein samples incubated in PBS.

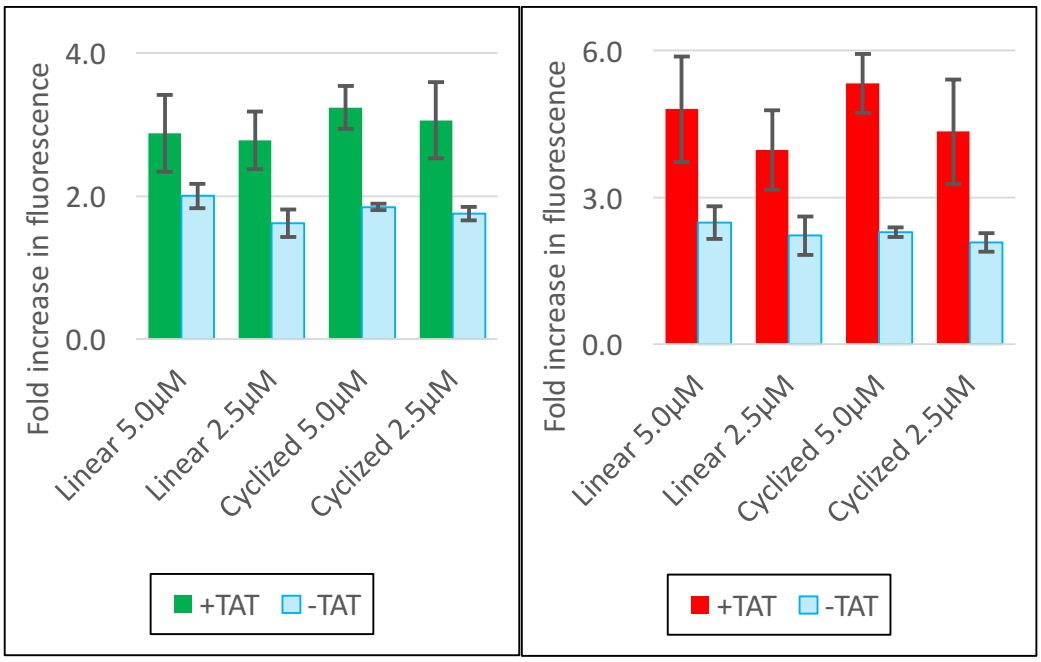


Figure 30. Transduction capacity of TAT-fluorescent protein fusions. 5μm and 2.5μm of the fluorescent protein fusions were incubated for 120 minutes with HEK293 cells. Extracellular fluorescence was quenched by addition of 10% volume 0.4% trypan blue. Intracellular fluorescence for eGFP (485ex/516em) and mCherry (580ex/610em) was measured for transduction of the TAT-fusion proteins. The results shown are are the average of three replicates with one standard deviation.

	Conc. (μm)	$k\left[\frac{1}{\text{min}}\right]$	Error_k	$\tau_{0.5}$ (min)
(C)TATeGFP	5.0	0.136	2.54E-02	5.09
(C)TATeGFP	2.5	0.182	7.69E-03	3.80
TATeGFP	5.0	0.132	1.44E-02	5.25
TATeGFP	2.5	0.113	1.46E-02	6.13
(C)eGFP	5.0	0.072	1.56E-02	9.62
(C)eGFP	2.5	0.143	7.64E-02	4.84
eGFP	5.0	0.179	5.19E-02	3.87
eGFP	2.5	0.868	5.66E-01	0.79
(C)TATmCherry	5.0	0.131	2.23E-02	6.13
(C)TATmCherry	2.5	0.113	6.04E-02	5.29
TATmCherry	5.0	0.127	1.73E+00	0.54
TATmCherry	2.5	0.149	8.74E-02	4.64
(C)mCherry	5.0	0.256	1.94E-01	2.70
(C)mCherry	2.5	0.101	9.83E-03	6.83
mCherry	5.0	0.171	1.02E-01	4.03
mCherry	2.5	0.847	4.43E-01	0.82

Table 13. First-order rate constant and calculated fluorescent half-life of eGFP and mCherry fluorometric assay. 5μm and 2.5μm of eGFP and mCherry fusion proteins were incubated alongside HEK293 cells. The extracellular fluorescence was modeled using exponential regression kinetics. The half-life is the amount of time required (minutes) to reach half of the maximum fluorescence signal.

	Conc. (µm)	F_{max} Experimental (±σ)	F_{max} Regression	±σ
(C)TATeGFP	5.0	$0.91 \pm 5.14x10^{-2}$	0.85	7.98E-02
(C)TATeGFP	2.5	$0.81 \pm 6.40x10^{-2}$	0.80	1.69E-02
TATeGFP	5.0	$1.00 \pm 5.94x10^{-2}$	0.95	5.23E-02
TATeGFP	2.5	$0.92 \pm 5.94x10^{-2}$	0.88	5.75E-02
(C)eGFP	5.0	$0.46 \pm 1.67x10^{-2}$	0.42	4.58E-02
(C)eGFP	2.5	$0.29 \pm 6.32x10^{-2}$	0.24	6.47E-02
eGFP	5.0	$0.41 \pm 7.84x10^{-2}$	0.37	5.41E-02
eGFP	2.5	$0.35 \pm 8.38x10^{-2}$	0.33	1.09E-01
(C)TATmCherry	5.0	$0.90 \pm 8.66x10^{-2}$	0.85	8.42E-02
(C)TATmCherry	2.5	$0.76 \pm 8.33x10^{-2}$	0.69	1.59E-01
TATmCherry	5.0	$1.00 \pm 1.78x10^{-2}$	0.90	6.15E-01
TATmCherry	2.5	$0.81 \pm 3.78x10^{-2}$	0.62	1.83E-01
(C)mCherry	5.0	$0.28 \pm 7.46x10^{-2}$	0.35	1.35E-01
(C)mCherry	2.5	$0.28 \pm 7.29x10^{-2}$	0.37	1.83E-02
mCherry	5.0	$0.28 \pm 2.05x10^{-2}$	0.39	1.17E-01
mCherry	2.5	$0.28 \pm 4.03x10^{-2}$	0.29	7.76E-02

Table 14. Maximum relative fluorescence of eGFP and mCherry Fluorometric Assay using experimental and regression analysis. 5µm and 2.5µm of eGFP and mCherry fusion proteins were incubated alongside HEK293 cells. The extracellular fluorescence was modeled using exponential regression kinetics.

4.5.2 Long-term *in vitro* Assessment of (C)TATeGFP

Incubation of the (C)TATeGFP protein with HEK293 cells over a 3-day period demonstrated the *in vitro* stability of the cyclized protein. After 24 and 48 hours, the lysate fraction of the cells identified presence of eGFP in the (C)TATeGFP samples and not in the (C)eGFP samples. The (C)TATeGFP samples at 72 hours could not identify the protein. There was a time-dependent decrease of eGFP in the lysate fraction of the (C)TATeGFP samples. The housekeeping protein, actin provided a loading control for all samples.

The *in vitro* transduction capacity of TATeGFP proteins in a cell culture environment has been previously studied. The detection of eGFP in the lysate fraction has been identified on an anti-GFP western blot after 48 hours [212] and 72 hours [213]. The results in this experiment are comparable to the linear eGFP models discussed in literature. Therefore, the cyclized eGFP protein retains the *in vitro* transduction capacity.

Although a western blot does not provide evidence of native-folding protein structure, the (C)TATeGFP protein retained transduction capacity for 48 hours. As discussed earlier, bioavailability is a critical parameter of therapeutic proteins. Upon administration, the human body's natural defense system efficiently degrades exogenous proteins. *In vivo* efficacy is heavily predicated on the half-life of a therapeutic protein. In this study, we have determined that a cyclic protein retains transduction potential for at least 48 hours *in vitro*.

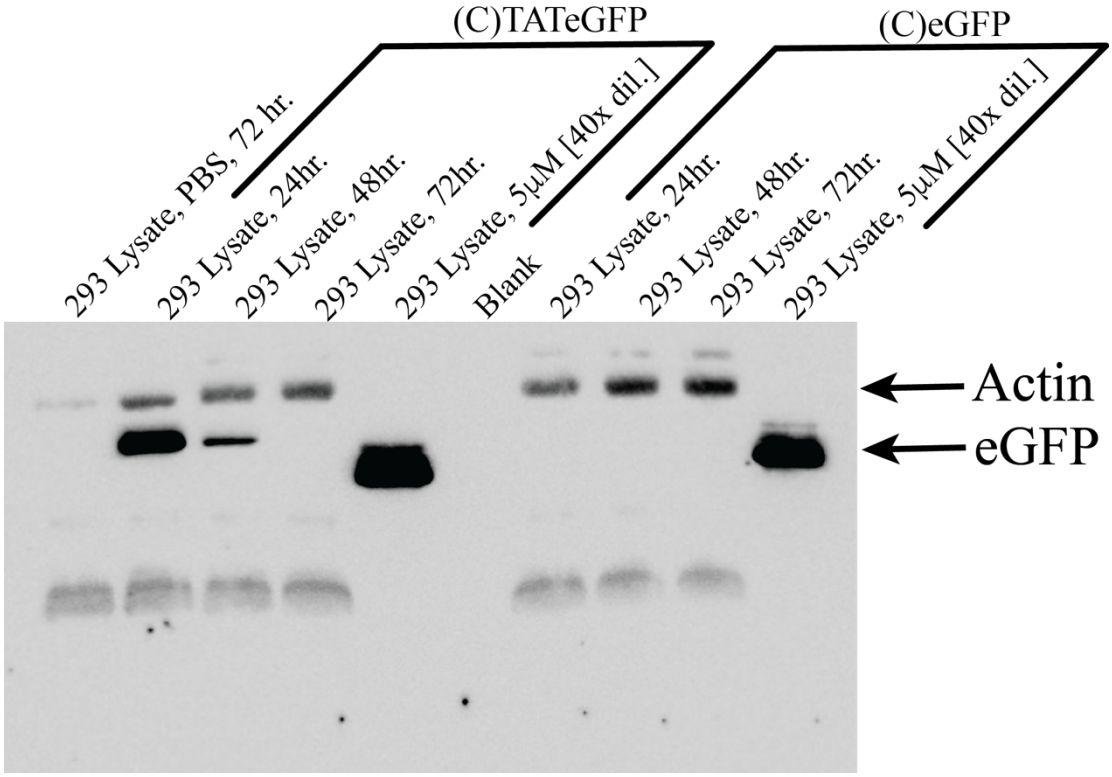


Figure 31. Anti-gfp and Anti-actin western blot of time-course samples of HEK293 lysates treated with (C)TATeGFP and (C)eGFP. 5μM of the purified protein was added to $5x10^5$ viable cells in a 6-well plate. After 24, 48, and 72 hours the cells were detached with trypsin and washed 3x with PBS. The lysate fraction was probed with anti-GFP and Anti-actin HRP-primary antibodies.

4.6 Conclusions

In this study, a cyclized, cell-penetrating peptide-protein system transduced the cell membrane and delivered a functional fluorometric reporter. The penetration of the macromolecule through the cell membrane exhibited low cytotoxicity. The kinetic rate and transduction efficiency of the cyclized fluorometric reporter was comparable to the linear-counterpart—a protein expressed without the intein-based system. The cell-penetrating peptide demonstrated robustness through delivery of two different fluorometric reporters, eGFP and mCherry, with comparable rate and efficiency. Finally, the enhanced stability of the cyclized fluorometric reporter, as discussed previously in Chapter 3, was orthogonally measured as the cyclical protein retained transduction capacity for at least 48 hours when co-cultured with mammalian cells.

There are two experimental discoveries in this chapter: the effective cellular uptake of a cyclized, fluorometric reporter, and the *in vitro* stability of the cyclized protein in 293 cells. As previously discussed in Chapter 3, the cyclized protein enhances the stability against protein degradation. This chapter demonstrates that inclusion of the TAT-peptide in the eGFP, intein fusion protein facilitates the cellular uptake, comparable to the linear counterpart. Transduction efficiency and, defined by the relative maximum fluorescence, and rate of transduction, measured by the fluorescence half-life were comparable to previous published results of the linear TATeGFP protein. An additional supporting piece of evidence that cyclized proteins retain cellular uptake capacity is the demonstration of *in vitro* transduction after 48 hours of incubation. There are no claims the eGFP protein retained the secondary

structure necessary for fluorescence, however, the fusion protein retained the capacity to transduce cells. The lysate fraction had measureable eGFP after 48 hours.

Chapter 5: Conclusions and Future Work

5.1 Conclusions

A unique cell-penetrating fluorometric reporter system has been engineered for enhanced stability while maintaining transduction potential. This study combined post-translational intein-splicing to form a cyclical protein with the cell transduction potential of the TAT-peptide. Fluorescent fusions retained native-state folding after transduction of the cell membrane. Orthogonal methods have been presented to confirm the cyclic, post-translational structure of the fluorometric reporters. The cyclical fluorometric reporters were resistant to exopeptidases and retained fluorescence at high concentrations of a chaotropic agent. The results of the experiment detail a fluorometric reporter system that has enhanced stability during protein expression, resistance to destabilization of hydrophobic interactions from chaotropic agents, resistance to exopeptidase digestion, and a rapid transduction across the cell membrane of eukaryotic cells.

A protein manufacturing process should produce a properly folded protein with high yields under a robust and reproducible platform. The protein expression phase sets the benchmark for overall yield of the process. It is critical to establish a production process that offers robustness in the relative stability of the protein. Protease degradation and thermal instability of exogenous proteins reduce yields and force execution of a harvest prior to the maximum yield of the process. A protein expressed as a cyclical structure creates an opportunity to increase yields during the protein expression phase. Increased yields in a protein manufacturing platform

reduce the manufacturing cost providing cost savings to the patient—the most critical stakeholder in the drug development process.

A cyclic, cell-penetrating, fluorometric reporter system could provide real-time tracking and identification of malignant tumors. It is unknown the full potential of the Maternin peptide. Through published results, it appears to be a peptide with immense potential. It is possible the peptide has selective function with cells expressing $CD133^{+}/CD44^{+}$ surface markers, uniquely identifying cancer cells. A fusion protein of Maternin, a cell-penetrating peptide and a fluorometric reporter could selectively identify, transduce and fluoresce cancer cells. As a diagnostic tool, the fusion protein could be orally ingested as the cyclical backbone could offer enhanced stability from the low pH environment and high concentration of proteases found in the digestive track. Upon entry into the bloodstream, the fluorometric reporter could be tracked with a non-invasive sensor as it circulates through the body. Upon positive identification, the fluorometric reporter penetrates the cell membrane and identifies the cancer cells. The system could also be engineered to remain in an endocytosis-induced vesicle which could enhance the residence time of the molecule ensuring proper identification and location.

5.2 Future Work

The future direction of this work is to study the integration of the Maternin peptide into a cyclical, cell-penetrating fluorometric reporter. The low efficacy of the cyclized Maternin-eGFP fusion protein led to the study of cell-penetrating peptides. To better understand the mechanisms of Maternin, the peptide could be inserted into

several locations on the eGFP backbone. The published crystal structures of eGFP illustrate exposed N-terminal and C-terminal amino acids. Therefore, additional sites could be employed of regions in the eGFP beta-barrel that are exposed—specifically on surface loop of eGFP: gfp157 or gfp172 [214]. A cyclic Maternin-eGFP could be a revolutionary diagnostic and therapeutic tool.

Appendices

Appendix A. Design, engineering, and description of plasmids

A.1. pET21a-Cint.14mer.NInt.6H

A pET21(a)+ vector with 564bp insert encoding nucleotide sequences for three target protein fragments: $Intein_C$, 14 amino acid peptide (Maternin), and $Intein_N$. Post-translational features of the translated sequence include an C-terminal His-tag (6 repeated residues) followed by a stop codon (TAA). Enzymatic digest sites included are:

Nde*I*	N-terminal
Hind*III*	Enclosed by $Intein_C$ and Maternin
EcoR*I*	Enclosed by Maternin and $Intein_N$
Xho*I*	C-terminal nucleotide sequence following stop codon (TAA)

Plasmid Construction:
The 564bp nucleotide sequence was synthesized by ThermoFisher (Carlsbad, CA) and inserted into a proprietary vector. The vector was digested with Nde*I* and Xho*I,* and agarose-gel purified. The digested fragment was ligated into a digested pET21(a)+ vector [Nde*I*/Xho*I*] and transformed into DH5-alpha competent cells (ThermoFisher). The competent cells were cultured in LB media containing 50μg/mL Ampicillin and plasmid was extracted from cell lysates. Purified plasmid was transformed into Rosetta(DE3) Competent Cells (Novagen).

Figure 32. 1.5% (w/v) Agarose gel separating a digested pET21(a)+ vector containing Cint.14mer.Nint.6H with NdeI / XhoI. Lane 1 is a 1kb DNA Ladder (Thermo) and Lane 2 is the digested plasmid. A band is observed between 500-750bp that indicates ligation of the of the 564bp target nucleotide sequence.

A.2 pET21a-Cint.14His.NInt.6H

A pET21(a)+ vector with 564bp insert encoding nucleotide sequences for three target protein fragments: $Intein_C$, 14 histidine residues, and $Intein_N$. Post-translational features of the translated sequence include an C-terminal His-tag (6 repeated residues) followed by a stop codon (TAA). Enzymatic digest sites included are:

Nde*I*	N-terminal
Hind*III*	Enclosed by $Intein_C$ and 14 histidine
EcoR*I*	Enclosed by 14 histidines and $Intein_N$
Xho*I*	C-terminal nucleotide sequence following stop codon (TAA)

Plasmid Construction:
The 564bp nucleotide sequence was synthesized by ThermoFisher (Carlsbad, CA) and inserted into a proprietary vector. The vector was digested with Nde*I* and Xho*I,* and agarose-gel purified. The digested fragment was ligated into a digested pET21(a)+ vector [Nde*I*/Xho*I*] and transformed into DH5-alpha competent cells (ThermoFisher). The competent cells were cultured in LB media containing 50μg/mL Ampicillin and plasmid was extracted from cell lysates. Purified plasmid was transformed into Rosetta(DE3) Competent Cells (Novagen).

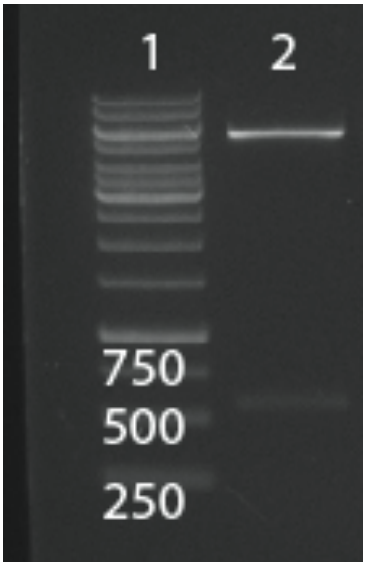


Figure 33. 1.5% (w/v) Agarose gel separating a digested pET21(a)+ vector containing Cint.14HIS.Nint.6H with NdeI / XhoI. Lane 1 is a 1kb DNA Ladder (Thermo) and Lane 2 is the digested plasmid. A band is observed between 500-750bp that indicates ligation of the 564bp target nucleotide sequence.

A.3. pET21a-Cint.gfp.NInt.6H

A pET21(a)+ vector with 1236bp insert encoding nucleotide sequences for three target protein fragments: $Intein_C$, egfp, and $Intein_N$. Post-translational features of the translated sequence include an C-terminal His-tag (6 repeated residues) followed by a stop codon (TAA). Enzymatic digest sites included are:

Nde*I*	N-terminal
Hind*III*	Enclosed by $Intein_C$ and egfp
EcoR*I*	Enclosed by egfp and $Intein_N$
Xho*I*	C-terminal nucleotide sequence following stop codon (TAA)

Plasmid Construction:
The 1236bp nucleotide sequence was synthesized by ThermoFisher (Carlsbad, CA) and inserted into a proprietary vector. The vector was digested with Nde*I* and Xho*I,* and agarose-gel purified. The digested fragment was ligated into a digested pET21(a)+ vector [Nde*I*/Xho*I*] and transformed into DH5-alpha competent cells (ThermoFisher). The competent cells were cultured in LB media containing 50μg/mL Ampicillin and plasmid was extracted from cell lysates. Purified plasmid was transformed into Rosetta(DE3) Competent Cells (Novagen).

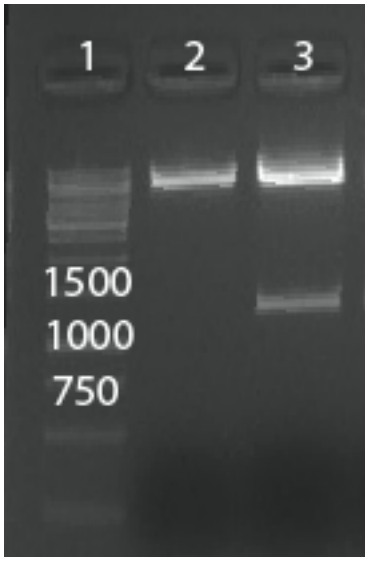


Figure 34. 1.5% (w/v) Agarose gel separating a digested pET21(a)+ vector containing Cint.gfp.Nint.6H with NdeI / XhoI. Lane 1 is a 1kb DNA Ladder (Thermo) and Lane 2 is a digested pET21(a)+ vector without the target insert. Lane 3 is the digested plasmid containing the target insert. A band is observed between 1000-1500bp that indicates ligation of the 1236bp target nucleotide sequence.

A.4. pET21a-Cint.egfp.6H.TEV.14mer.TEV.Nint

A pET21(a)+ vector with 1320bp insert encoding nucleotide sequences for three target protein fragments: $Intein_C$, egfp, and $Intein_N$. Post-translational features of the translated sequence include an N-terminal His-tag (6 repeated residues) at the C-terminus of the egfp nucleotide sequence. Then the 14mer peptide is flanked by TEV-protease recognition sites (ENLYFQS). A stop codon (TAA) follows the $Intein_N$ nucleotide sequence. Enzymatic digest sites included are:

Nde*I*	N-terminal
Hind*III*	Enclosed by $Intein_C$ and egfp
EcoR*I*	Enclosed by egfp and $Intein_N$
Xho*I*	C-terminal nucleotide sequence following stop codon (TAA)

Plasmid Construction:

A 938bp nucleotide sequence was synthesized by ThermoFisher (Carlsbad, CA) to encode the nucleotide sequence for translation of **egfp.6H.TEV.14mer.TEV**. At the N-terminus, five nucleotides (CGCCG) preceded the HindIII recognition site and following the EcoRI recognition site to allow for effective enzymatic digestion. The vector was digested with HindIII and EcoRI, and agarose-gel purified. The digested fragment was ligated into a digested **pET21a-Cint.14His.NInt.6H** vector [HindIII / EcoRI] and transformed into DH5-alpha competent cells (ThermoFisher). The competent cells were cultured in LB media containing 50μg/mL Ampicillin and plasmid was extracted from cell lysates. Purified plasmid was transformed into Rosetta(DE3) Competent Cells (Novagen).

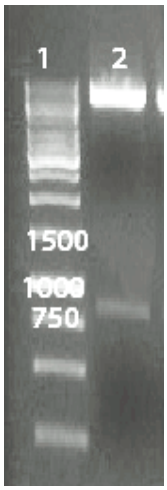


Figure 35. 1.5% (w/v) Agarose gel separating a digested pET21(a)+ vector containing Cint. gfp.6H.TEV.14mer.TEV.Nint with HindIII / EcoRI. Lane 1 is a 1kb DNA Ladder (Thermo) and Lane 2 is the digested plasmid. A band is observed between 750-1000bp that indicates ligation of the 934bp target nucleotide sequence.

A.5. pET21a-Cint.egfp.6H.TEV.14mer.TEV.NInt.AIDA1

A pET21(a)+ vector with 2817bp insert encoding nucleotide sequences for four target protein fragments: Intein$_C$, egfp, and Intein$_N$, and AIDA-1. Post-translational features of the translated sequence include **pET21a-Cint.egfp.6H.TEV.14mer.TEV.NInt** with a C-terminal AIDA-1 autotransporter linker sequence. A stop codon (TAA) follows the AIDA-1 nucleotide sequence. Enzymatic digest sites included are:

Nde*I*	N-terminal
Hind*III*	Enclosed by Intein$_C$ and egfp
EcoR*I*	Enclosed by egfp and Intein$_N$
BamH*I*	Enclosed by Intein$_N$ and AIDA-1
Xho*I*	C-terminal nucleotide sequence following stop codon (TAA)

Plasmid Construction:

The 1497bp nucleotide sequence encoding AIDA-1 was PCR amplified from a pET200 plasmid from J. Terrell (Forward CCGGGATTCGGTACCGTTAACA ATAATGG) and (Reverse CCGGAAGCTATATTTGATACCCAGTGCACCG CTAATGG.) The primers added a BamH*I* recognition site at the N-terminus and a stop codon (TAA) followed by Xho*I* at the C-terminus. The 1320bp **Cint.egfp.6H.TEV.14mer.TEV.NInt** insert was PCR amplified from **pET21a-Cint.egfp.6H.TEV.14mer.TEV.NInt** (Forward GCGCATATGGTGAAAGTTATTGG) and Reverse GCGTTTGATGGTGCCTGCATCCAGCAGCGG). The N-terminal Nde*I* was retained and the C-terminal amino acid sequence of InteinN was followed by a BamH*I* recognition site. The AIDA-1 PCR amplicon was digested with BamH*I* and Xho*I*, the **Cint.egfp.6H.TEV.14mer.TEV.NInt** amplicon was digested with Nde*I* and BamH*I* and ligated into a Nde*I* / Xho*I* digested pET21(a)+ vector and transformed into DH5-alpha competent cells (ThermoFisher). The competent cells were cultured in LB media containing 50µg/mL Ampicillin and plasmid was extracted from cell lysates. Purified plasmid was transformed into Rosetta(DE3) Competent Cells (Novagen).

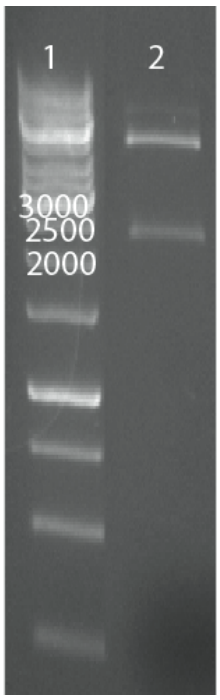


Figure 36. 1.5% (w/v) Agarose gel separating a digested pET21(a)+ vector containing pET21a-Cint.gfp.6H.TEV.14mer.TEV.NInt.AIDA1 with HindIII / XhoI. Lane 1 is a 1kb DNA Ladder (Thermo) and Lane 2 is the digested plasmid. A band is observed between 2500-3000bp that indicates ligation of the 2817bp sequence.

A.6. pET21a-SP.Cint.egfp.6H.TEV.14mer.TEV.NInt.AIDA1 and pET21a-Cint.egfp.6H.TEV.14mer.TEV.NInt.AIDA1.

A pET21(a)+ vector with 2973bp insert encoding nucleotide sequences for four target protein fragments: $Intein_C$, egfp, and $Intein_N$, and AIDA-1. Post-translational features of the translated sequence include an N-terminal AIDA-1 signal peptide, and **pET21a-Cint.egfp.6H.TEV.14mer.TEV.NInt.AIDA1.** Enzymatic digest sites included are:

Nde*I*	N-terminal
Sal*I*	Enclosed by Signal Peptide and $Intein_C$
Hind*III*	Enclosed by $Intein_C$ and egfp
EcoR*I*	Enclosed by egfp and $Intein_N$
BamH*I*	Enclosed by $Intein_N$ and AIDA-1
Xho*I*	C-terminal nucleotide sequence following stop codon (TAA)

Plasmid Construction:

The 150bp nucleotide sequence encoding the AIDA-1 signal peptide was PCR amplified from a pET200 plasmid from J. Terrell (Forward CCGCATATGAACAAAGCCTACAGCATCATTTGG) and (Reverse

CCGGTCGACTGCAAATGCATTACCAATGG.) The primers added a Nde*I* recognition site at the N-terminus and a Sal*I* at the C-terminus. The 2817bp **Cint.egfp.6H.TEV.14mer.TEV.NInt .AIDA1** insert was PCR amplified from **pET21a- SP.Cint.egfp.6H.TEV.14mer.TEV.NInt.AIDA1** (Forward GCGAGTCGACATGGTTAAAGTTATTGG) and Reverse GCGTTTGATGGTGCCTGCATCCAGCAGCGG). An N-terminal Sal*I* was inserted and the C-terminal amino acid sequence was maintained. The Signal Peptide PCR amplicon was digested with Nde*I* and Sal*I*, the **Cint.egfp.6H.TEV.14mer.TEV.NInt.AIDA1** amplicon was digested with Sal*I* and Xho*I* and ligated into a Nde*I* / Xho*I* digested pET21(a)+ vector and transformed into DH5-alpha competent cells (ThermoFisher). The competent cells were cultured in LB media containing 50µg/mL Ampicillin and plasmid was extracted from cell lysates. Purified plasmid was transformed into Rosetta(DE3) Competent Cells (Novagen).

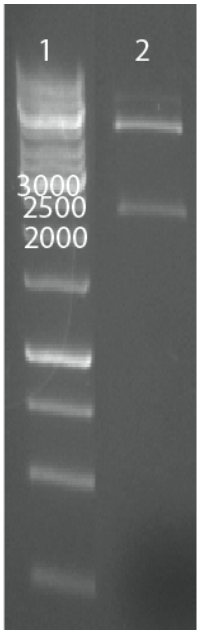


Figure 37. 1.5% (w/v) Agarose gel separating a digested pET21(a)+ vector containing pET21a-Cint.gfp.6H.TEV.14mer.TEV.NInt.AIDA1 with HindIII / XhoI. Lane 1 is a 1kb DNA Ladder (Thermo) and Lane 2 is the digested plasmid. A band is observed between 2500-3000bp that indicates ligation of the 2817bp sequence.

A.7. pET21a-Cint.egfp.Thr.6H.Nint, pET21a-Cint.egfpnoc.Thr.6H.Nint (C49G and C71G), and pET21a-Cint.egfpind.Thr.6H.Nint (C49G, V69P, Q70L, F72L)

A pET21(a)+ vector with 1236bp insert encoding nucleotide sequences for three target protein fragments: $Intein_C$, egfp, and $Intein_N$. Post-translational features of the translated sequence include an C-terminal Thrombin recognition site, a His-tag (6 repeated residues). A stop codon (TAA) follows the $Intein_N$ nucleotide sequence. Enzymatic digest sites included are:

Nde*I*	N-terminal
Hind*III*	Enclosed by $Intein_C$ and egfp
EcoR*I*	Enclosed by egfp and $Intein_N$
Xho*I*	C-terminal nucleotide sequence following stop codon (TAA)

Plasmid Construction:
414bp nucleotide sequences was synthesized by ThermoFisher (Carlsbad, CA) and inserted into a proprietary vector. The sequences were flanked by enzymatic recognition sites HindIII and Cla*I* at the N-terminus and C-terminus respectively. The vector was digested with HindIII and Cla*I,* and agarose-gel purified. The digested fragment was ligated into a digested pET21a-Cint.egfp.6H.TEV.14mer.TEV.NInt
vector [HindIII/Cla*I*] and transformed into DH5-alpha competent cells (ThermoFisher). Sequences were confirmed by nucleotide sequencing using T7promoter primer (MacrogenUSA)

Figure 38. **1.5% (w/v) Agarose gel separating a digested pET21(a)+ vector containing pET21a-Cint.gfp.Thr.6H.NInt; pET21a-Cint.gfpnoc.Thr.6H.NInt*;* pET21a-Cint.gfpind.Thr.6H.NInt with Nde*I* / Xho*I*.** Lane 1 is a 1kb DNA Ladder (Thermo) and Lane 2 is a digested pET21a-Cint.gfp.Thr.6H.NInt; Lane 3 is digested pET21a-Cint.gfpnoc.Thr.6H.NInt*;* Lane 4 is digested pET21a-Cint.gfpind.Thr.6H.NInt. All vectors were digested with Nde*I* and Xho*I*. Bands are observed between 1000-1500bp that indicates ligation of the 1236 target nucleotide sequence.

Appendix B. Clonogenic Assay Raw Data.

	Compound Name-Concentration	Violet Stained Colonies		Plating Efficiency		Survival Fraction	
		Mean	**St. Dev**	**PE**	**St. Dev**	**SF**	**St Dev.**
HTB-123 Cell line	(C)eGFP-1ug/ml	164	4.62	82.0%	2.31E-02	63.52%	7.13E-02
	(C)eGFP-5ug/ml	165	4.16	82.5%	2.08E-02	63.90%	7.06E-02
	(C)eGFP-10ug/ml	162	2.16	81.0%	1.08E-02	62.74%	6.84E-02
	(C)Maternin-eGFP-1ug/ml	164.5	3.11	82.3%	1.55E-02	63.71%	6.93E-02
	(C)Maternin-eGFP-5ug/ml	162	1.41	81.0%	7.07E-03	62.74%	6.79E-02
	(C)Maternin-eGFP-10ug/ml	167.75	2.22	83.9%	1.11E-02	64.97%	6.84E-02
	Synthetic Maternin-1ug/ml	41.75	2.06	20.9%	1.03E-02	16.17%	6.83E-02
	Synthetic Maternin-5ug/ml	32	2.16	16.0%	1.08E-02	12.39%	6.84E-02
	Synthetic Maternin-10ug/ml	21.75	1.71	10.9%	8.54E-03	8.42%	6.80E-02
	Taxol-1ug/ml	28.5	1.29	14.3%	6.45E-03	11.04%	6.78E-02
	Taxol - 5ug/ml	13.5	1.29	6.75%	6.45E-03	5.23%	6.78E-02
	Taxol-10ug/ml	8.75	1.71	4.38%	8.54E-03	3.39%	6.80E-02
	Control	181	2.16	90.5%	1.08E-02	70.10%	6.84E-02
	Compound Name-Concentration	**Violet Stained Colonies**		**Plating Efficiency**		**Survival Fraction**	
		Mean	**St. Dev**	**PE**	**St. Dev**	**SF**	**St Dev.**
PC-3 Cell line	(C)eGFP-1ug/ml	136	2.63	67.9%	1.31E-02	92.66%	2.61E-02
	(C)eGFP-5ug/ml	137	4.03	68.4%	2.02E-02	93.34%	3.02E-02
	(C)eGFP-10ug/ml	136	2.16	68.0%	1.08E-02	92.83%	2.50E-02
	(C)Maternin-eGFP-1ug/ml	137	2.06	68.4%	1.03E-02	93.34%	2.48E-02
	(C)Maternin-eGFP-5ug/ml	133	3.20	66.6%	1.60E-02	90.96%	2.77E-02
	(C)Maternin-eGFP-10ug/ml	135	4.76	67.5%	2.38E-02	92.15%	3.28E-02
	Synthetic Maternin-10ug/ml	20.8	0.957	10.4%	4.79E-03	14.16%	2.30E-02
	Synthetic Maternin-5ug/ml	16.3	1.50	8.13%	7.50E-03	11.09%	2.38E-02
	Synthetic Maternin-1ug/ml	8.50	1.29	4.25%	6.45E-03	5.80%	2.35E-02

	Compound Name-Concentration	Violet Stained Colonies		Plating Efficiency		Survival Fraction	
		Mean	**St. Dev**	**PE**	**St. Dev**	**SF**	**St Dev.**
	Taxol-1ug/ml	17.5	2.08	8.75%	1.04E-02	11.95%	2.48E-02
	Taxol - 5ug/ml	14.8	1.71	7.38%	8.54E-03	10.07%	2.41E-02
	Taxol-10ug/ml	8.50	1.29	4.25%	6.45E-03	5.80%	2.35E-02
	Control	147	4.51	73.3%	2.25E-02	100%	3.19E-02
	Compound Name-Concentration	**Violet Stained Colonies**		**Plating Efficiency**		**Survival Fraction**	
		Mean	**St. Dev**	**PE**	**St. Dev**	**SF**	**St Dev.**
KSY-1 Cell line	(C)eGFP-1ug/ml	114	3.40	57.1%	1.70E-02	86.88%	2.23E-02
	(C)eGFP-5ug/ml	113	2.63	56.6%	1.31E-02	86.12%	1.95E-02
	(C)eGFP-10ug/ml	115	4.80	57.3%	2.40E-02	87.07%	2.80E-02
	(C)Maternin-eGFP-1ug/ml	115	3.50	57.6%	1.75E-02	87.64%	2.27E-02
	(C)Maternin-eGFP-5ug/ml	116	2.63	57.9%	1.31E-02	88.02%	1.95E-02
	(C)Maternin-eGFP-10ug/ml	115	2.94	57.5%	1.47E-02	87.45%	2.06E-02
	Synthetic Maternin-1ug/ml	38.0	1.83	19.0%	9.13E-03	28.90%	1.71E-02
	Synthetic Maternin-5ug/ml	20.5	1.29	10.3%	6.45E-03	15.59%	1.58E-02
	Synthetic Maternin-10ug/ml	13.0	1.83	6.50%	9.13E-03	9.89%	1.71E-02
	Taxol-1ug/ml	32.5	2.38	16.3%	1.19E-02	24.71%	1.87E-02
	Taxol - 5ug/ml	22.0	2.94	11.0%	1.47E-02	16.73%	2.06E-02
	Taxol-10ug/ml	12.3	1.71	6.13%	8.54E-03	9.32%	1.68E-02
	Control	132	2.89	65.8%	1.44E-02	100%	2.04E-02

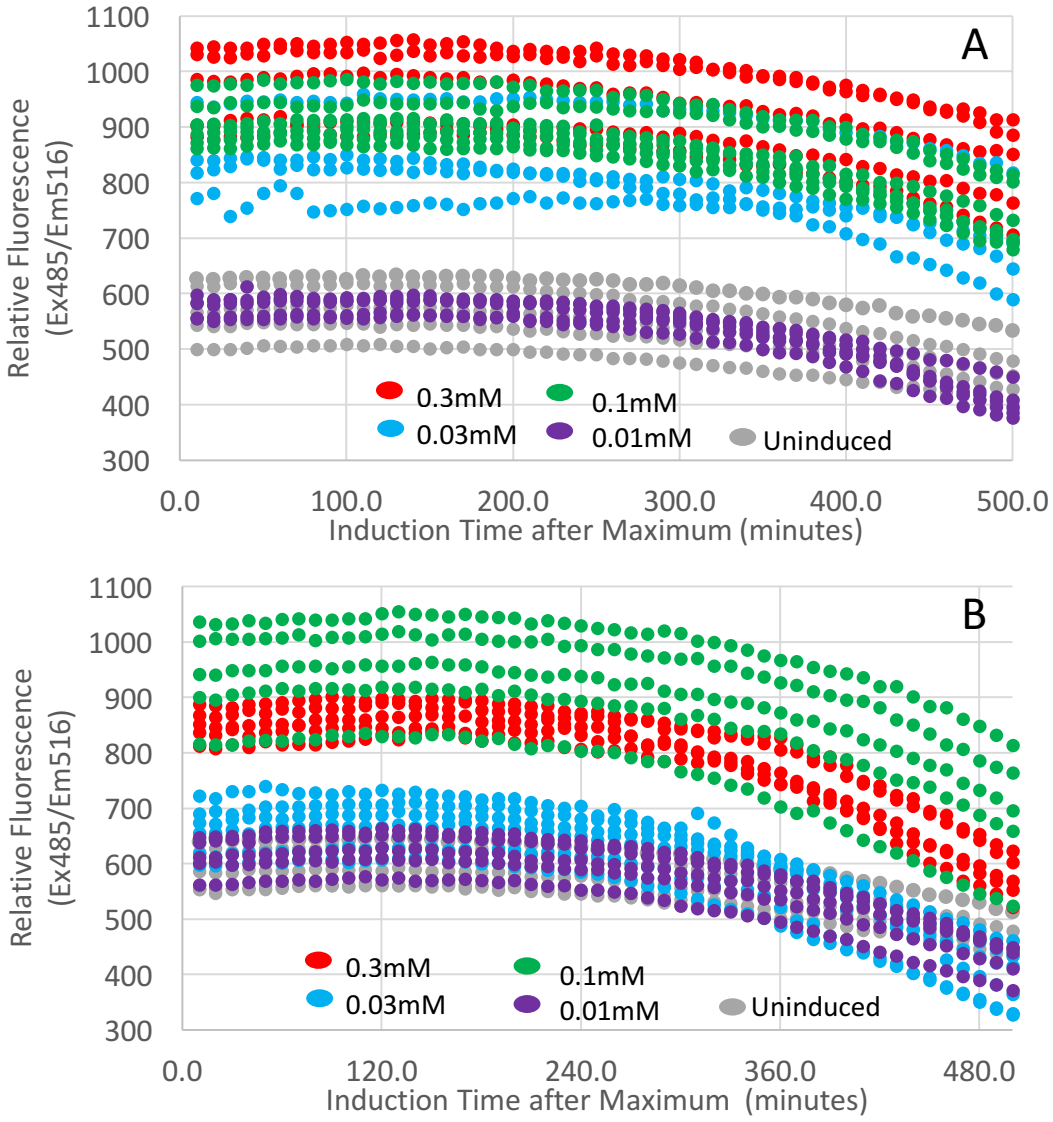
Table 15. Plating Effeciency and Survival Fraction of Compounds tested on HTB-123, PC-3, and KSY-1 Cells. Control (PBS), Taxol, Synthetic Maternin, (C)Maternin-eGFP, and (C)eGFP were incubated at 10μg/mL, 5μg/mL, 1μg/mL for 7 days. Colonies (>25 cells) were scored against control as reference.

Appendix C. Fluorescence and Optical Density Measurements during Protein Expression of eGFP, TATeGFP, and (C)TATeGFP

C.1 Maximum Fluorescence and Equilibrium Unfolding Kinetics Data

Construct	IPTG conc.	Relative Fluorescence Units [RFU] (485ex/516em)					Avg	St. Dev
		1	**2**	**3**	**4**	**5**		
eGFP	**N/A**	740	680	810	720	800	**750**	**54.8**
	0.3mM	890	1100	980	920	860	**950**	**94.9**
	0.1mM	840	830	910	870	930	**876**	**43.4**
	0.03mM	870	920	850	840	960	**888**	**50.7**
	0.01mM	650	650	680	740	710	**686**	**39.1**
TATeGFP	**N/A**	840	790	860	770	790	**810**	**38.1**
	0.3mM	940	960	870	970	870	**922**	**48.7**
	0.1mM	800	1010	1000	830	880	**902**	**93.9**
	0.03mM	660	680	710	780	800	**726**	**61.5**
	0.01mM	730	760	770	770	750	**756**	**16.7**
(C)TATeGFP	**N/A**	620	600	660	660	840	**676**	**95.3**
	0.3mM	920	870	860	990	1030	**934**	**74.4**
	0.1mM	930	770	940	880	870	**878**	**67.6**
	0.03mM	850	780	890	750	680	**790**	**82.8**
	0.01mM	520	640	650	730	590	**626**	**77.7**

Table 16. Maximum fluorescence of eGFP, TATeGFP, and (C)TATeGFP constructs during induction over half-log increments of 0.01mM, 0.03mM, 0.1mM, and 0.3mM IPTG. Fluorescence was measured every 10 minutes and the maximum signal was determined for the three eGFP variants over four IPTG concentrations.


A
Relative Fluorescence (Ex485/Em516)
1100
1000
900
800
700
600
500
400
300
0.0
100.0
200.0
300.0
400.0
500.0
Induction Time after Maximum (minutes)
0.3mM
0.1mM
0.03mM
0.01mM
Uninduced
B
Relative Fluorescence (Ex485/Em516)
1100
1000
900
800
700
600
500
400
300
0.0
120.0
240.0
360.0
480.0
Induction Time after Maximum (minutes)
0.3mM
0.1mM
0.03mM
0.01mM
Uninduced

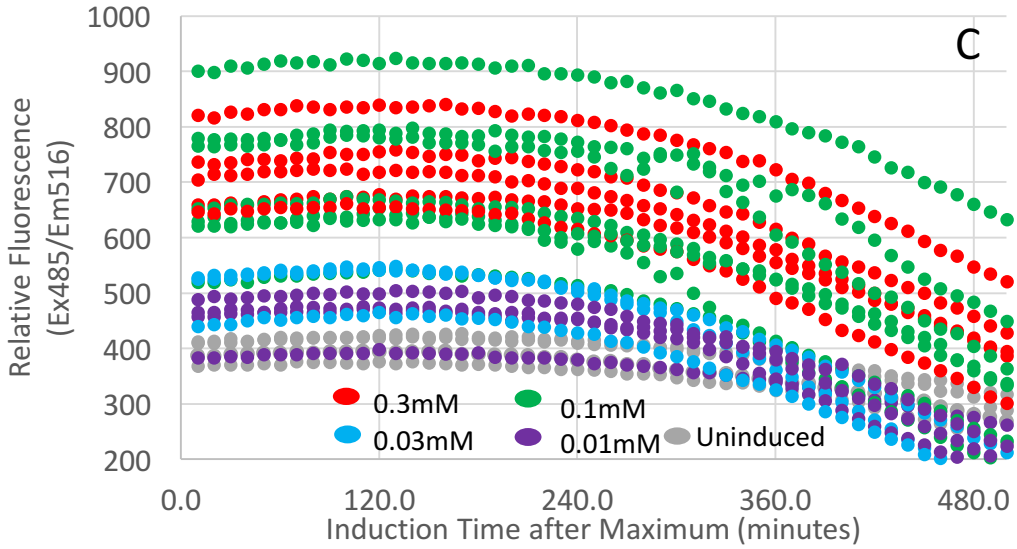


Figure 39. Relative Fluorescence after maximum during induction of eGFP, TATeGFP, and (C)TATeGFP using half-log concentrations of 0.01mM, 0.03mM, 0.1mM, and 0.3mM IPTG. (C)TATeGFP (Frame A), TATeGFP (Frame B), and eGFP (Frame C) all experience higher expression as the IPTG concentration is increased from 0.01mM, 0.03mM, 0.1mM, and 0.3mM IPTG. After the maximum flourescence is observed during induction, equilibrium exists between protein expression and proper folding of eGFP and the degredation that is induced from intracellular and extraceullar factors.

C.2 Exponential Regression Analysis

For regression analysis of exponential functions, Microsoft Excel's Solver function was used to minimize the sum of squares. The standard error of the exponential function coefficients were calculated from Bayesian probability theory [215]:

Standard Error of rate constant:

$$\Delta k = \frac{\sigma}{F_{max}}\sqrt{8k^3\Delta t}\text{ , t~time scale of measurement} \quad \text{Equation 28}$$

Standard Error of Maximum Fluorescence:

$$\Delta F_{max} = \sigma\sqrt{2k\Delta t} \quad \text{Equation 29}$$

Appendix D. Effect of divalent ions as a catalyst for post-translational intein processing

Homing endonucleases require a co-factor for efficient cleavage and site-specificity of supercoiled DNA [216]. Intein endonucleases require a divalent ion as a cofactor. The intein endonuclease from the DNA polymerase enzyme of *Thermococcus fumicolans* is 5-10 fold more active in the presence of Mn^{2+} over Mg^{2+} [217]. Intein-induced cleavage of proteins is most efficient under reducing conditions [218] and strongly inhibited by divalent ions [219].

D.1. Materials and Methods

Plasmid pET21a-Cint.egfp.6H.TEV.14mer.TEV.Nint [**A.4. pET21a-Cint.egfp.6H.TEV.14mer.TEV.Nint]** was used to express (C)eGFP as described in [**2.3.1 Protein Expression].** The frozen cell pellet was fractioned into two aliquots. One aliquot was resuspended in IMAC Load/Lyse buffer, pH 6.5 and the second in IMAC Load/Lyse buffer, pH 8.5 **[G.2 IMAC Column Buffers]** The resuspended cell pellets were lysed as described in **[2.3.2 IMAC Purification]** and the clarified lysate fraction was separated into ten, 5mL fractions. Fractions (n=3) were supplemented with the following a reducing agent and/or divalent cation as described in **Table 17**.

Lysate Fraction	Reducing Agent	Divalent Cation
1	1mM TCEP	25mM Ca^{2+}
2	50mM DTT	25mM Ca^{2+}
3	---	25mM Ca^{2+}
4	1mM TCEP	25mM Mg^{2+}
5	50mM DTT	25mM Mg^{2+}
6	---	25mM Mg^{2+}
7	1mM TCEP	1mM Zn^{2+}
8	50mM DTT	1mM Zn^{2+}
9	---	1mM Zn^{2+}
10	---	---

Table 17. Divalent ion and reducing agent supplimentation of (C)eGFP lysate fractions for on-column intein processing. The 10 fractions (n=3) were purified using an IMAC column to assess the intein processing effeciency.

The 5mL supplemented fractions were loaded onto charged IMAC columns and washed once with IMAC wash buffer. The columns were stored at 4°C for 16 hours. The columns were eluted with IMAC elution buffer. The elution fraction was analyzed using a Biotek Neo2 for absorbance at 280nm (total protein) and 475nm (eGFP fluorescence emission maximum). The elution fractions were loaded onto a 4-15% reducing SDS-PAGE gel (Biorad) and separated at 200V for 45 minutes. The gel was stained in 0.1% coomassie blue and de-stained with ethanol/acetic acid. Band densitometry analysis was performed on four key intein process fragments: unprocessed intein (50kDa), eGFP-C_{intein} (34kDa), eGFP (30kDa), and N_{intein} (14kDa).

To quantify the efficiency of intein processing using orthogonal methods, the optical density and band densitometry were used to compare reducing agents, divalent ions, and pH conditions. For optical density, the intein processing efficiency was measured by normalizing eGFP fluorescence by:

$$Normalized\,\frac{A_{475}}{A_{280}} = \frac{\frac{A_{475}}{A_{280}}}{\left[\frac{A_{475}}{A_{280}}\right]_{max}} \qquad \text{Equation 30}$$

For band densitometry, a ratio was calculated to quantify the intein processing efficiency:

$$\%\ of\ total\ intein\ fragments = \frac{A_i}{\sum_i A_i} \qquad \text{Equation 31}$$

D.2. Results and Analysis

IMAC purified samples of (C)eGFP expressed protein lysates, supplemented with reducing agents and divalent ions, over acidic and basic conditions, exhibited differences in the efficiency that a translated intein sequence is processed. As discussed in the literature review and the introduction section to Appendix C, intein processing has referenced protocols including divalent ions and or reducing agents to enhance the efficiency that a translated intein sequence processes the final intein fragments. For this experiment, the goal was to enhance the processing of the intein, resulting in a cyclized-eGFP protein. Moreover, results of this experiment would outline the most efficient method for protein production and purification for subsequent intein-related analysis.

After purification of the lysate fractions, the first method used to evaluate intein processing was the relative fluorescence of the elution samples. As illustrated in **2.3.4 Size Exclusion Purification of eGFP-variants**, the relative fluorescence of unprocessed intein fragments is minimal. Therefore, if total protein can be quantified by absorbance at 280nm and eGFP fluorescence has a maximum absorbance at 475nm, a ratio of these two wave lengths will correlate to relative concentration of processed, cyclized eGFP. In **Figure 40**, there is no significant difference of intein

processing under acidic (pH 6.5) or basic (pH 8.5) conditions. However, there is a significant difference upon supplementation of divalent ions and reducing agents. Based on the data in the figure, the optimal condition is supplementing with 1mM DTT, and 25mM $CaCl_2$. pH 6.5. Intein processing efficiency is reduced when supplementing with Zn^{2+}, and is reversible by supplementing with a reducing agent.

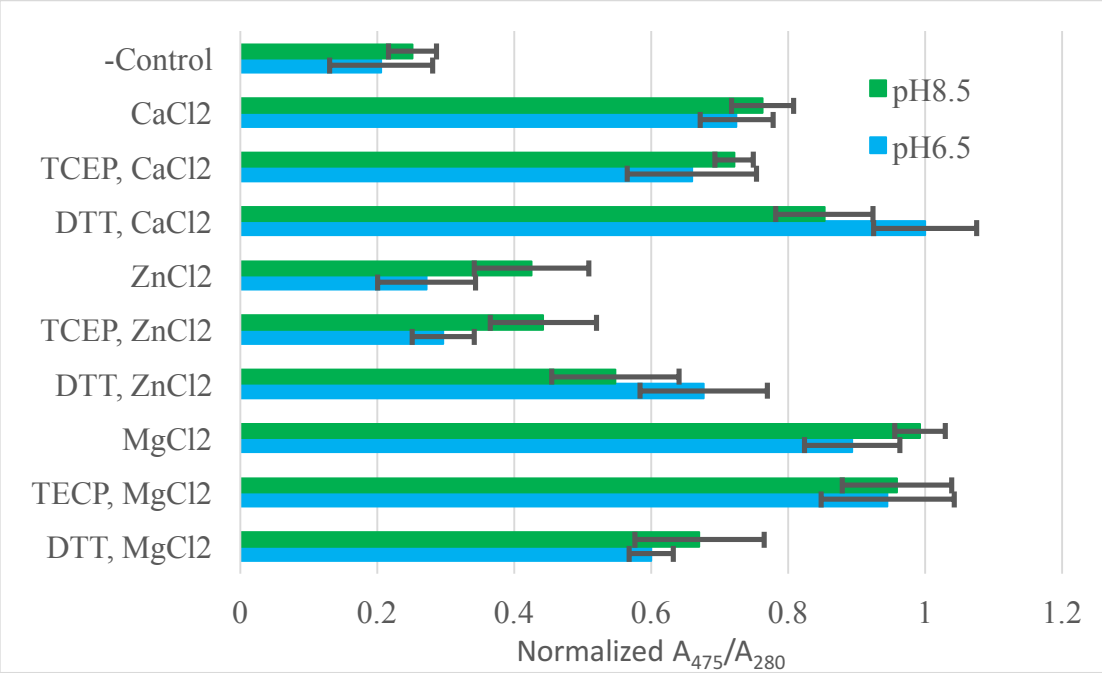


Figure 40. Divalent and reducing agent analysis of on-column intein processing. (C)eGFP cell pellets (n=3) were lysed using one of two reducing agents (1mM TCEP or 50mM DTT), one of three divalent ions (25mM $CaCl_2$, 25mM $MgCl_2$, or 1mM $ZnCl_2$), or 2 pH conditions (pH 6.5 or pH 8.5), and a negative control and purified using an IMAC column. After the binding step, the column was washed and then incubated at 4°C for 16 hours before washing and eluting the sample. The elution samples were measured for absorbance at 280nm (total protein) and 475nm (eGFP). For the graph, the absorbance at 475 was normalized by 280nm. The results shown are are the average of three replicates with one standard deviation.

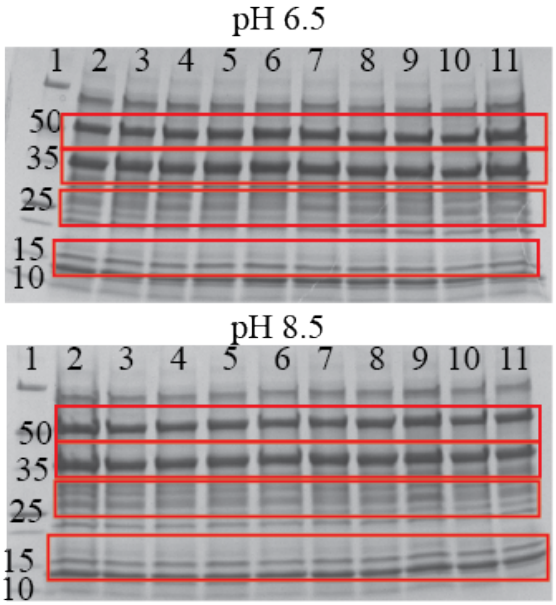
From **Figure 3**, the processing of the Lactam Intein results in two products: a N_{intein} fragment and a eGFP-C_{intein} fragment. Therefore, if the process were inhibited at this step, these two fragments should be formed in relatively equal molar quantities.

When the lactam product is further processed, it results in two products, a C_{intein} fragment and a cyclized eGFP protein. Therefore, upon full intein processing, the molar ratio of N_{intein} to eGFP-C_{intein} should be >1. This is difficult to correlate in **Figure 42** as the N_{intein} fragment does not contain a His-tag. Any lactam processing prior to this experiment would produce an N_{intein} fragment that would collect in the column flowthrough or wash sample. Analysis of molar ratios of the N_{intein} to eGFP-C_{intein} would not be accurate unless the flowthrough and wash fractions were included in the molar balance.

While it is difficult to observe intein processing efficiency from the SDS-PAGE gel illustrated in **Figure 41**, the band densitometry analysis does provide basis for supporting hypothesis that Zn^{2+} is inhibitory to intein processing. A relative intein efficiency can be calculated to compare the samples tested in this experiment. A dense unprocessed intein protein band with respect to a lower absorbance at 280nm would indicate a high percentage of unprocessed inteins. Therefore, if Zn^{2+} were inhibitory, it would have the highest ratio. If we consider it an efficiency ratio, then it can be calculated as illustrated in

$$Relative\ Intein\ Effeciency = 1 - \frac{\frac{A_i}{A_{280}}}{\left[\frac{A_i}{A_{280}}\right]_{max}} \qquad \text{Equation 32}$$

Based on this ratio, Zn^{2+} is inhibitory to intein processing, but can be reversed with supplementing DTT.



Lane	Sample
1	Ladder
2	DTT, $CaCl_2$
3	TCEP, $CaCl_2$
4	$CaCl_2$
5	DTT, $ZnCl_2$
6	TCEP, $ZnCl_2$
7	$ZnCl_2$
8	DTT, $MgCl_2$
9	TCEP, $MgCl_2$
10	$MgCl_2$
11	(-) control

Figure 41. Band densitometry analysis using reducing SDS-PAGE gel of intein fragments. Elution samples were separated using a reducing-protein gel and the four intein fragments: unprocessed intein (50kDa), eGFP-Cintein (34kDa), eGFP (30kDa), and Nintein (14kDa), were analyzed using band densitometry analysis on a Biorad Chemidoc Imager.

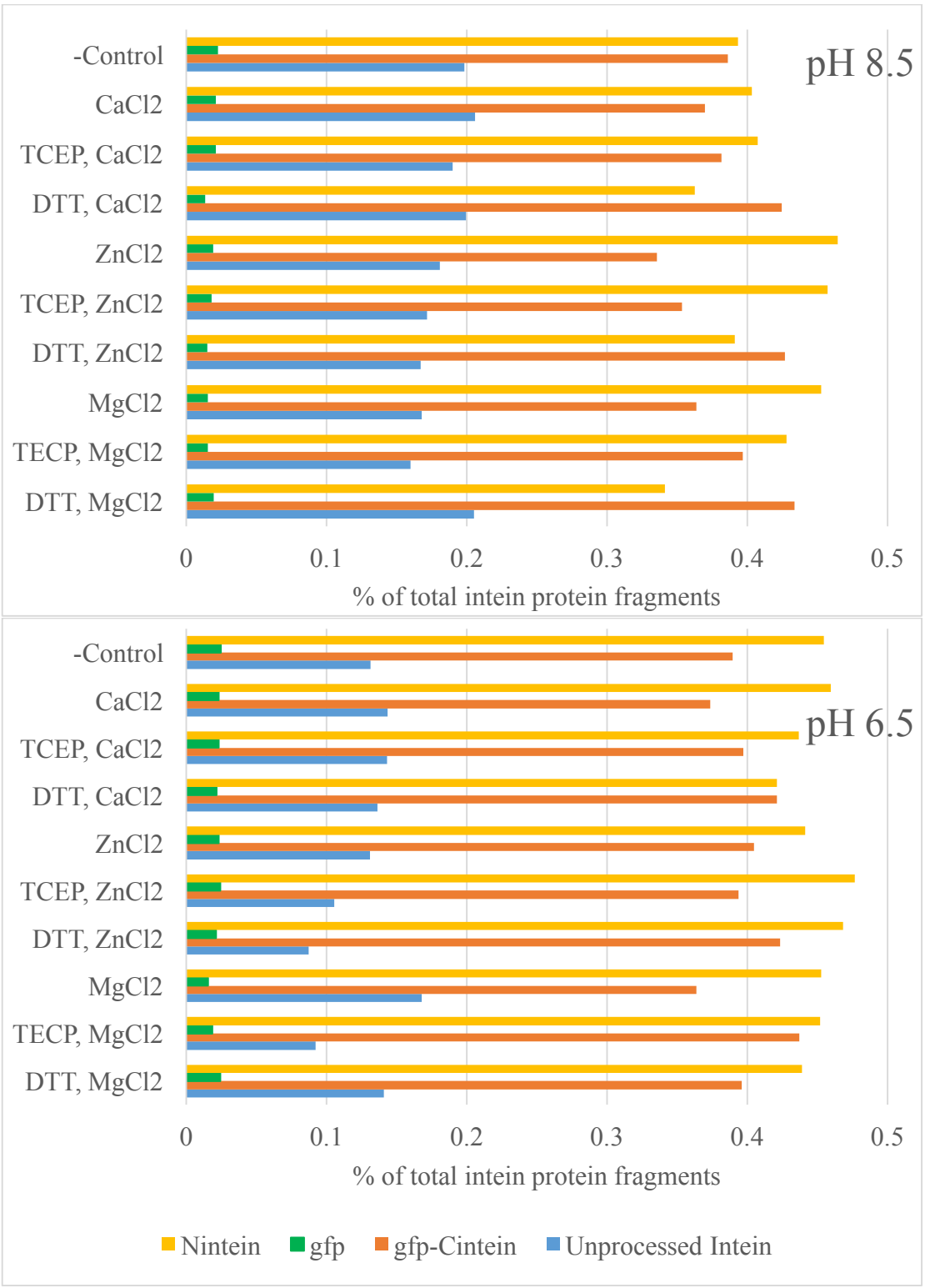


Figure 42. Intein process of on-column samples using reducing agents, divalent ions, and pH. Elution samples of overnight-bound intein samples supplimented with one of two reducing agents (1mM TCEP or 50mM DTT), one of three divalent ions

(25mM $CaCl_2$, 25mM $MgCl_2$, or 1mM $ZnCl_2$), or 2 pH conditions (pH 6.5 or pH 8.5), and a negative control and purified using an IMAC column. The elution sample was separated using a reducing SDS-PAGE gel and stained with coomassie blue. The stained gel (**Figure 41**) was analyzed for protein band area using a Biorad Chemidoc Imaging System. The four bands analysed: Unprocessed Intein (50kDa), eGFP-Cintein (34kDa), eGFP (30kDa), and Nintein (14kDa).

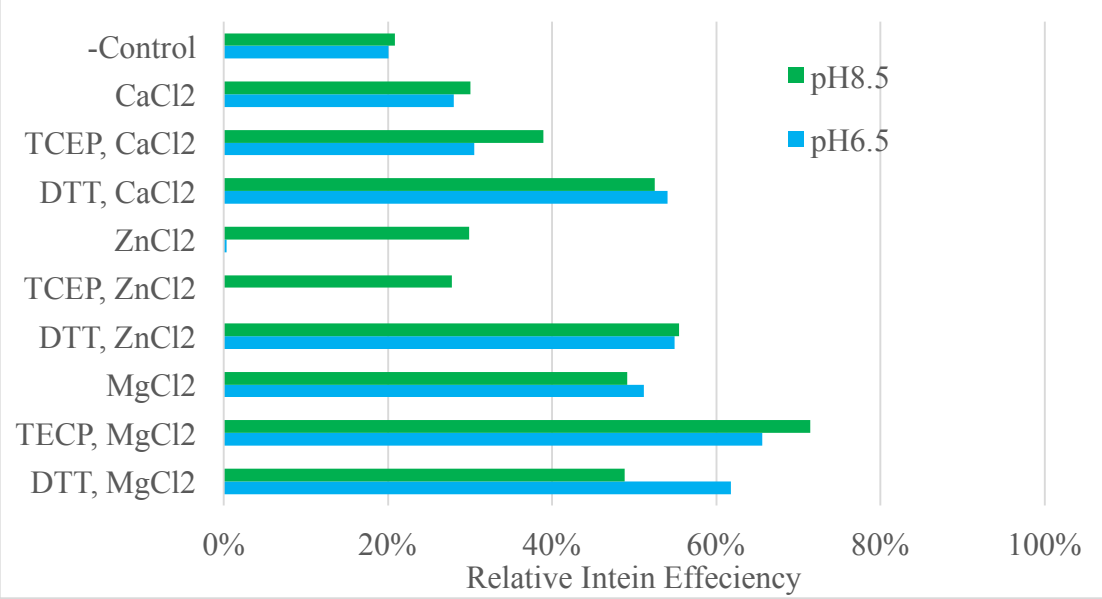


Figure 43. Relative intein effeciency of (C)eGFP samples supplimented with divalent ions or reducing agents at pH 6.5 and pH 8.5. Using the densitometry analysis from **Figure 41**, and the total protein absorbance at 280nm from **Figure 40**, a relative intein effeciency can be calculated. Adding Zn^{2+} to samples inhibits intein processing relative to Ca^{2+} and Mg^{2+}. This can be reversed by adding DTT to the Zn^{2+} samples.

D.3. Conclusions

The motivation for this experiment was to investigate optimal conditions for intein processing. Performing an experiment to measure intein processing is difficult based on the number of steps required for precise analytical measurements. This experiment used semi-pure samples to perform fluorescence and densitometry analysis. Performing a molar balance on the intein fragments would be a streamlined approach for intein processing efficiency, however, some of the fragments would not

bind and elute under the same conditions. Therefore, orthogonal measurements were employed to select a divalent ion, reducing agent, and pH for further studies that required purified inteins.

Based on literature review, Zn^{2+} was inhibitory, and Mg^{2+}, Ca^{2+} and either DTT or TCEP would enhance intein processing. The fluorescence to total protein ratio (**Equation 30**) confirmed that Zn2+ was inhibitory and could be reversible with DTT or TCEP. Using densitometry analysis, coupled with total protein absorbance at 280nm, Zn^{2+} inhibited intein processing was confirmed. All lysis and purification steps will be executed at pH 8.5, and supplemented with 25mM Ca^{2+} and 1mM TCEP.

Appendix E. Effect of coding ligation-enhancing intein fragments for internal cyclization of target proteins

The *SSp* DnaE intein contains two translated fragments: the DnaE-N and DnaE-C proteins which catalyze the excision of a cyclized protein. The natural amino acid coding sequence of the DnaE-N protein has a Cys at the N-terminal position and the DnaE-C has a Gly or Ala on the C-terminal position [72]. The protein cyclization efficiency can be enhanced by coding for a Ser Gly as the N-terminal and Pro Lys as the C-terminal amino acids of the target cyclic protein [79]. In this experiment, a hypothesis is tested that including an engineered amino acid sequence in the target cyclic protein can induce internal ligation. The wild type eGFP sequence had all Cys deleted from the 238-amino acid sequence and replaced by Gly: C49G, and C71G. The N-terminal Cys on the DnaE-N is responsible for the internal thiol nucleophile attack. Deletion of the Cys will prevent any internal residues from executing a nucleophile attack. Then, the engineered ligation site is inserted: Pro Lys Cys Lys. This sequence contains the two C-terminal amino acids from the target cyclized protein and the two N-terminal amino acids from the DnaE-C protein: V69P Q70L G71C F72L. If the hypothesis were true, a 15kDa protein could be observed which would indicate the internal ligation sequence activated the split-intein protein splicing mechanism. Three proteins studied in this experiment were the wild type eGFP (eGFP), wild type eGFP, C49G, C71G, (C)eGFP-GG, and wild type eGFP C49G, V69P, Q70L, and F72L,(C)eGFP-PLCL.

D.1. Materials and Methods

The plasmids containing coding sequences for the three target proteins: (C)eGFP, (C)eGFP-GG, and (C)eGFP-PLCL were constructed as described in **A.7.** pET21a-Cint.egfp.Thr.6H.Nint, pET21a-Cint.egfpnoc.Thr.6H.Nint (C49G and C71G), and pET21a-Cint.egfpind.Thr.6H.Nint (C49G, V69P, Q70L, F72L) The proteins were expressed and purified as described in **2.3.1** Protein Expression**, and 2.3.2** IMAC Purification**.** To confirm accurate amino acid modifications, purified proteins were analyzed for fluorescence. It was hypothesized the deletion of the amino acids in (C)eGFP-GG and (C)eGFP-PLCL would disrupt the beta-barrel chain of wild type eGFP and inhibit fluorescence. After purifying the proteins, the resulting elution fractions would be analyzed for erroneous protein fractions at 15kDa indicating internal ligation. The elution samples were analyzed using a reducing SDS-PAGE and MALDI-TOF analysis.

D.2. Experimental Data

The elution fractions of the three proteins did not identify any erroneous proteins to the wild type (C)eGFP elution patterns. The elution patterns for the three proteins all appeared relatively similar as observed in Figure 44. The relative sensitivity of SDS-PAGE, using Coomassie Brilliant Blue R-250 is 30ng [220], therefore if the internal ligation was inefficient, only erroneous proteins could be identified using MALDI-TOF. It can be concluded the inclusion of artificial ligation sequences does not effectively induce internal ligation in the intein, post-translational processing.

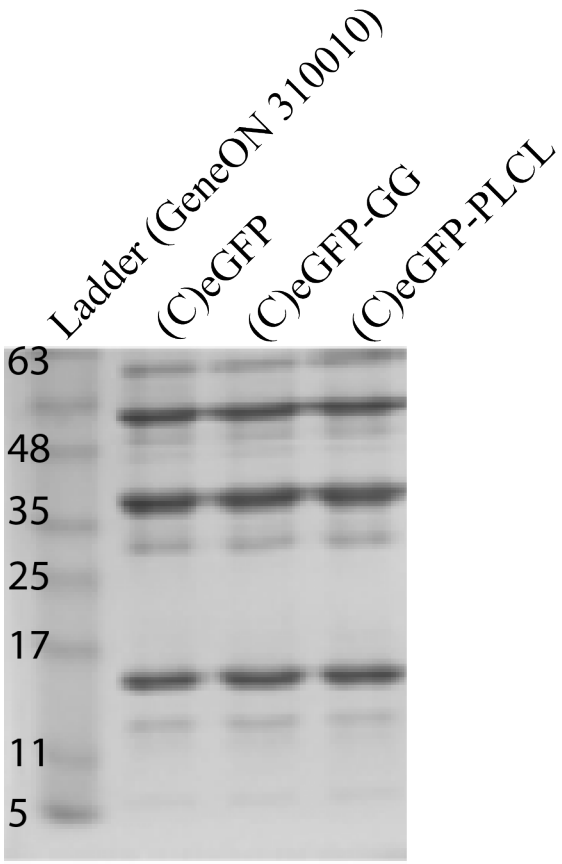


Figure 44. 4-20% Tris-Glycine Gel of Induced Intein, IMAC Elution Samples. No induction of internal ligation from the artifical codon was observed in the protein gels. The 4-20% gel can enhance observation of low molecular weight proteins.

Appendix G. Media and Buffer Formulations

G.1. LMR Media (Minimal Media) Formulation

Media formulation, modified from MTM media.[221]

1000mL solution of 20x (50mL/L) Nitrogen Phosphate Stock (NPS): 50g NH_4Cl, 14.2g Na_2SO_4, 136g KH_2PO_4, 142g Na_2HPO_4

1000mL solution of 50x Carbon Source Stock (CSS): 300g glycerol, 125g D-glucose

100mL solution of 10,000x (0.1mL/L) Trace Metals Stock (TMS): 50mL of 0.1M $FeCl_2$ in 0.01MHCl, 2mL of 1M $CaCl_2$, 1mL of 1M $MnCl_2$, 1mL of 1M $ZnCl_2$, 1mL of 0.2M $CoCl_2$, 2mL of 0.1M $CuCl_2$, 1mL of- 0.2M $NiCl_2$, 2mL of 0.1M Na_2MoO_4, 2mL of 0.1M Na_2SeO_3, 2mL of 0.1M H_3BO_3

100mL solution of 1000x (1mL/L) Vitamin Stock (VS): 4mL- 5mM vitamin B12 (0.2 mM), 2 mL of 10 mM thiamine– HCL, 2mL- 10mM nicotinic acid (0.2 mM), 2mL- 10mM pyridoxine-HCl (0.2 mM), 2mL -10mM *p*-aminobenzoic acid (0.2 mM), 5mL- 100µM folic acid (5 µM), 5mL- 100µM riboflavin (5 µM)

2L of LMR: 1.85L water, 2mL 1M $MgSO_4$, 200µL TMS, 40mL CSS, 100mL NPS, 2mL VS. Adjust to pH6.8 if necessary.

G.2 IMAC Column Buffers

IMAC Load/Lyse Buffer: 50mM Tris-base, 500mM NaCl, 20mM Imidizol
IMAC Wash Buffer: 50mM Tris-base, 500mM NaCl, 50mM Imidizol
IMAC Elution Buffer: 50mM Tris-base, 500mM NaCl, 300mM Imidizol

Appendix H. Evaluation of sodium azide on cell transduction effeciency

The exact mechanism of cell transduction is not defined for peptide-protein, TAT-fusion systems. However, sodium azide has been shown to inhibit endocytosis of cell penetrating peptides [222]. In theory, sodium azide inhibits the energy-dependent endocytosis by oxidative phosphorylation within the cell membrane. The inhibition of oxidative phosphorylation blocks the production of ATP, the energy required for

endocytosis [223, 224]. In this experiment, sodium azide (10µm and 1µm) was added to the cell culture media during co-incubation of (C)eGFP or (C)TATeGFP with CHO-cells. After incubation of 15 and 60-minutes, the extracellular fluorescence was quenched with trypan blue. From Figure 45, at 10µm, sodium azide reduces the relative fluorescence of cells incubated without a fluorescent protein. At 1µm, sodium azide does not have a significant effect on the transduction of GFP.

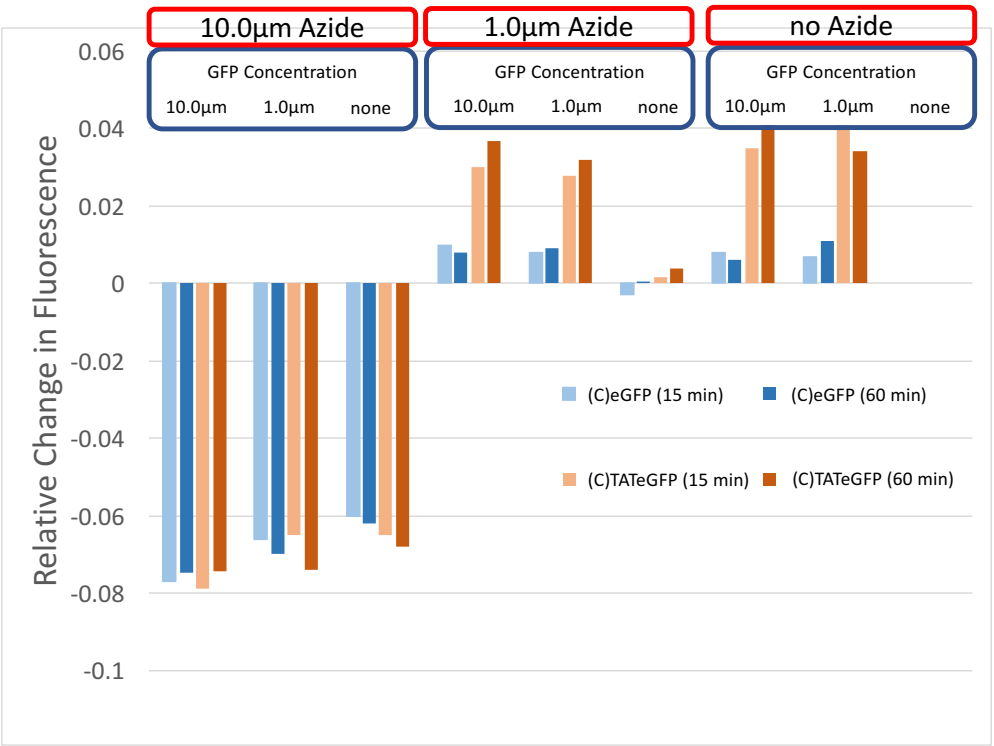


Figure 45 - Sodium Azide does not inhibit endocytosis of intein-derived TAT-peptide protein fusions. At 10µm, sodium azide reduces the relative fluorescence of cells incubated without a fluorescent protein. At 1µm, sodium azide does not have a significant effect on the transduction of GFP.

Bibliography

1. Baneyx, F., *Recombinant protein expression in Escherichia coli.* Curr Opin Biotechnol, 1999. **10**(5): p. 411-21.
2. Summers, D., *Timing, self-control and a sense of direction are the secrets of multicopy plasmid stability.* Mol Microbiol, 1998. **29**(5): p. 1137-45.
3. Williams, S.G., et al., *Repressor titration: a novel system for selection and stable maintenance of recombinant plasmids.* Nucleic Acids Res, 1998. **26**(9): p. 2120-4.
4. Cranenburgh, R.M., et al., *Escherichia coli strains that allow antibiotic-free plasmid selection and maintenance by repressor titration.* Nucleic Acids Res, 2001. **29**(5): p. E26.
5. Garmory, H.S., et al., *Antibiotic-free plasmid stabilization by operator-repressor titration for vaccine delivery by using live Salmonella enterica Serovar typhimurium.* Infect Immun, 2005. **73**(4): p. 2005-11.
6. Carpousis, A.J. and J.D. Gralla, *Interaction of RNA polymerase with lacUV5 promoter DNA during mRNA initiation and elongation. Footprinting, methylation, and rifampicin-sensitivity changes accompanying transcription initiation.* J Mol Biol, 1985. **183**(2): p. 165-77.
7. Daber, R., et al., *Structural analysis of lac repressor bound to allosteric effectors.* J Mol Biol, 2007. **370**(4): p. 609-19.
8. Grossman, T.H., et al., *Spontaneous cAMP-dependent derepression of gene expression in stationary phase plays a role in recombinant expression instability.* Gene, 1998. **209**(1-2): p. 95-103.
9. Glascock, C.B. and M.J. Weickert, *Using chromosomal lacIQ1 to control expression of genes on high-copy-number plasmids in Escherichia coli.* Gene, 1998. **223**(1-2): p. 221-31.
10. Schumacher, M.A., et al., *Crystal structure of LacI member, PurR, bound to DNA: minor groove binding by alpha helices.* Science, 1994. **266**(5186): p. 763-70.
11. Hansen, L.H., S. Knudsen, and S.J. Sorensen, *The effect of the lacY gene on the induction of IPTG inducible promoters, studied in Escherichia coli and Pseudomonas fluorescens.* Curr Microbiol, 1998. **36**(6): p. 341-7.
12. Veigl, M.L., et al., *Effect of isopropyl-beta-D-thiogalactopyranosid induction of the lac operon on the specificity of spontaneous and doxorubicin-induced mutations in Escherichia coli.* Environ Mol Mutagen, 1995. **26**(1): p. 16-25.
13. Dreisigmeyer, D.W., et al., *Determinants of bistability in induction of the Escherichia coli lac operon.* IET Syst Biol, 2008. **2**(5): p. 293-303.
14. Forrer, P. and R. Jaussi, *High-level expression of soluble heterologous proteins in the cytoplasm of Escherichia coli by fusion to the bacteriophage lambda head protein D.* Gene, 1998. **224**(1-2): p. 45-52.

15. Pryor, K.D. and B. Leiting, *High-level expression of soluble protein in Escherichia coli using a His6-tag and maltose-binding-protein double-affinity fusion system.* Protein Expr Purif, 1997. **10**(3): p. 309-19.
16. Hannig, G. and S.C. Makrides, *Strategies for optimizing heterologous protein expression in Escherichia coli.* Trends Biotechnol, 1998. **16**(2): p. 54-60.
17. Sasaki, F., et al., *A high-affinity monoclonal antibody against the FLAG tag useful for G-protein-coupled receptor study.* Anal Biochem, 2012. **425**(2): p. 157-65.
18. Futatsumori-Sugai, M., et al., *Utilization of Arg-elution method for FLAG-tag based chromatography.* Protein Expr Purif, 2009. **67**(2): p. 148-55.
19. Lee, S.Y., *High cell-density culture of Escherichia coli.* Trends Biotechnol, 1996. **14**(3): p. 98-105.
20. Yee, L. and H.W. Blanch, *Recombinant protein expression in high cell density fed-batch cultures of Escherichia coli.* Biotechnology (N Y), 1992. **10**(12): p. 1550-6.
21. Makrides, S.C., *Strategies for achieving high-level expression of genes in Escherichia coli.* Microbiol Rev, 1996. **60**(3): p. 512-38.
22. Murby, M., M. Uhlen, and S. Stahl, *Upstream strategies to minimize proteolytic degradation upon recombinant production in Escherichia coli.* Protein Expr Purif, 1996. **7**(2): p. 129-36.
23. Vera, A., et al., *The conformational quality of insoluble recombinant proteins is enhanced at low growth temperatures.* Biotechnol Bioeng, 2007. **96**(6): p. 1101-6.
24. Structural Genomics, C., et al., *Protein production and purification.* Nat Methods, 2008. **5**(2): p. 135-46.
25. Grant, S.G., et al., *Differential plasmid rescue from transgenic mouse DNAs into Escherichia coli methylation-restriction mutants.* Proc Natl Acad Sci U S A, 1990. **87**(12): p. 4645-9.
26. Kovarik, A., et al., *Transposition of IS10 from the host Escherichia coli genome to a plasmid may lead to cloning artefacts.* Mol Genet Genomics, 2001. **266**(2): p. 216-22.
27. Durfee, T., et al., *The complete genome sequence of Escherichia coli DH10B: insights into the biology of a laboratory workhorse.* J Bacteriol, 2008. **190**(7): p. 2597-606.
28. Daegelen, P., et al., *Tracing ancestors and relatives of Escherichia coli B, and the derivation of B strains REL606 and BL21(DE3).* J Mol Biol, 2009. **394**(4): p. 634-43.
29. Rosano, G.L. and E.A. Ceccarelli, *Recombinant protein expression in Escherichia coli: advances and challenges.* Front Microbiol, 2014. **5**: p. 172.
30. Wagner, S., et al., *Tuning Escherichia coli for membrane protein overexpression.* Proc Natl Acad Sci U S A, 2008. **105**(38): p. 14371-6.
31. Studier, F.W. and B.A. Moffatt, *Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes.* J Mol Biol, 1986. **189**(1): p. 113-30.

32. Fernandez-Castane, A., et al., *Evidencing the role of lactose permease in IPTG uptake by Escherichia coli in fed-batch high cell density cultures.* J Biotechnol, 2012. **157**(3): p. 391-8.
33. Lakowicz, J.R., *Principles of fluorescence spectroscopy*. 2006, Springer: New York. p. 1 online resource (xxvi, 954 p.).
34. Albani, J.R., *Principles and applications of fluorescence spectroscopy*. 2007, Oxford ; Ames, Iowa: Blackwell Science. viii, 255 p., [4] p. of plates.
35. Bacia, K., E. Haustein, and P. Schwille, *Fluorescence correlation spectroscopy: principles and applications.* Cold Spring Harb Protoc, 2014. **2014**(7): p. 709-25.
36. Meech, S.R., *Excited state reactions in fluorescent proteins.* Chem Soc Rev, 2009. **38**(10): p. 2922-34.
37. Niswender, K.D., et al., *Quantitative imaging of green fluorescent protein in cultured cells: comparison of microscopic techniques, use in fusion proteins and detection limits.* J Microsc, 1995. **180**(Pt 2): p. 109-16.
38. Jung, G., J. Wiehler, and A. Zumbusch, *The photophysics of green fluorescent protein: influence of the key amino acids at positions 65, 203, and 222.* Biophys J, 2005. **88**(3): p. 1932-47.
39. Reddy, G., Z. Liu, and D. Thirumalai, *Denaturant-dependent folding of GFP.* Proc Natl Acad Sci U S A, 2012. **109**(44): p. 17832-8.
40. Chang, D.H., et al., *Cooperativity and Site-Selectivity of Intramolecular Hydrogen Bonds on the Fluorescence Quenching of Modified GFP Chromophores.* J Org Chem, 2015. **80**(24): p. 12431-43.
41. Elliott, G., J. McGrath, and E. Crockett-Torabi, *Green fluorescent protein: A novel viability assay for cryobiological applications.* Cryobiology, 2000. **40**(4): p. 360-9.
42. Chalfie, M., *Green fluorescent protein.* Photochem Photobiol, 1995. **62**(4): p. 651-6.
43. Miyawaki, A., *Green fluorescent protein glows gold.* Cell, 2008. **135**(6): p. 987-90.
44. Shimomura, O., F.H. Johnson, and Y. Saiga, *Extraction, purification and properties of aequorin, a bioluminescent protein from the luminous hydromedusan, Aequorea.* J Cell Comp Physiol, 1962. **59**: p. 223-39.
45. Prasher, D.C., et al., *Primary structure of the Aequorea victoria green-fluorescent protein.* Gene, 1992. **111**(2): p. 229-33.
46. Chalfie, M., et al., *Green fluorescent protein as a marker for gene expression.* Science, 1994. **263**(5148): p. 802-5.
47. Heim, R., D.C. Prasher, and R.Y. Tsien, *Wavelength mutations and posttranslational autoxidation of green fluorescent protein.* Proc Natl Acad Sci U S A, 1994. **91**(26): p. 12501-4.
48. Heim, R. and R.Y. Tsien, *Engineering green fluorescent protein for improved brightness, longer wavelengths and fluorescence resonance energy transfer.* Curr Biol, 1996. **6**(2): p. 178-82.
49. Zimmer, M., *GFP: from jellyfish to the Nobel prize and beyond.* Chem Soc Rev, 2009. **38**(10): p. 2823-32.

50. Carlson, H.J., D.W. Cotton, and R.E. Campbell, *Circularly permuted monomeric red fluorescent proteins with new termini in the beta-sheet.* Protein Sci, 2010. **19**(8): p. 1490-9.
51. Sakaue-Sawano, A., et al., *Visualizing spatiotemporal dynamics of multicellular cell-cycle progression.* Cell, 2008. **132**(3): p. 487-98.
52. Li, J., et al., *Impediments to secretion of green fluorescent protein and its fusion from Saccharomyces cerevisiae.* Biotechnol Prog, 2002. **18**(4): p. 831-8.
53. Albano, C.R., et al., *Green fluorescent protein as a real time quantitative reporter of heterologous protein production.* Biotechnol Prog, 1998. **14**(2): p. 351-4.
54. Cha, H.J., T. Gotoh, and W.E. Bentley, *Simplification of titer determination for recombinant baculovirus by green fluorescent protein marker.* Biotechniques, 1997. **23**(5): p. 782-4, 786.
55. Andersen, J.B., et al., *New unstable variants of green fluorescent protein for studies of transient gene expression in bacteria.* Appl Environ Microbiol, 1998. **64**(6): p. 2240-6.
56. Albano, C.R., et al., *High throughput studies of gene expression using green fluorescent protein-oxidative stress promoter probe constructs: the potential for living chips.* J Biomol Screen, 2001. **6**(6): p. 421-8.
57. Tsien, R.Y., *The green fluorescent protein.* Annu Rev Biochem, 1998. **67**: p. 509-44.
58. Pietrokovski, S., *Intein spread and extinction in evolution.* Trends Genet, 2001. **17**(8): p. 465-72.
59. Barnard, G.C., et al., *Integrated recombinant protein expression and purification platform based on Ralstonia eutropha.* Appl Environ Microbiol, 2005. **71**(10): p. 5735-42.
60. Bradley, L.H., et al., *An intein-based genetic selection allows the construction of a high-quality library of binary patterned de novo protein sequences.* Protein Eng Des Sel, 2005. **18**(4): p. 201-207.
61. Chong, S., et al., *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, 1996. **271**(36): p. 22159-22168.
62. Mills, K.V., et al., *Protein purification via temperature-dependent, intein-mediated cleavage from an immobilized metal affinity resin.* Anal Biochem, 2006. **356**(1): p. 86-93.
63. Mills, K.V., M.A. Johnson, and F.B. Perler, *Protein splicing: how inteins escape from precursor proteins.* J Biol Chem, 2014. **289**(21): p. 14498-14505.
64. Mujika, J.I., X. Lopez, and A.J. Mulholland, *Modeling protein splicing: reaction pathway for C-terminal splice and intein scission.* J Phys Chem B, 2009. **113**(16): p. 5607-5616.
65. Xu, M.Q., H. Paulus, and S. Chong, *Fusions to self-splicing inteins for protein purification.* Methods Enzymol, 2000. **326**: p. 376-418.
66. Topilina, N.I. and K.V. Mills, *Recent advances in in vivo applications of intein-mediated protein splicing.* Mob DNA, 2014. **5**(1): p. 5.

67. Frischkorn, K., et al., *Investigation of mycobacterial recA function: protein introns in the RecA of pathogenic mycobacteria do not affect competency for homologous recombination.* Mol Microbiol, 1998. **29**(5): p. 1203-14.
68. Derbyshire, V. and M. Belfort, *Lightning strikes twice: intron-intein coincidence.* Proc Natl Acad Sci U S A, 1998. **95**(4): p. 1356-7.
69. Chong, S., et al., *Modulation of protein splicing of the Saccharomyces cerevisiae vacuolar membrane ATPase intein.* J Biol Chem, 1998. **273**(17): p. 10567-77.
70. Southworth, M.W., et al., *Purification of proteins fused to either the amino or carboxy terminus of the Mycobacterium xenopi gyrase A intein.* Biotechniques, 1999. **27**(1): p. 110-4, 116, 118-20.
71. Wood, D.W., et al., *A genetic system yields self-cleaving inteins for bioseparations.* Nat Biotechnol, 1999. **17**(9): p. 889-892.
72. Wu, H., Z. Hu, and X.Q. Liu, *Protein trans-splicing by a split intein encoded in a split DnaE gene of Synechocystis sp. PCC6803.* Proc Natl Acad Sci U S A, 1998. **95**(16): p. 9226-31.
73. Evans, T.C., Jr., et al., *Protein trans-splicing and cyclization by a naturally split intein from the dnaE gene of Synechocystis species PCC6803.* J Biol Chem, 2000. **275**(13): p. 9091-4.
74. Aranko, A.S., et al., *In vivo and in vitro protein ligation by naturally occurring and engineered split DnaE inteins.* PLoS One, 2009. **4**(4): p. e5185.
75. Zheng, Y., et al., *Mutual synergistic protein folding in split intein.* Biosci Rep, 2012. **32**(5): p. 433-442.
76. Bowers, A.A., *Biochemical and biosynthetic preparation of natural product-like cyclic peptide libraries.* Med. Chem. Commun., 2012. **3**: p. 905-915.
77. Scott, C.P., et al., *Structural requirements for the biosynthesis of backbone cyclic peptide libraries.* Chem Biol, 2001. **8**(8): p. 801-15.
78. Tavassoli, A. and S.J. Benkovic, *Split-intein mediated circular ligation used in the synthesis of cyclic peptide libraries in E. coli.* Nat Protoc, 2007. **2**(5): p. 1126-33.
79. Nilsson, L.O., M. Louassini, and E. Abel-Santos, *Using siclopps for the discovery of novel antimicrobial peptides and their targets.* Protein Pept Lett, 2005. **12**(8): p. 795-9.
80. Frutos, S., et al., *Branched intermediate formation stimulates peptide bond cleavage in protein splicing.* Nat Chem Biol, 2010. **6**(7): p. 527-33.
81. Scott, C.P., et al., *Production of cyclic peptides and proteins in vivo.* Proc Natl Acad Sci U S A, 1999. **96**(24): p. 13638-43.
82. Belfort, M. and SpringerLink (Online service), *Homing endonucleases and inteins*, in *Nucleic acids and molecular biology, 16*. 2005, Springer: Berlin ; New York. p. xix, 377 p.
83. Zettler, J., V. Schutz, and H.D. Mootz, *The naturally split Npu DnaE intein exhibits an extraordinarily high rate in the protein trans-splicing reaction.* FEBS Lett, 2009. **583**(5): p. 909-14.
84. Zettler, J., et al., *SPLICEFINDER - a fast and easy screening method for active protein trans-splicing positions.* PLoS One, 2013. **8**(9): p. e72925.

85. Elleuche, S. and S. Poggeler, *Inteins, valuable genetic elements in molecular biology and biotechnology.* Appl Microbiol Biotechnol, 2010. **87**(2): p. 479-89.
86. Copolovici, D.M., et al., *Cell-penetrating peptides: design, synthesis, and applications.* ACS Nano, 2014. **8**(3): p. 1972-94.
87. Foerg, C., et al., *Differentiation restricted endocytosis of cell penetrating peptides in MDCK cells corresponds with activities of Rho-GTPases.* Pharm Res, 2007. **24**(4): p. 628-42.
88. Pooga, M. and U. Langel, *Classes of Cell-Penetrating Peptides.* Methods Mol Biol, 2015. **1324**: p. 3-28.
89. Ishihara, T., et al., *Intracellular delivery of siRNA by cell-penetrating peptides modified with cationic oligopeptides.* Drug Deliv, 2009. **16**(3): p. 153-9.
90. Eiriksdottir, E., et al., *Secondary structure of cell-penetrating peptides controls membrane interaction and insertion.* Biochim Biophys Acta, 2010. **1798**(6): p. 1119-28.
91. Foerg, C., et al., *Decoding the entry of two novel cell-penetrating peptides in HeLa cells: lipid raft-mediated endocytosis and endosomal escape.* Biochemistry, 2005. **44**(1): p. 72-81.
92. Lindgren, M., et al., *Cell-penetrating peptides.* Trends Pharmacol Sci, 2000. **21**(3): p. 99-103.
93. Howl, J., et al., *Bioportide: an emergent concept of bioactive cell-penetrating peptides.* Cell Mol Life Sci, 2012. **69**(17): p. 2951-66.
94. Lukanowska, M., J. Howl, and S. Jones, *Bioportides: bioactive cell-penetrating peptides that modulate cellular dynamics.* Biotechnol J, 2013. **8**(8): p. 918-30.
95. Madani, F., et al., *Mechanisms of cellular uptake of cell-penetrating peptides.* J Biophys, 2011. **2011**: p. 414729.
96. Floren, A., I. Mager, and U. Langel, *Uptake kinetics of cell-penetrating peptides.* Methods Mol Biol, 2011. **683**: p. 117-28.
97. Lecorche, P., et al., *Cellular uptake and biophysical properties of galactose and/or tryptophan containing cell-penetrating peptides.* Biochim Biophys Acta, 2012. **1818**(3): p. 448-57.
98. Desai, P.R., et al., *(31)P solid-state NMR based monitoring of permeation of cell penetrating peptides into skin.* Eur J Pharm Biopharm, 2014. **86**(2): p. 190-9.
99. Maler, L., *Solution NMR studies of cell-penetrating peptides in model membrane systems.* Adv Drug Deliv Rev, 2013. **65**(8): p. 1002-11.
100. Tanaka, G., et al., *CXCR4 stimulates macropinocytosis: implications for cellular uptake of arginine-rich cell-penetrating peptides and HIV.* Chem Biol, 2012. **19**(11): p. 1437-46.
101. Pae, J., et al., *Translocation of cell-penetrating peptides across the plasma membrane is controlled by cholesterol and microenvironment created by membranous proteins.* J Control Release, 2014. **192**: p. 103-13.
102. Kawaguchi, Y., et al., *Syndecan-4 Is a Receptor for Clathrin-Mediated Endocytosis of Arginine-Rich Cell-Penetrating Peptides.* Bioconjug Chem, 2016. **27**(4): p. 1119-30.

103. Wender, P.A., et al., *The design of guanidinium-rich transporters and their internalization mechanisms.* Adv Drug Deliv Rev, 2008. **60**(4-5): p. 452-72.
104. Richard, J.P., et al., *Cellular uptake of unconjugated TAT peptide involves clathrin-dependent endocytosis and heparan sulfate receptors.* J Biol Chem, 2005. **280**(15): p. 15300-6.
105. Tyagi, M., et al., *Internalization of HIV-1 tat requires cell surface heparan sulfate proteoglycans.* J Biol Chem, 2001. **276**(5): p. 3254-61.
106. Adkisson, C.D., et al., *What extent of pancreatic resection do patients with MEN-1 require?* JOP, 2012. **13**(4): p. 402-8.
107. Richard, J.P., et al., *Cell-penetrating peptides. A reevaluation of the mechanism of cellular uptake.* J Biol Chem, 2003. **278**(1): p. 585-90.
108. Guo, Z., et al., *Cell-penetrating peptides: Possible transduction mechanisms and therapeutic applications.* Biomed Rep, 2016. **4**(5): p. 528-534.
109. Wender, P.A., et al., *The design, synthesis, and evaluation of molecules that enable or enhance cellular uptake: peptoid molecular transporters.* Proc Natl Acad Sci U S A, 2000. **97**(24): p. 13003-8.
110. Rogers, W.J. and P. Basu, *Factors regulating macrophage endocytosis of nanoparticles: implications for targeted magnetic resonance plaque imaging.* Atherosclerosis, 2005. **178**(1): p. 67-73.
111. Ter-Avetisyan, G., et al., *Cell entry of arginine-rich peptides is independent of endocytosis.* J Biol Chem, 2009. **284**(6): p. 3370-8.
112. Appelbaum, J.S., et al., *Arginine topology controls escape of minimally cationic proteins from early endosomes to the cytoplasm.* Chem Biol, 2012. **19**(7): p. 819-30.
113. Li, M., et al., *Discovery and characterization of a peptide that enhances endosomal escape of delivered proteins in vitro and in vivo.* J Am Chem Soc, 2015. **137**(44): p. 14084-93.
114. Huang, H.W., F.Y. Chen, and M.T. Lee, *Molecular mechanism of Peptide-induced pores in membranes.* Phys Rev Lett, 2004. **92**(19): p. 198304.
115. Jenssen, H., P. Hamill, and R.E. Hancock, *Peptide antimicrobial agents.* Clin Microbiol Rev, 2006. **19**(3): p. 491-511.
116. Varkouhi, A.K., et al., *Endosomal escape pathways for delivery of biologicals.* J Control Release, 2011. **151**(3): p. 220-8.
117. Oliveira, S., et al., *Fusogenic peptides enhance endosomal escape improving siRNA-induced silencing of oncogenes.* Int J Pharm, 2007. **331**(2): p. 211-4.
118. Wiley, D.C. and J.J. Skehel, *The structure and function of the hemagglutinin membrane glycoprotein of influenza virus.* Annu Rev Biochem, 1987. **56**: p. 365-94.
119. Horth, M., et al., *Theoretical and functional analysis of the SIV fusion peptide.* EMBO J, 1991. **10**(10): p. 2747-55.
120. Han, X., et al., *Membrane structure and fusion-triggering conformational change of the fusion domain from influenza hemagglutinin.* Nat Struct Biol, 2001. **8**(8): p. 715-20.
121. Lonn, P., et al., *Enhancing Endosomal Escape for Intracellular Delivery of Macromolecular Biologic Therapeutics.* Sci Rep, 2016. **6**: p. 32301.

122. Lonn, P. and S.F. Dowdy, *Cationic PTD/CPP-mediated macromolecular delivery: charging into the cell.* Expert Opin Drug Deliv, 2015. **12**(10): p. 1627-36.
123. Wadia, J.S., R.V. Stan, and S.F. Dowdy, *Transducible TAT-HA fusogenic peptide enhances escape of TAT-fusion proteins after lipid raft macropinocytosis.* Nat Med, 2004. **10**(3): p. 310-5.
124. Qian, Z., et al., *Discovery and Mechanism of Highly Efficient Cyclic Cell-Penetrating Peptides.* Biochemistry, 2016. **55**(18): p. 2601-12.
125. Ramsey, J.D. and N.H. Flynn, *Cell-penetrating peptides transport therapeutics into cells.* Pharmacol Ther, 2015. **154**: p. 78-86.
126. Vives, E., et al., *Effects of the Tat basic domain on human immunodeficiency virus type 1 transactivation, using chemically synthesized Tat protein and Tat peptides.* J Virol, 1994. **68**(5): p. 3343-53.
127. Umezawa, N., et al., *Translocation of a beta-peptide across cell membranes.* J Am Chem Soc, 2002. **124**(3): p. 368-9.
128. Mitchell, D.J., et al., *Polyarginine enters cells more efficiently than other polycationic homopolymers.* J Pept Res, 2000. **56**(5): p. 318-25.
129. Futaki, S., et al., *Arginine-rich peptides. An abundant source of membrane-permeable peptides having potential as carriers for intracellular protein delivery.* J Biol Chem, 2001. **276**(8): p. 5836-40.
130. Vives, E., et al., *TAT peptide internalization: seeking the mechanism of entry.* Curr Protein Pept Sci, 2003. **4**(2): p. 125-32.
131. Upadhya, A. and P.C. Sangave, *Hydrophobic and electrostatic interactions between cell penetrating peptides and plasmid DNA are important for stable non-covalent complexation and intracellular delivery.* J Pept Sci, 2016. **22**(10): p. 647-659.
132. Herce, H.D., et al., *Arginine-rich peptides destabilize the plasma membrane, consistent with a pore formation translocation mechanism of cell-penetrating peptides.* Biophys J, 2009. **97**(7): p. 1917-25.
133. Hallbrink, M., et al., *Cargo delivery kinetics of cell-penetrating peptides.* Biochim Biophys Acta, 2001. **1515**(2): p. 101-9.
134. Mager, I., et al., *Assessing the uptake kinetics and internalization mechanisms of cell-penetrating peptides using a quenched fluorescence assay.* Biochim Biophys Acta, 2010. **1798**(3): p. 338-43.
135. Tseng, Y.L., J.J. Liu, and R.L. Hong, *Translocation of liposomes into cancer cells by cell-penetrating peptides penetratin and tat: a kinetic and efficacy study.* Mol Pharmacol, 2002. **62**(4): p. 864-72.
136. Silhol, M., et al., *Different mechanisms for cellular internalization of the HIV-1 Tat-derived cell penetrating peptide and recombinant proteins fused to Tat.* Eur J Biochem, 2002. **269**(2): p. 494-501.
137. Frankel, A.D. and C.O. Pabo, *Cellular uptake of the tat protein from human immunodeficiency virus.* Cell, 1988. **55**(6): p. 1189-93.
138. Mann, D.A. and A.D. Frankel, *Endocytosis and targeting of exogenous HIV-1 Tat protein.* EMBO J, 1991. **10**(7): p. 1733-9.
139. Snyder, E.L., et al., *Treatment of terminal peritoneal carcinomatosis by a transducible p53-activating peptide.* PLoS Biol, 2004. **2**(2): p. E36.

140. Zhang, T., et al., *Inhibiting bladder tumor growth with a cell penetrating R11 peptide derived from the p53 C-terminus.* Oncotarget, 2015. **6**(35): p. 37782-91.
141. Ueda, Y., et al., *Induction of autophagic cell death of glioma-initiating cells by cell-penetrating D-isomer peptides consisting of Pas and the p53 C-terminus.* Biomaterials, 2012. **33**(35): p. 9061-9.
142. U.S. National Institutes of Health. *A Phase 2a Study to Evaluate the Safety and Efficacy of AZX100 in Trocar Sites of Arthroscopic Shoulder Surgery Patients*. Study Record Detail 2012 September 10, 2012 [cited 2016 September 24]; ClinicalTrials.gov Identifier: NCT00811577]. Available from: https://clinicaltrials.gov/ct2/show/NCT00811577.
143. U.S. National Institutes of Health. *Efficacy and Safety Study of Botulinum Toxin Type A for Moderate to Severe Crow's Feet Lines (RADIANT)*. 2014 April 14, 2014 [cited 2016 September 24]; ClinicalTrials.gov Identifier: NCT01776606]. Available from: https://clinicaltrials.gov/ct2/show/NCT01776606.
144. U.S. National Institutes of Health. *Dose-escalating Safety and Preliminary Efficacy of DaxibotulinumtoxinA for Injection in Cervical Dystonia*. 2016 August 2, 2016 [cited 2016 September 24]; ClinicalTrials.gov Identifier: NCT02706795]. Available from: https://clinicaltrials.gov/ct2/show/NCT02706795.
145. U.S. National Institutes of Health. *Safety and Efficacy of Botulinum Toxin Type A Topical Gel for Primary Axillary Hyperhidrosis*. 2016 August 2, 2016 [cited 2016 September 24]; ClinicalTrials.gov Identifier: NCT02565732]. Available from: https://clinicaltrials.gov/ct2/show/NCT02565732.
146. Fielden, M.R., et al., *Nonclinical Safety Profile of Etelcalcetide, a Novel Peptide Calcimimetic for the Treatment of Secondary Hyperparathyroidism.* Int J Toxicol, 2016. **35**(3): p. 294-308.
147. Use, C.f.M.P.f.H. *Summary of opinion (initial authorisation) for Parsabiv (etelcalcetide)*. 2016 15 September 2016 [cited 2016 September 24]; EMA/CHMP/587633/2016]. Available from: http://www.ema.europa.eu/docs/en_GB/document_library/Summary_of_opinion_-_Initial_authorisation/human/003995/WC500212899.pdf.
148. Health, U.S.N.I.o. *Head-to-Head Study of Etelcalcetide (AMG 416) and Cinacalcet*. 2016 July 22, 2016 [cited 2016 September 24]; ClinicalTrials.gov Identifier: NCT01896232]. Available from: https://clinicaltrials.gov/ct2/show/NCT01896232.
149. Flynn, C.R., et al., *Internalization and intracellular trafficking of a PTD-conjugated anti-fibrotic peptide, AZX100, in human dermal keloid fibroblasts.* J Pharm Sci, 2010. **99**(7): p. 3100-21.
150. Stone, H.F., et al., *Characterization of diffusion and duration of action of a new botulinum toxin type A formulation.* Toxicon, 2011. **58**(2): p. 159-67.
151. Krautwald, S., et al., *Inhibition of regulated cell death by cell-penetrating peptides.* Cell Mol Life Sci, 2016. **73**(11-12): p. 2269-84.
152. Jeffries, C.M., et al., *Stabilization of a binary protein complex by intein-mediated cyclization.* Protein Sci, 2006. **15**(11): p. 2612-8.

153. Williams, N.K., et al., *Stabilization of native protein fold by intein-mediated covalent cyclization.* J Mol Biol, 2005. **346**(4): p. 1095-108.
154. Johnson, C.M., *Differential scanning calorimetry as a tool for protein folding and stability.* Arch Biochem Biophys, 2013. **531**(1-2): p. 100-9.
155. Neira, J.L. and J. Gomez, *The conformational stability of the Streptomyces coelicolor histidine-phosphocarrier protein. Characterization of cold denaturation and urea-protein interactions.* Eur J Biochem, 2004. **271**(11): p. 2165-81.
156. Thompson, J.B., et al., *The backbone conformational entropy of protein folding: experimental measures from atomic force microscopy.* J Mol Biol, 2002. **322**(3): p. 645-52.
157. Arnau, J., Conni Lauritzen, Gitte E. Petersen, and John Pedersen, *Reprint Of: Current Strategies for the Use of Affinity Tags and Tag Removal for the Purification of Recombinant Proteins.* Protein Expression and Purification, 2005.
158. Christianson, D.W., P.R. David, and W.N. Lipscomb, *Mechanism of carboxypeptidase A: hydration of a ketonic substrate analogue.* Proc Natl Acad Sci U S A, 1987. **84**(6): p. 1512-5.
159. Caron, N.J., S.P. Quenneville, and J.P. Tremblay, *Endosome disruption enhances the functional nuclear delivery of Tat-fusion proteins.* Biochem Biophys Res Commun, 2004. **319**(1): p. 12-20.
160. Marth, C., et al., *Effects of Taxol on choriocarcinoma cells.* Am J Obstet Gynecol, 1995. **173**(6): p. 1835-42.
161. Riechelmann, R.P. and E.D. Saad, *A systematic review on drug interactions in oncology.* Cancer Invest, 2006. **24**(7): p. 704-12.
162. Gill, P.S., et al., *The effects of preparations of human chorionic gonadotropin on AIDS-related Kaposi's sarcoma.* N Engl J Med, 1996. **335**(17): p. 1261-9.
163. Audie, J. and J. Swanson, *Advances in the prediction of protein-peptide binding affinities: implications for peptide-based drug discovery.* Chem Biol Drug Des, 2013. **81**(1): p. 50-60.
164. Philippe, G., et al., *Development of cell-penetrating peptide-based drug leads to inhibit MDMX:p53 and MDM2:p53 interactions.* Biopolymers, 2016. **106**(6): p. 853-863.
165. Darlington, G.J., *Viability staining of Mammalian cell cultures.* CSH Protoc, 2007. **2007**: p. pdb prot4769.
166. Browne, S.M. and M. Al-Rubeai, *Defining viability in mammalian cell cultures.* Biotechnol Lett, 2011. **33**(9): p. 1745-9.
167. Masuda, H. and T. Asahara, *Clonogenic assay of endothelial progenitor cells.* Trends Cardiovasc Med, 2013. **23**(4): p. 99-103.
168. Franken, N.A., et al., *Clonogenic assay of cells in vitro.* Nat Protoc, 2006. **1**(5): p. 2315-9.
169. Rafehi, H., et al., *Clonogenic assay: adherent cells.* J Vis Exp, 2011(49).
170. Wang, S., *Anchorage-independent growth of prostate cancer stem cells.* Methods Mol Biol, 2009. **568**: p. 151-60.
171. Welte, Y., et al., *Patient derived cell culture and isolation of CD133(+) putative cancer stem cells from melanoma.* J Vis Exp, 2013(73): p. e50200.

172. Quere, R., et al., *High levels of the adhesion molecule CD44 on leukemic cells generate acute myeloid leukemia relapse after withdrawal of the initial transforming event.* Leukemia, 2011. **25**(3): p. 515-26.
173. McFarlane, S., et al., *CD44 increases the efficiency of distant metastasis of breast cancer.* Oncotarget, 2015. **6**(13): p. 11465-76.
174. Lunardi-Iskandar, Y., et al., *Effects of a urinary factor from women in early pregnancy on HIV-1, SIV and associated disease.* Nat Med, 1998. **4**(4): p. 428-34.
175. Hawe, A., et al., *Forced degradation of therapeutic proteins.* J Pharm Sci, 2012. **101**(3): p. 895-913.
176. .S. Dept. of Health and Human Services, F.a.D.A., Center for Drug Evaluation and Research : Center for Biologics Evaluation and Research Rockville, MD 2003, *Guidance for industry [electronic resource] : Q1A(R2) stability testing of new drug substances and products*. Center for Drug Evaluation and Research (U.S.). and Center for Biologics Evaluation and Research (U.S.). and International Conference on Harmonisation.
177. Manning, M.C., et al., *Stability of protein pharmaceuticals: an update.* Pharm Res, 2010. **27**(4): p. 544-75.
178. Philo, J.S., *A critical review of methods for size characterization of non-particulate protein aggregates.* Curr Pharm Biotechnol, 2009. **10**(4): p. 359-72.
179. Food and H.H.S. Drug Administration, *International Conference on Harmonisation; guidance on Q5E Comparability of Biotechnological/Biological Products Subject to Changes in Their Manufacturing Process; availability. Notice.* Fed Regist, 2005. **70**(125): p. 37861-2.
180. Pucci, F. and M. Rooman, *Physical and molecular bases of protein thermal stability and cold adaptation.* Curr Opin Struct Biol, 2017. **42**: p. 117-128.
181. *Quality of biotechnological products: stability testing of biotechnological/biological products. Annex to the ICH Harmonised Tripartite Guideline for the Stability Testing of New Drug Substances and Products.* Dev Biol Stand, 1998. **93**: p. 211-9.
182. Ahmad, F., S. Yadav, and S. Taneja, *Determining stability of proteins from guanidinium chloride transition curves.* Biochem J, 1992. **287 (Pt 2)**: p. 481-5.
183. Halter, M., et al., *Automated live cell imaging of green fluorescent protein degradation in individual fibroblasts.* Cytometry A, 2007. **71**(10): p. 827-34.
184. Kitsera, N., A. Khobta, and B. Epe, *Destabilized green fluorescent protein detects rapid removal of transcription blocks after genotoxic exposure.* Biotechniques, 2007. **43**(2): p. 222-7.
185. Chirico, G., F. Cannone, and A. Diaspro, *Unfolding time distribution of GFP by single molecule fluorescence spectroscopy.* Eur Biophys J, 2006. **35**(8): p. 663-74.
186. Jung K-C, P.J.-B., Maeng P-J, Kim H-J, *Fluorescence quenching of green fluorescent protein during denaturation by guanidine.* Bulletin of the Korean chemical society, 2005. **26**(3): p. 413-417.

187. Scheiner, S. and W.N. Lipscomb, *Molecular orbital studies of enzyme activity. 4. Hydrolysis of peptides by carboxypeptidase A*. J Am Chem Soc, 1977. **99**(10): p. 3466-72.
188. Park, J.D. and D.H. Kim, *Cysteine derivatives as inhibitors for carboxypeptidase A: synthesis and structure-activity relationships.* J Med Chem, 2002. **45**(4): p. 911-8.
189. Zhloba, A.A., et al., *[Arginine and lysine as products of basic carboxypeptidase activity associated with fibrinolysis].* Biomed Khim, 2013. **59**(5): p. 570-7.
190. Lee, H.M. and J.F. Riordan, *Does carboxypeptidase Y have intrinsic endopeptidase activity?* Biochem Biophys Res Commun, 1978. **85**(3): p. 1135-42.
191. Li, F., et al., *Cell culture processes for monoclonal antibody production.* MAbs, 2010. **2**(5): p. 466-79.
192. Marini, G., et al., *Experimental design approach in recombinant protein expression: determining medium composition and induction conditions for expression of pneumolysin from Streptococcus pneumoniae in Escherichia coli and preliminary purification process.* BMC Biotechnol, 2014. **14**: p. 1.
193. Larentis, A.L., et al., *Evaluation of pre-induction temperature, cell growth at induction and IPTG concentration on the expression of a leptospiral protein in E. coli using shaking flasks and microbioreactor.* BMC Res Notes, 2014. **7**: p. 671.
194. Navarro Llorens, J.M., A. Tormo, and E. Martinez-Garcia, *Stationary phase in gram-negative bacteria.* FEMS Microbiol Rev, 2010. **34**(4): p. 476-95.
195. Ou, J., et al., *Stationary phase protein overproduction is a fundamental capability of Escherichia coli.* Biochem Biophys Res Commun, 2004. **314**(1): p. 174-80.
196. Pletnev, P., et al., *Survival guide: Escherichia coli in the stationary phase.* Acta Naturae, 2015. **7**(4): p. 22-33.
197. Prouty, W.F. and A.L. Goldberg, *Effects of protease inhibitors on protein breakdown in Escherichia coli.* J Biol Chem, 1972. **247**(10): p. 3341-52.
198. Harcum, S.W. and W.E. Bentley, *Response dynamics of 26-, 34-, 39-, 54-, and 80-kDa proteases in induced cultures of recombinant Escherichia coli.* Biotechnol Bioeng, 1993. **42**(6): p. 675-85.
199. Heim, R., A.B. Cubitt, and R.Y. Tsien, *Improved green fluorescence.* Nature, 1995. **373**(6516): p. 663-4.
200. Leveau, J.H. and S.E. Lindow, *Predictive and interpretive simulation of green fluorescent protein expression in reporter bacteria.* J Bacteriol, 2001. **183**(23): p. 6752-62.
201. Saeed, I.A. and S.S. Ashraf, *Denaturation studies reveal significant differences between GFP and blue fluorescent protein.* Int J Biol Macromol, 2009. **45**(3): p. 236-41.
202. Kilpatrick, *NIST Mass and Fragment Calculator*. 2012, Kilpatrick et al.: Protein Purification and Expression 85
203. Colussi, T.M., et al., *Initiation of translation in bacteria by a structured eukaryotic IRES RNA.* Nature, 2015. **519**(7541): p. 110-3.

204. Vethanayagam, J.G. and A.M. Flower, *Decreased gene expression from T7 promoters may be due to impaired production of active T7 RNA polymerase.* Microb Cell Fact, 2005. **4**(1): p. 3.
205. Wahlers, A., et al., *Influence of multiplicity of infection and protein stability on retroviral vector-mediated gene expression in hematopoietic cells.* Gene Ther, 2001. **8**(6): p. 477-86.
206. Nischan, N., et al., *Covalent attachment of cyclic TAT peptides to GFP results in protein delivery into live cells with immediate bioavailability.* Angew Chem Int Ed Engl, 2015. **54**(6): p. 1950-3.
207. Hall, D.J., et al., *Transduction of a dominant-negative H-Ras into human eosinophils attenuates extracellular signal-regulated kinase activation and interleukin-5-mediated cell viability.* Blood, 2001. **98**(7): p. 2014-21.
208. Caron, N.J., et al., *Intracellular delivery of a Tat-eGFP fusion protein into muscle cells.* Mol Ther, 2001. **3**(3): p. 310-8.
209. Cermenati, G., et al., *The CPP Tat enhances eGFP cell internalization and transepithelial transport by the larval midgut of Bombyx mori (Lepidoptera, Bombycidae).* J Insect Physiol, 2011. **57**(12): p. 1689-97.
210. Munoz, N.M. and A.R. Leff, *Highly purified selective isolation of eosinophils from human peripheral blood by negative immunomagnetic selection.* Nat Protoc, 2006. **1**(6): p. 2613-20.
211. Del Gaizo, V. and R.M. Payne, *A novel TAT-mitochondrial signal sequence fusion protein is processed, stays in mitochondria, and crosses the placenta.* Mol Ther, 2003. **7**(6): p. 720-30.
212. Yang, Y., et al., *HIV-1 TAT-mediated protein transduction and subcellular localization using novel expression vectors.* FEBS Lett, 2002. **532**(1-2): p. 36-44.
213. Flinterman, M., et al., *Delivery of therapeutic proteins as secretable TAT fusion products.* Mol Ther, 2009. **17**(2): p. 334-42.
214. Paramban, R.I., R.C. Bugos, and W.W. Su, *Engineering green fluorescent protein as a dual functional tag.* Biotechnol Bioeng, 2004. **86**(6): p. 687-97.
215. Neil, J.J. and G.L. Bretthorst, *On the use of Bayesian probability theory for analysis of exponential decay data: an example taken from intravoxel incoherent motion experiments.* Magn Reson Med, 1993. **29**(5): p. 642-7.
216. Thion, L., et al., *The two-step cleavage activity of PI-TfuI intein endonuclease demonstrated by matrix-assisted laser desorption ionization time-of-flight mass spectrometry.* J Biol Chem, 2002. **277**(47): p. 45442-50.
217. Saves, I., et al., *Inteins of Thermococcus fumicolans DNA polymerase are endonucleases with distinct enzymatic behaviors.* J Biol Chem, 2000. **275**(4): p. 2335-41.
218. Mills, K.V. and H. Paulus, *Reversible inhibition of protein splicing by zinc ion.* J Biol Chem, 2001. **276**(14): p. 10832-8.
219. Guan, D. and Z. Chen, *Affinity Purification of Proteins in Tag-Free Form: Split Intein-Mediated Ultrarapid Purification (SIRP).* Methods Mol Biol, 2017. **1495**: p. 1-12.

220. Choi, J.K. and G.S. Yoo, *Fast protein staining in sodium dodecyl sulfate polyacrylamide gel using counter ion-dyes, Coomassie brilliant blue R-250 and neutral red.* Arch Pharm Res, 2002. **25**(5): p. 704-8.
221. Lorsch, J., *Methods in Enzymology. Laboratory methods in enzymology: protein part B. Preface.* Methods Enzymol, 2014. **539**: p. xv.
222. Wang, H., et al., *Highly Efficient Delivery of Functional Cargoes by a Novel Cell-Penetrating Peptide Derived from SP140-Like Protein.* Bioconjug Chem, 2016. **27**(5): p. 1373-81.
223. Drin, G., et al., *Studies on the internalization mechanism of cationic cell-penetrating peptides.* J Biol Chem, 2003. **278**(33): p. 31192-201.
224. Lewis, H.D., et al., *Creation of a novel peptide with enhanced nuclear localization in prostate and pancreatic cancer cell lines.* BMC Biotechnol, 2010. **10**: p. 79.