

Stony Brook University



OFFICIAL COPY

The official electronic file of this thesis or dissertation is maintained by the University Libraries on behalf of The Graduate School at Stony Brook University.

© All Rights Reserved by Author.

**N/C Terminal Relocation, Truncation, and Native Chemical Ligation;
Accessing the Chromophore of Green Fluorescent Protein**

A Dissertation Presented

by

Eduard Hendrik Melief

to

The Graduate School

In Partial fulfillment of the

Requirements

for the Degree of

Doctor of Philosophy

in

Biochemistry and Structural Biology

Stony Brook University

August 2010

Copyright by
Eduard Hendrik Melief
2010

Stony Brook University

The Graduate School

Eduard Hendrik Melief

We, the dissertation committee for the above candidate for the

Doctor of Philosophy degree,

hereby recommend acceptance of this dissertation.

Peter J. Tonge - Dissertation Advisor

Professor, Department of Chemistry

Erwin London - Chairperson

Professor, Department of Biochemistry and Cell Biology

Carlos L. Simmerling – Committee Member

Associate Professor, Department of Chemistry

James B. Bliska - Outside Member

Professor, Department of Molecular Genetics and Microbiology,
Stony Brook University

This dissertation is accepted by the Graduate School

Lawrence Martin
Dean of the Graduate School

Abstract of the Dissertation

**N/C Terminal Relocation, Truncation, and Native Chemical Ligation; Accessing
the Chromophore of Green Fluorescent Protein**

by

Eduard Hendrik Melief

Doctor of Philosophy

in

Biochemistry and Structural Biology

Stony Brook University

2010

The Green Fluorescent Protein (GFP), from the jellyfish *Aequorea Victoria*, is a vital imaging tool in cellular and molecular biology. The *p*-hydroxybenzylidene imidazole chromophore is formed post-translationally within the beta-barrel protein from three centrally located amino acids; S65, Y66, and G67. In order to probe the mechanism of chromophore formation and facilitate the introduction of unnatural amino acids into the chromophore, a permuted variant of GFP was constructed in which the existing N- and C- termini of the protein are linked and a new N-terminus is created close to the tripeptide chromophore region. Spectroscopic studies demonstrate that this GFP variant closely resembles wild-type GFP. Following purification from inclusion bodies, the GFP variant efficiently undergoes folding and *de novo* chromophore formation.

N-Terminal truncation of the permuted GFP protein resulted in a form of GFP to which short peptides can be ligated, thereby facilitating the incorporation of isotopically labeled and unnatural amino acids into the chromophore. A novel and efficient method for the generation of such a truncation variant under the partially denaturing conditions was developed. The production of thioester peptides for ligation to the truncated GFP variant using Fmoc solid phase synthesis was explored and successfully implemented. Ligation reactions of the bacterially expressed truncation GFP variant to the synthetically generated thioester peptides under denaturing conditions were refined using existing methodology and *de novo* folding of the resultant ligation product was carried out.

Modified unnatural amino-acids and dipeptides were synthetically generated and incorporated into the thioester peptide to explore formation of the chromophore. In order to probe the excited state proton transfer mechanism essential for the fluorescence of GFP, a fluorinated tyrosine substituted into the Y66 position that lowers the pKa of the resulting chromophore phenol was biosynthetically generated and synthetically protected for introduction into a thioester peptide by solid phase synthesis.

Further studies using the techniques developed will allow novel chromophores to be studied inside the protein matrix. Lastly, this general methodology may be extended to other protein systems, thereby expanding the chemical space within a protein system beyond the genetically encoded amino acid set.

Dedication

To Christina, for all the inspiration, motivation,
and unending love & support.

To Jacob, who makes my life richer every single day.

To all of the parents in my life, your gentle prodding
and advice have been invaluable.

Table of Contents

List of Abbreviations and Symbols.....	ix
List of Figures.....	xiii
List of Tables.....	xvi
Acknowledgements.....	xvii
Chapter 1: Introduction to Green Fluorescent Protein and Methodologies of Study.....	1
1. History of Green Fluorescent Protein.....	2
2. Structure of GFP and Proposed Biosynthetic Pathway.....	5
3. Unnatural amino acids and circular permutation as tools for elucidation of chromophore biosynthesis.....	13
4. Aim of Study.....	19
5. References.....	21
Chapter 2: Generation of GFP Circular Permutants and Characterization of Variants.....	25
1. Introduction to approach and limitations of GFP Circular Permutants.....	26
2. Methods	
Cloning and Mutagenesis.....	28
Protein Expression and Purification.....	31
Insoluble Protein Expression and <i>De Novo</i> Folding.....	32
SDS-PAGE and Western Blotting.....	33
3. Results and Discussion	
Expression and Evaluation of Circular Permutants.....	35
50-GFP α A206K exhibits absorption and fluorescent emission similar to wtGFP.....	41
Circular permutant 50-GFP α A206K is structurally similar to wtGFP.....	43
4. Summary.....	53
5. References.....	55

Chapter 3:	Development of bacterially expressed truncation protein with an N-terminal cysteine, synthesis of a thioester peptide using Fmoc chemistry, native chemical ligation, <i>de novo</i> folding of ligation products, and characterization of ligated 50-GFP α A206K.....	57
1.	Introduction.....	58
	Thioester peptide synthesis strategy.....	61
	Native Chemical Ligation Strategy.....	67
2.	Methods	
	Cloning and mutagenesis.....	70
	Protein Expression and Purification.....	71
	Insoluble Expression of Protein and Endoproteolytic Cleavage.....	72
	Synthesis of peptide thioesters via a protected peptide intermediate synthesized on chlorotrityl resin.....	77
	Native chemical ligation and <i>de novo</i> folding of peptide thioester with 50T70GFP α A206K.....	78
3.	Results and Discussion	
	N-terminal cysteine truncation protein generated by intein cleavage.....	80
	N-terminal cysteine truncation protein generated by endoproteolysis.....	83
	Synthesis of peptide thioesters via O-allyl protected side chain anchored glutamine.....	88
	Synthesis of peptide thioesters via a resin-free side chain protected peptide intermediate.....	91
	Native chemical ligation and <i>de novo</i> folding.....	93
4.	Summary.....	98
5.	References.....	102
Chapter 4:	Synthesis of peptide thioesters for native chemical ligation using Fmoc chemistry and synthesis of unnatural amino acids for incorporation into thioester peptides.....	106
1.	Introduction.....	107
2.	Methods	
	Insoluble expression of protein and endoproteolytic cleavage.....	110
	Tyrosinyl-glycine Pseudo-dipeptide synthesis.....	112
	Synthesis of peptide thioesters via a protected peptide intermediate synthesized on chlorotrityl resin.....	116
	Native chemical ligation and <i>de novo</i> folding.....	118
3.	Results and Discussion.....	119
4.	Summary.....	129
5.	References.....	131

Chapter 5: Use of unnatural amino acids to modulate the photophysical properties of green fluorescent protein.....	132
1. Introduction.....	133
2. Methods	
Insoluble expression of protein and endoproteolytic cleavage.....	135
Fluorinated tyrosine synthesis and protection for SPPS.....	136
Synthesis of peptide thioesters via a protected peptide intermediate synthesized on chlorotrityl resin.....	140
Native chemical ligation and <i>de novo</i> folding.....	142
3. Results and Discussion.....	144
4. Conclusions and Future Directions.....	155
5. References.....	160
List of References.....	162
Appendix 1: Pseudo dipeptide synthesis intermediate mass spectra.....	172

List of Abbreviations and Symbols

3,5FY	3,5 fluorotyrosine
β -ME	β -mercaptoethanol
Å	Angstrom
A	steady state neutral, ground state
A*	steady state neutral, excited state
aaRS	aminoacyl-tRNA synthetase
Ala (A)	alanine
Amp	ampicillin
Amp ^r	ampicillin resistance
AMU	atomic mass unit
Arg (R)	arginine
Asn (N)	asparagine
Asp (D)	aspartic acid
ASU	Arizona State University
B	steady state anion, ground state
B*	steady state anion, excited state
BAL	backbone amide linker
BCIP	5-Bromo-4-chloro-3-indolyl phosphate
Boc	tert-butoxycarbonyl
C	Celcius
CDI	carbonyldiimidazole
cm	centimeter
cps	counts per second
Cys (C)	cysteine
DCC	N,N'-dicyclohexylcarbodiimide
DIPEA	N,N'-diisopropylethylamine
DMAP	4-dimethylaminopyridine
DMF	N,N-dimethylformamide
DTT	dithiothreitol
EDTA	ethylenediaminetetraacetic acid
EPL	expressed protein ligation
ESI	Electrospray Ionization
ESPT	excited state proton transfer
Fmoc	9-fluorenylmethylloxycarbonyl
FRET	Forster resonance energy transfer partner
FT-IR	Fourier transform ion cyclotron resonance
g	grams

<i>g</i>	standard gravity
GFP	Green Fluorescent Protein
Gln (Q)	glutamine
Glu (E)	glutamic acid
Gly (G)	glycine
HAL	Histidine Ammonia Lyase
HBTU	2-(1H-benzotriazole-1-yl)-1,1,3,3-tetramethylaminium hexafluorophosphate
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
His (H)	histidine
HPLC	high performance liquid chromatography
I	excited state
I*	ground state
Ile (I)	isoleucine
INN	chloramphenicol
IPTG	isopropyl-beta-D-thiogalactopyranoside
IR	infrared
kDa	kiloDalton
KIE	Kinetic Isotope Effect
L	liter
LB	Luria bertani media
Leu (L)	leucine
Lys (K)	lysine
M	mass
MALDI	matrix assisted laser desorption/ionization
MES	4-morpholinoethanesulfonic acid
MESNA	mercaptoethanesulfonic acid
mg	milligram
μ	micro
μ g	microgram
min	minute
mL	millileter
mM (mmol)	millimolar
MPA	mercaptopropionic acid ethyl ester
MPAA	mercaptophenyl acetic acid
MS	mass spectrometry
MWCO	molecular weight cut-off
MY	<i>meta</i> -tyrosine
NBT	nitroblue tetrazolium
nm	nanometer
NME	<i>N</i> -methyl
NMM	<i>N</i> -methylnmorpholine
ns	nanosecond
NTA	nitrilotriacetic acid

PAGE	polyacrylamide gel electrophoresis
PAL-PEG-PS	5-(4'-Fmoc-aminomethyl-3',5'-dimethoxy-phenoxy)-valeric acid]/poly(ethylene glycol)/polystyrene
PBS	phosphate buffered saline
PCR	polymerase chain reaction
PDB ID	protein data bank identification number
PDVF	microporous polyvinylidene fluoride
Phe (F)	phenolalanine
π	pi
pKa	acid dissociation constant
ps	picosecond
PSE	pseudodipeptide
PyBOP	benzotriazol-1-yl-oxytripyrrolidinophoshonium hexafluorophosphate
RP-HPLC	reverse phase high performance liquid chromatography
RPM	revolutions per minute
SDS	sodiumdodecylsulfate
Ser (S)	serine
SPPS	solid phase peptide synthesis
TAG	amber codon
TBS	tris buffered saline
TBTU	tetrafluoroborate
tbu	tert-butyl
TCA	trichloroacetic acid
TCEP	tris-2-carboxyethyl phosphine
Tet ^f	tetracyclin resistance
TFA	trifluoroacetic acid
THF	tetrahydrofuran
Thr (T)	threonine
TIS	triisopropylsilane
TLC	thin layer chromatography
TPL	tyrosine pheol lyase
TRF	time resolved fluorometry
TRIR	time resolved infrared spectroscopy
TRIS	tris(hydroxymethyl)aminomethane
tRNA	transfer ribonucleic acid
TRP (W)	tryptophan
trt	teryl
Tyr (Y)	tyrosine
UV	ultraviolet
V	volts
v	volume
Val (V)	valine

w	weight
W (sonication)	Watts
wt	wild type

List of Figures

Chapter 1	
1.1	Photocomplex within <i>A. Victoria</i> and GFP.....2
1.2	Absorption spectra of GFP.....3
1.3	X-ray crystal structure of wtGFP and its chromophore.....6
1.4	Photocycle of GFP.....8
1.5	Current models of chromophore formation.....11
1.6	Methods for incorporation of unnatural amino acids.....14
Chapter 2	
2.1	Purification profile and absorption spectra for 48-GFP.....37
2.2	Purification, absorption, and fluorescence of 50-GFP.....39
2.3	Western blot analysis of 50-GFP degradation products and redesign of the linker.....40
2.4	Characterization of wtGFP and 50-GFP α A206K; purification, absorption, and fluorescence spectra42
2.5	Irreversible photoconversion of 50-GFP α A206K and wtGFP due to UV induced decarboxylation at Glu222.....43
2.6	Structural overlay of wtGFP and 50-GFP α A206K.....45
2.7	Raman spectrum of wtGFP and 50-GFP α A206K.....46
2.8	Flow injection ESI-MS of purified 50-GFP α A206K in positive mode48
2.9	Comparison of <i>de novo</i> folded to natively expressed 50-GFP α A206K.....49
2.10	Optimizing <i>de novo</i> folding of the 50-GFP α A206K circular permutant by varying denatured protein concentrations.....50
2.11	Optimization of <i>de novo</i> folding by modulating folding buffer dilution factor.....51
Chapter 3	
3.1	Mechanism of intein splicing and native chemical ligation.....60
3.2	Two methods of thioester peptide generation using Fmoc-based solid phase synthesis.....63
3.3	Attempted generation of a thioester peptide by using a side-chain anchored C-terminal protected amino acid scaffold.....64
3.4	Synthesis of peptide thioester via a resin free protected

peptide intermediate.....	66
3.5 Native chemical ligation reaction promoted by <i>in situ</i> transesterification with reactive thiols.....	68
3.6 Characterization and purification of 50T70GFP α A206K- <i>Ssp</i> DNAB intein fusion construct.....	81
3.7 Chitin purification of the 50T70GFP α A206K- <i>Ssp</i> DnaB intein fusion construct induced at 16°C.....	82
3.8 Factor Xa endoproteolysis of 50-GFP α A206K-factor Xa	84
3.9 Enterokinase endoproteolysis of 50-GFP α A206K-entero.....	85
3.10 Large scale enterokinase endoproteolysis of 50-GFP α A206K-entero generating 50T70GFP α A206K.....	87
3.11 MALDI mass spectrometry of the 21-amino acid O-allyl protected peptide.....	89
3.12 Unsuccessful esterification reactions on the carboxylic acid of the side chain anchored peptide.....	90
3.13 Characterization of the synthesis of thioester peptides on chlorotrityl resin.....	92
3.14 Native chemical ligation of the peptide thioester to 50T70GFP α A206K.....	94
3.15 Absorption and fluorescence spectra of folded ligation product.....	96
 Chapter 4	
4.1 Current models of GFP chromophore formation.....	107
4.2 Proposed modifications to the Tyr66 backbone to interrupt chromophore formation.....	108
4.3 Synthesis of tyrosine-glycine pseudodipeptide product.....	112
4.4 HPLC purification of the N-methyl tyrosine containing thioester peptide.....	121
4.5 HPLC purification of the pseudo-dipeptide containing thioester peptide.....	122
4.6 MALDI mass spectrometry of the 21-amino acid O-allyl protected peptide.....	124
4.7 Absorption spectrum of the wild-type, N-methyl tyrosine, and pseudo-dipeptide ligation products after <i>de novo</i> folding.....	125
4.8 MALDI mass spectrum analysis of folded wild type peptide, N-methyl tyrosine, and pseudo-dipeptide ligation products.....	126

Chapter 5

5.1	Proposed unnatural amino acids to incorporate into the GFP chromophore and modulate photophysics.....	134
5.2	Synthesis of Fmoc-L-3,5-difluorotyrosine(tBu)-OH from L-3,5 difluorotyrosine.....	138
5.3	3,5-difluorotyrosine biosynthesis using tyrosine phenol lyase.....	145
5.4	Native chemical ligation of peptide thioesters incorporating unnatural amino acids to 50T70GFP α A206K.....	148
5.5	Absorption and fluorescence of folded ligation products.....	149
5.6	pH titration of 3,5-difluorotyrosine, wild-type ligation products compared to natively expressed wtGFP and 50-GFP α A206K.....	152

List of Tables

Chapter 2		
2.1	Primers used to generate circular permutants.....	30
Chapter 3		
3.1	Protein and peptide amounts used in native chemical ligation reaction.....	93
Chapter 4		
4.1	Protein and peptide amounts used in native chemical ligation reactions.....	123
Chapter 5		
5.1	Overall yields of peptide thioesters generated for native chemical ligation.....	146
5.2	Protein and peptide amounts used in native chemical ligation reactions.....	147

Acknowledgements

This thesis would not have been complete without the tireless efforts of Christina L. Melief in editing several drafts of this thesis. Thank you for making sure all of it was there.

My advisor, Peter J. Tonge, for the freedom and financial support to attempt this daunting project and see it to completion.

My thesis committee; Dr. Erwin London, Dr. Carlos Simmerling, Dr. James Bliska, and Dr. Daniel Raleigh for the patience in meetings and discussions when this project was not going as well as it should.

My fellow fluorescent protein biochemists;

Dr. Debbie Stoner-Ma; for introducing me to GFP and teaching me the basics.

Dr. Andy Jaye; from whom I have learned much about Raman spectroscopy and yet, not enough.

Allison Haigney; for technical help in generating fluorinated tyrosines and support through the long lab hours.

Richard Brust; for contributions in optimizing the folding of GFP.

Tae Soo Kim; for contributions in the folding of GFP and its kinetics.

Other Tonge Lab Members, past and present;

Dr. Nicole Kruh, for the cloning techniques and general molecular biology help.

Dr. Carl Machutta, for troubleshooting, creative thinking, and logical analysis of many aspects of this project.

Dr. Todd Sullivan, for the start in organic synthesis

Chris Am Ende, for technical and practical synthetic advice

Dr. Gopal Bommineni, for technical aid in the synthesis of amino acid adducts

Janine Borgaro and Kanishk Kapilashrami for general lab help and resources.

Mentors

Dr. Richard Fluck, for believing in me and starting me down a career in science.

Drs. Janet Fischer and Peter Fields, for my first start as an independent scientist.

Mrs. Sobkow and Mrs. Reed, my 10th grade chemistry and 7th grade science teacher, respectively. Thank you for nurturing the minds of many budding scientists.

Chapter 1

Introduction to Green Fluorescent Protein and Methodologies of Study

History of Green Fluorescent Protein.....	2
Structure of GFP and Proposed Biosynthetic Pathways.....	5
Unnatural amino acids and circular permutation as tools for Elucidation of chromophore biosynthesis.....	13
Aim of Study.....	19
References.....	21

History of Green Fluorescent Protein

The Green Fluorescent Protein (GFP) from the Pacific jellyfish *Aequorea victoria* has proven a revolutionary biotechnological tool for the study of proteins and protein dynamics within the cell. GFP was originally discovered and isolated from jellyfish that appear green in the 1960's by Osamu Shimomura in complex with its *in vivo* Förster resonance energy transfer (FRET) partner, aequorin. Squeezates of the isolated rim of the jellyfish containing the green coloring were found to be a single protein complex [1]. When isolated and treated with calcium, this protein complex emitted green fluorescence at 508 nm. Subsequently, GFP was found to be the source of the green fluorescence emission at 508 nm; it does so by absorbing emitted 466 nm light from excited complex partner aequorin (Figure 1.1) [1]. Isolation of GFP and

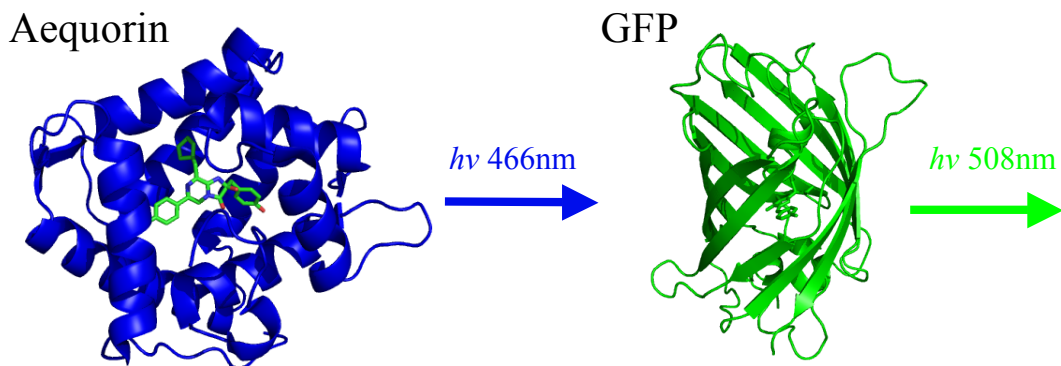


Figure 1.1 Photocomplex within *A. Victoria*. Aequorin (blue) binds coelenterazine and with the addition of O_2 , covalently binds. Addition of Ca^{2+} to this aequorin complex liberates 466 nm light, which is absorbed by GFP and in turn emits light at 508 nm.

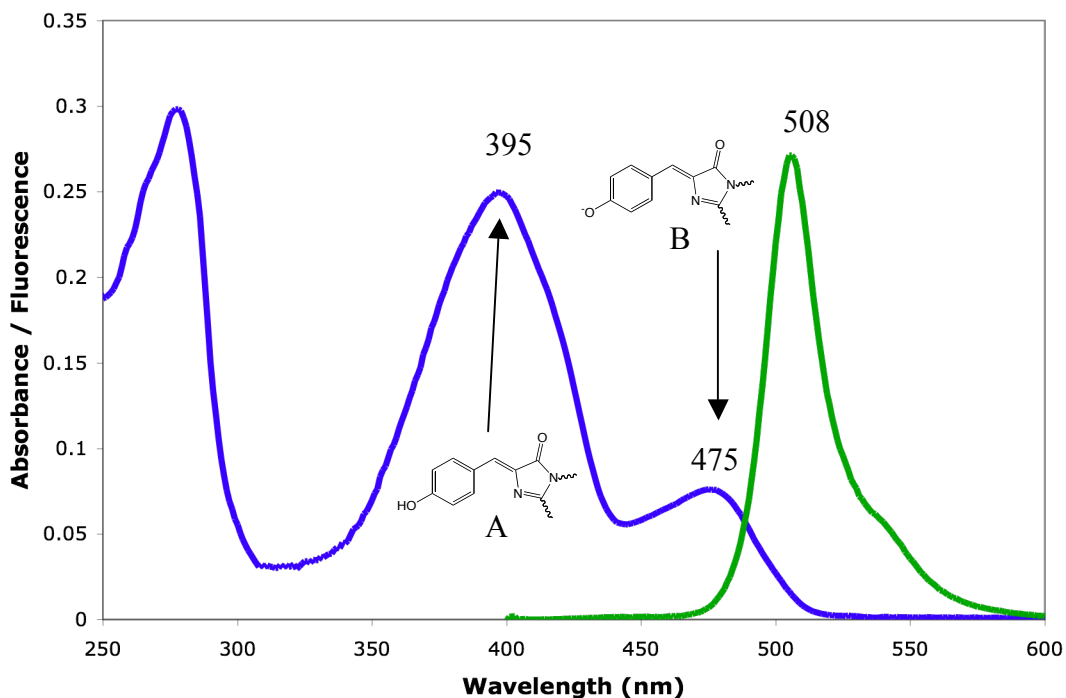


Figure 1.2: Absorption spectrum of GFP. The absorption spectrum of GFP shows a strong absorption at 395 nm (ϵ 21,000 $M^{-1}cm^{-1}$) and at 475 nm (ϵ 7,150 $M^{-1}cm^{-1}$)[2], overlapping with aequorin 466 nm emission, then strongly emits at 508 nm (fluorescence signal in arbitrary units scaled to absorption spectra). The two absorption bands correspond to a neutral 4-hydroxybenzylidene imidazolinone of GFP (A) that becomes ionized upon excitation (B).

absorption characterization of the protein revealed two absorption bands at 395 nm and 475 nm, overlapping well with the emission of aequorin (Figure 1.2) [3]. These observations were based on UV visible absorption spectra but did not offer insight into the structural interaction between aequorin and GFP, nor how GFP emitted the fluorescence.

Since GFP is from a biological source, great interest was invested in the biologically-generated chemical structure of the light-emitting molecule within GFP. Proteolysis of heat denatured GFP, chemical extraction, and repeated rounds of thin

layer chromatography (TLC) purification revealed a compound that, after undergoing acid hydrolysis, corresponded by mass and absorption maxima to the conjugated 4-(*p*-hydroxybenzylidene)-5-imidazolone chromophore [4]. Further work by Prasher *et al.* showed that when the GFP gene was isolated from the *A. victoria* cDNA and expressed heterologously, the green fluorescence of GFP remained, proving that the fluorescent nature of this protein was intrinsic to the 238 amino-acid primary sequence [5]. Several years later, Chalfie *et al.* harnessed the intrinsic fluorescence of GFP as a genetically encoded molecular tracer for study in *Caenorhabditis elegans* by first heterologously expressing the protein in *Escherichia coli* [6]. The same year, the heterologous expression potential of this protein was confirmed by Inouye *et al* [6, 7]. Since this initial work, the heterologous expression of GFP has exploded to almost every facet of cellular biology as a non-invasive *in vivo* visualization tool [8]. As a result, many have extensively studied and modified the original GFP structure for improved spectral characteristics and function.

Structure of GFP and Proposed Biosynthetic Pathways

The crystal structure of wild-type GFP (wtGFP) was solved in 1996 and revealed a 238 amino-acid beta-barrel with a central strained helix containing the chromophore (Figure 1.3A) [9]. Surrounding the chromophore is a network of amino-acid side chains (Gln69, Arg96, His148, Thr203, Ser205, and Glu222) that sequester the chromophore from bulk solvent and stabilize the excited state anion form, resulting in high quantum emission yield upon excitation (Figure 1.3B) [9, 10]. The crystal structure confirms the chemistry of the chromophore as a post-translational modification of centrally located Ser65-Tyr66-Gly67 and place the chromophore in the middle of the GFP amino-acid sequence, as was postulated from the primary sequence of the protein after a cDNA isolate from *A. victoria* was cloned and sequenced [5, 11]. Additionally, a hydrogen-bonding network was observed between the phenol of Tyr66 with ordered water adjacent to the chromophore, Ser 205 and Glu222 (Figure 1.3B) [12].

A.



B.

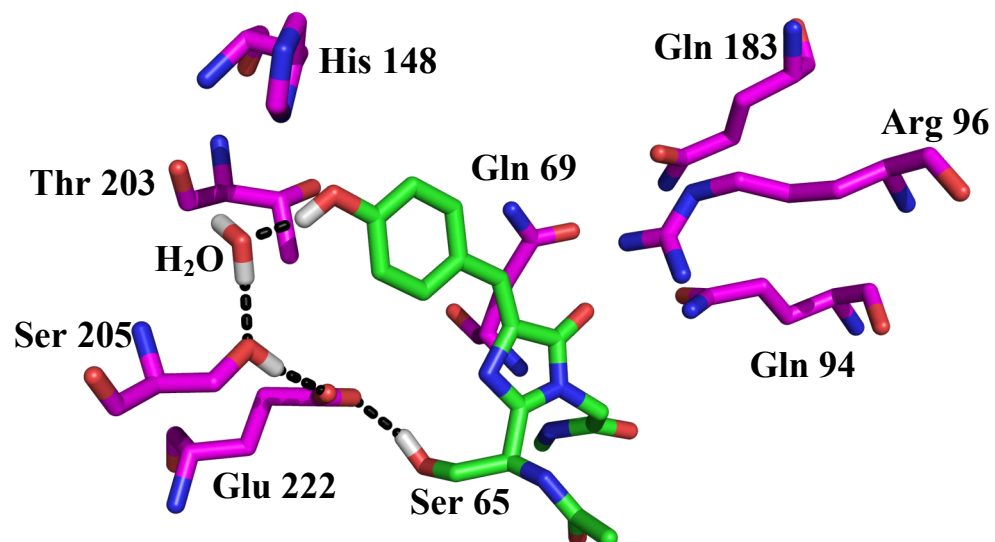


Figure 1.3. X-ray crystal structure of wtGFP and its chromophore. Structures are based on the first published structure of wtGFP, PDB ID 1GFL. Residues that surround and interact with the chromophore are in magenta, with an intrinsic hydrogen bonding network depicted as dashed black lines.

Time-resolved fluorescence (TRF) studies allow insight into the dynamics of the chromophore during excitation by monitoring the emission of fluorescence over the picosecond to nanosecond time scale. GFP emits a short 460 nm emission correlated to the emission of the neutral form of the chromophore, which dissipates within 30 ps while rise and emission at 509 nm from the anionic form of the chromophore lasts on the order of nanoseconds [13-15]. These results suggest multiple chromophore charge states within the hydrogen-bonding network of the protein.

This idea of charge states and cycling of GFP absorption and emission were interpreted to be from three different states [12, 15]. Under this model, the steady state neutral chromophore (A) upon excitation (A*) can directly emit 460 nm light and relax back to the neutral chromophore (A) or undergo an excited state proton transfer (ESPT) resulting in the formation of a phenolate in the excited chromophore (I*). I* then emits 509 nm light to return to ground state (I) and cycles back to the neutral steady state (A) by quick reprotonation. A small population of the chromophore exists in the anionic ground state (B), where excitation at 476 nm leads to an excited B* state, which then can emit 504 nm light and return to ground B state. The difference between A and B state is proposed to be a structural rearrangement of the hydrogen bonding network. However, at very slow rate, Chatteraj *et al.* suggested that in both the excited and ground state, I, which contains the phenolate on Tyr66, can be converted to the steady state anion (B) by rearrangement of this hydrogen bonding network (Figure 1.4) [15].

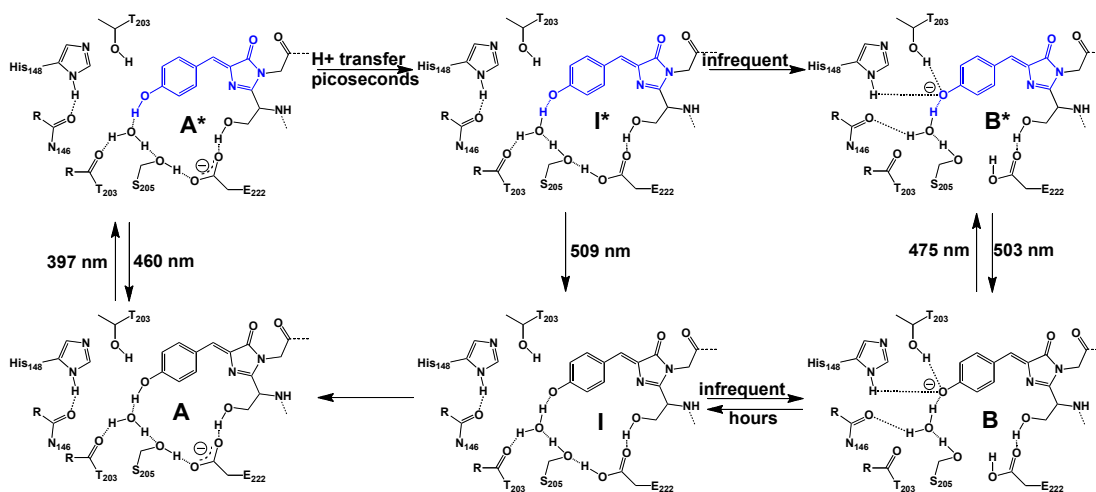


Figure 1.4. Photocycle of GFP. Excited state chromophores detailed in blue, ground state chromophores detailed in black. (Structures adapted from Stoner-Ma *et al.* [16] and based on charge states detailed in Chatteraj *et al.* [26].)

As would be expected, GFP chromophore spectral characteristics are pH dependent even though the protein is stable between pH 6-10 [17]. For this reason, GFP may not only be a tracer or tag, but also a tightly defined physiological pH sensor [18]. The phenol of Tyr66 can be ionized to form the phenolate at pKa 8.1, which sharply changes the maximal absorption from 384 to 448 nm [19]. The anionic variant of GFP, S65T is pH titratable and adopts a neutral form in acidic solutions [20].

Mutations of residues in and around the chromophore of GFP showed that the absorption and emission can be altered to differing wavelengths. The S65T mutation shifts the absorption spectrum to the anionic form, presumably by disrupting the hydrogen bonding to Glu222, thereby forcing this Glu residue to remain neutral and permitting the chromophore phenol to form an anion more easily through the

hydrogen bonding network. However, in wtGFP, the Ser65 maintains this Glu as an anion, thereby promoting a neutral species on the phenol of the chromophore through the hydrogen bonding network [2, 21-23]. Interestingly, this mutation causes a four-fold increase in the rate of chromophore formation [2, 24]. Driving the absorption spectrum to the neutral form (λ max 399 nm), the T203I mutant disrupts hydrogen bonding to the phenol of the chromophore thereby promoting a neutral chromophore that may still ionize in the excited state and emit at 511 nm [25, 26]. Mutations at residue Thr203 to an aromatic amino acid, most notably tyrosine, creates a π -stacking interaction with the phenol in the chromophore that shifts both the absorption and emission spectra to the red approximately 20 nm, thereby creating yellow fluorescent protein [9, 27-29]. Mutation of the Tyr66 to other aromatic amino acids shifts the excitation and emission into the blue range; Y66W (λ max 436 nm, emission 476 nm), Y66H (λ max 383 nm, emission 447 nm), and Y66F (λ max 360 nm, emission 442 nm) that are often used in cellular studies as FRET partners with green or yellow emitting variants of GFP [26, 30, 31].

With the identification of the residues that are post-translationally modified to form the chromophore and molecular oxygen being the only required external cofactor, two models for the formation of the GFP chromophore based on biochemical and crystallographic studies of both wtGFP and created variants were proposed. Prior to extensive mutant structural and biochemical studies, the pathway was thought to be a simple spontaneous backbone cyclization event coupled with the dehydration of the serine carbonyl, followed by slow oxidation over a four hour

period (Figure 1.5A) [26]. The first model generated from mutant GFP studies, proposed by the Getzoff group, suggested an initial cyclization of the peptide backbone when the amide of Gly67 is brought in close proximity to the carbonyl of Ser65 in a strained alpha helix during folding of the protein. Subsequent dehydration of the cyclized product would create a semi-stable intermediate, which is then oxidized by molecular oxygen via hydroperoxide intermediate generating the mature chromophore (Figure 1.5B) [24, 26, 30]. The proposed rate-limiting step of this pathway is oxidation of the dehydrated cyclized intermediate. This model arose from the crystal structure of the S65A/Y66S mutant of GFP. These mutations mimic the post-translational modification of Histidine Amomia Lyase (HAL), which also carries out a backbone cyclization and dehydration event between Gly144 and Ala142. A lack of distinct electronic density around the carbonyl carbon of Ser65 in GFP suggests that these residues are in a mixed population of hydrated and dehydrated species [32, 33]. These results suggest that the dehydration mechanism is fast and reversible, leaving the slow oxidation at the C α of Tyr66 as the rate limiting final step in chromophore maturation for these mutants.

The crystal structures of the Y66L variant of GFP suggests an alternate pathway, as proposed by the Wachter group. This structure shows a hydrated species

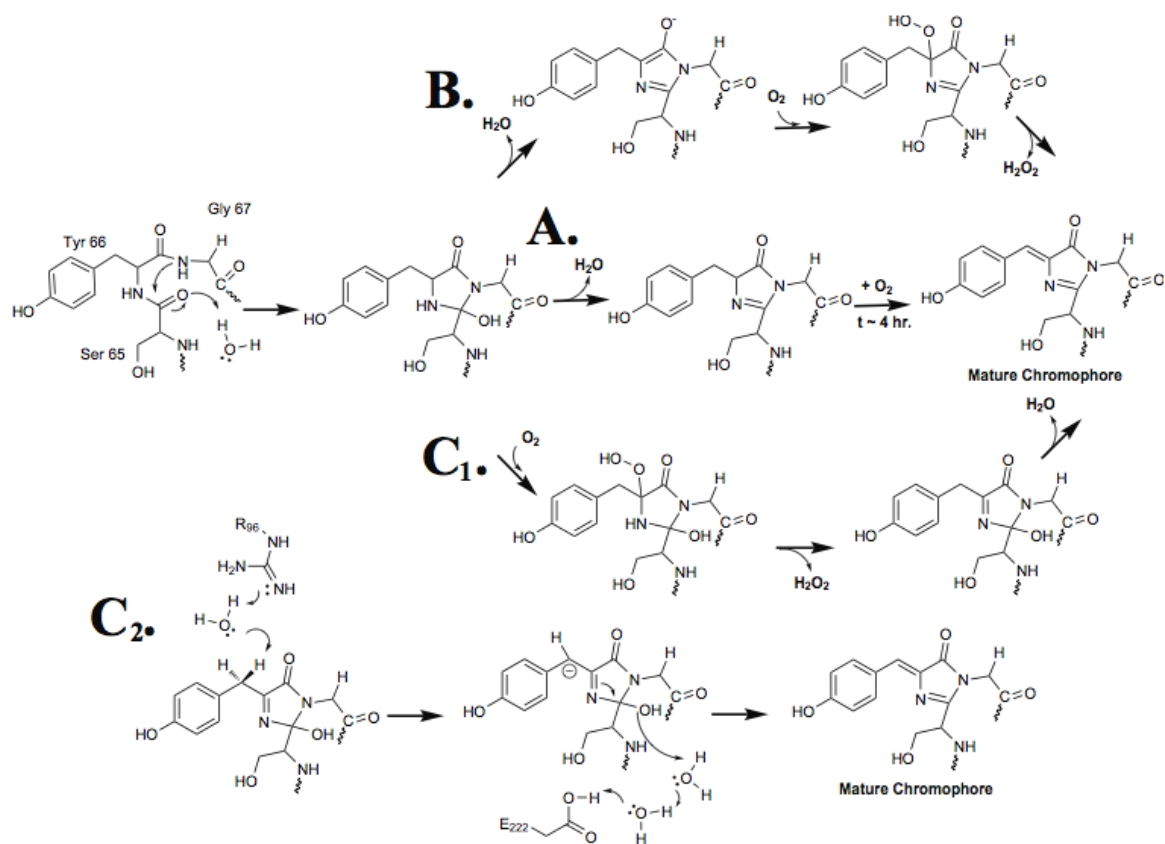


Figure 1.5. Current models of chromophore formation.

A. Initial models of chromophore formation suggest a cyclization event followed by dehydration to a cyclic imine. Slow aerobic oxidation yields the final conjugated chromophore. **B.** the model developed from the crystal structure of S65A/Y66S shows backbone cyclization and a reversible dehydration step. Oxidation of the dehydrated species commits the structure to the final conjugate chromophore. **C₁.** an alternate model based on the Y66L GFP variant shows a cyclization, oxidation, and dehydration mechanism **C₂.** KIE studies of a F64L/S65T/E222Q mutant show that after oxidation, the final dehydration step is stimulated by proton abstraction from C β of the tyrosine moiety, causing internal rearrangement of the imidazolynl ring and ejection of the hydroxyl group.

with a “fully reduced imidazole with an attached carbonyl oxygen, analogous to the imidazolinone ring of the mature chromophore (partially reduced imidazole with an attached carbonyl oxygen)” [34]. Additionally, Fourier transform ion cyclotron resonance electrospray ionization mass spectrometry (FT-ICR-ESI-MS) confirmed the protein retains the hydrated form of the chromophore with a mass addition of 18 Da [34]. These data suggests a cyclization-oxidation-dehydration mechanism for the formation of the mature GFP chromophore (Figure 1.5C₁). Further experiments in which the C β of Y66 is deuterated in a F64L/S65T/F99S/M153T/V163A/A206K mutant show a Kinetic Isotope Effect (KIE) where replacement of the hydrogens with deuterium shows a slowing of carbon-deuterium scission relative to carbon-hydrogen scission only after the peroxide species is formed. These data suggested an Arg96-base mediated proton abstraction from C β of the tyrosine after hydroxylated cyclic imine formation (due to oxidation, peroxide leaving) [35]. This step is succeeded by imine rearrangement to eject the hydroxide moiety thereby forming the final dehydrated mature chromophore (Figure 1.5C₂) [35].

As detailed above, the models for chromophore formation proposed thus far are based structural and biochemical studies of mutants, not the wild-type chromophore sequence itself. To remedy the issue of multiple chromophore formation models from an array of mutants, smaller changes to the chromophore forming residues, such as incorporation of unnatural amino acids, can be made. These more subtle changes allow examination of each step in chromophore formation within the context of the wild type protein structure.

Unnatural amino acids and circular permutation as tools for elucidation of chromophore biosynthesis

There are currently two available methods for unnatural amino acid substitution of residues within a protein sequence. One methodology, developed by the Schultz lab and aimed at expanding the genetic code, involves the re-engineering of cellular protein synthesis machinery to generate and use an orthogonal tRNA that matches the rarely used amber stop codon (TAG) (Figure 1.6A). This tRNA is charged by an engineered or evolved tRNA synthetase [36, 37]. The amber codon introduced into the target protein gene at the desired codon position allows incorporation of the unnatural amino acid. This methodology has been applied to GFP at amino acid positions both distant from the chromophore and at the critical Tyr66 position [38, 39]. The unnatural amino acids previously substituted for Tyr66 that modulated the absorption and fluorescent properties of GFP include *p*-amino-L-phenylalanine, *p*-methoxy-L-phenylalanine, *p*-iodo-L-phenylalanine, and *p*-bromo-L-phenylalanine. All of these modifications are limited to substitutions of the phenol oxygen of the tyrosine and modifications of the amino-acid side chain that influence only half of the GFP chromophore. Exploration of the nature of chromophore formation within GFP demands a more flexible synthetic system allowing modification of both the main

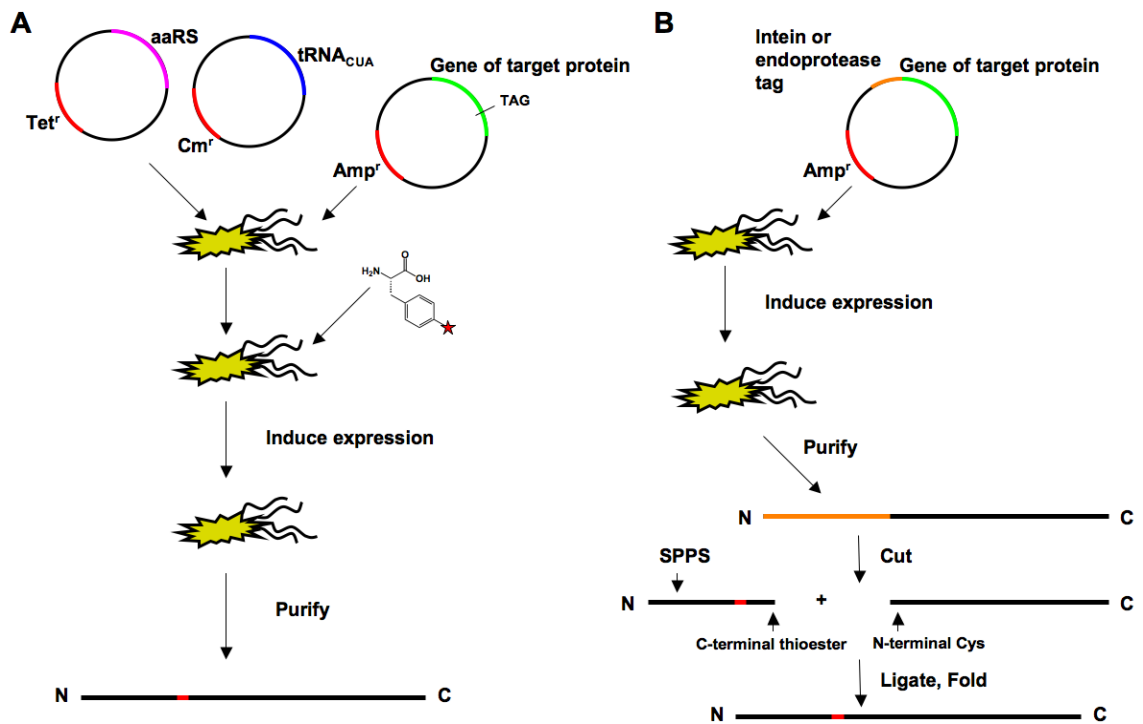


Figure 1.6. Method for incorporation of unnatural amino acids into proteins.

A. Exploiting the rarely used TAG amber codon, cells are transformed with plasmids encoding both the tRNA evolved to contain the anticodon matching the amber codon and a plasmid that encodes the tRNA synthetase capable of accepting the unnatural amino-acid and charging the evolved tRNA. The gene for the protein of interest, with the amber codon mutated in at the desired position, is transformed into these cells as well. Cells are then fed the modified amino acid (denoted here as a tyrosine derivative with differing functionalities at the *para* position), expression is induced, and protein is purified from cells. **B.** Plasmid containing the gene of the protein of interest with either an N-terminal intein or endoprotease recognition sequence is transformed into cells, expressed, and purified. The protein is then cut by addition of an endoprotease or through the action of the tagged intein. Through SPPS, the first section of the desired protein is synthesized as a C-terminal thioester peptide with the modified amino acid incorporated at the desired position. The cut protein, with a requisite N-terminal cysteine, is then ligated to this thioester peptide and the ligation product is folded to yield the desired protein.

chain backbone and the side chain, thereby allowing modification to both the imidazolinone and *p*-hydroxybenzylidene portion of the chromophore.

A second and more amenable method for unnatural amino-acid incorporation involves a semi-synthetic approach where a majority of the protein is bacterially expressed with an N-terminal cysteine (Figure 1.6B). The portion carrying the unnatural amino-acid is synthetically incorporated into a peptide thioester via solid phase peptide synthesis (SPPS), and the two pieces are ligated together via an intermolecular transthioesterification and rearrangement [40]. This reaction is carried out using thiol catalyst to promote the transesterification reaction. This reaction can be performed under native conditions if the site of ligation is solvent accessible, but may also be carried out in denaturing solution (usually $\geq 6\text{M}$ guanidine•HCl). This methodology, called expressed protein ligation (EPL), has the advantage of allowing a much broader range of chemical modification to the amino-acids incorporated into the synthetic peptide. Due to the greater chemical flexibility, we have chosen the EPL method to generate synthetic peptide thioesters with modified residues at the chromophore forming positions, and ligate them to the remainder of the GFP protein. The ligation product is then folded and the resulting chromophore structure may be studied by a variety of techniques including UV-visible absorption spectroscopy, mass spectroscopy, ground state raman spectroscopy, x-ray crystallography, steady state, time-resolved fluorescence, and infrared spectroscopy.

Since the chromophore forming residues of GFP are in the center of the protein's primary sequence, a method to shift the start of the primary sequence closer to these residues via a circular permutant was developed. By starting the primary sequence within a few amino acids of the chromophore forming residues, only a small portion of the protein must be truncated at the N-terminus to eliminate the chromophore forming residues up to the first cysteine within the sequence. This truncated portion of peptide sequence is then synthetically generated, incorporating the desired unnatural amino acid at the correct position, and ligated to the truncated circular permutant at the cysteine. This ligated protein is then *de novo* folded, yielding the GFP protein with the unnatural amino acid tyrosine at the Tyr66 position.

Based on the mechanisms involved in chromophore formation, two unnatural amino acids that modify the protein backbone were chosen to interrupt and probe the chromophore formation mechanism. By attaching a methyl group to the backbone amine of the tyrosine (N-methyl tyrosine), this amine is present in a tertiary form, thereby preventing further unsaturation of the backbone Ser65 carbonyl carbon-Tyr 66 amine nitrogen bond that occurs when the carbonyl of Ser65 is dehydrated. This modification therefore traps the chromophore in a hydrated state, allowing an investigation via mass spectroscopy, as was done with anaerobic and aerobic samples of wtGFP, of the oxidation mechanism at the C α of Tyr66 (i.e. is dehydration required to permit oxidation and double bond formation between C α and C β of the Tyr66 side chain) [26]. Likewise, the oxidation addition at the C α of Tyr66 is proposed to proceed via a carbonyl enolate at the Tyr66 carbonyl [35]. Removal of

this carbonyl oxygen would prevent enolate formation and oxidation at the adjacent C α of Tyr66, thereby evaluating whether the dehydration step (mass loss of 18 Da via mass spectrometry) can happen prior to and independent of the oxidation step. The removal of the carbonyl group of Tyr66 may even prevent the initial cyclization reaction, as some models propose an enolate at this position for adjacent amine of Gly67 to attack the carbonyl of Ser65 [32]. Any post translational modification observed by mass change in the protein from the expected sequence would indicate that cyclization is possible without an enolate at the carbonyl of Tyr66. Since these modifications may not produce a detectible absorption or fluorescence signal due to interrupted chromophore formation, mass spectral analysis will glean as much information about the intermediate chromophore formation state as possible. Ideally, crystal structures of the unnatural amino acid incorporated into the GFP scaffold would provide the clearest image of the trapped chromophore intermediate.

Other proposed amino acid modifications to the chromophore tyrosine would modulate the photochemistry of the chromophore. The semi-synthetic method proposed here allows for site-specific incorporation of tyrosine derivatives, thereby avoiding any effects of global incorporation of modified amino acids in all 11 tyrosine residues in GFP [20, 41]. Movement of the *p*-hydroxyl group of the tyrosine around the ring to the *meta* position, thereby disrupting the hydrogen bonding network of the chromophore within the wtGFP context, while still retaining all of the functional groups on the chromophore, would test the importance of *p*-hydroxyl position on the chromophore. Presumably, a chromophore with a hydroxyl group

outside of the hydrogen bonding network would behave as a completely neutral species, with absorption only at 397 nm (or slightly less) and potentially little fluorescence, unless the excited state *meta* phenol is able to eject the proton onto another acceptor residue. The isosteric substitution of hydrogens for fluorine atoms in the phenol ring of the chromophore would reduce the pKa of the phenol of the tyrosine by electron withdrawing inductive effects, thereby forcing an anionic chromophore into the wild-type GFP hydrogen-bonding network [20, 42]. This would most likely rearrange the hydrogen bonding network to accommodate the easily ionized chromophore and produce a predominantly anionic absorption spectrum. pH titrations monitored via fluorescence of this fluorinated tyrosine derivative will reveal the extent to which the pKa of the phenol tyrosine is modulated. Additionally, the use of time resolved fluorescence would give insight into some of the fast changes (on the picosecond timescale) within the fluorescence spectrum of GFP and this fluorinated derivative where presumably, the ionization of the chromophore phenol would most likely populate the B-state of the hydrogen bonding network.

Aim of Study

While the development of GFP as a biomolecular visualization tool has progressed extensively since its initial characterization and discovery, the scientific debate over the exact chemical mechanisms that lead to the formation of the chromophore is ongoing. Exploration the chemical space within the chromophore of GFP and dissection the chromophore forming steps requires a methodology that allows full chemical access to the residues that comprise the chromophore. While the tRNA-amber codon method for the introduction of unnatural amino-acids is an attractive means of introducing amino-acid side chains that are compatible with this system, the relatively limited space within the GFP barrel excludes the use of a large majority of amber codon compatible unnatural amino-acids. Additionally, the imidazolinone ring of the chromophore is derived from the amino-acid backbone. Therefore, main-chain modified amino acids useful in exploring the modifications to the imidazolinone ring would most likely not be functionally incorporated by the tRNA synthetase onto the tRNA or be successfully used by the ribosome. For these reasons, we chose to develop a purely synthetic method to introduce unnatural amino-acids (including those that modify that standard peptide bond in the residues that will compose the imidazolinone ring) into a peptide and then ligate the peptide to the remaining, bacterially expressed, protein. The use of this methodology will probe the mechanism of chromophore formation and potentially improve upon the existing

models, thereby giving a greater understanding to the formation of this chromophore as a whole. This method also allows a much broader set of modifications to the chromophore structure. Modifications to the tyrosine residue, including relocation of the tyrosine hydroxyl and fluorination of the tyrosine phenol that decreases the pKa of the phenol group, probe the importance of the hydrogen bonding network responsible for efficient fluorescent emission in the GFP protein.

References

1. 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.
2. Heim, R., A.B. Cubitt, and R.Y. Tsien, *Improved green fluorescence*. Nature, 1995. **373**(6516): p. 663-664.
3. Morise, H., et al., *Intermolecular energy transfer in the bioluminescent system of Aequorea*. Biochemistry, 1974. **13**(12): p. 2656-62.
4. Shimomura, O., *Structure of the chromophore of Aequorea green fluorescent protein*. Febs Lett., 1979. **104**(2): p. 220-222.
5. Prasher, D.C., et al., *Primary structure of the Aequorea victoria green-fluorescent protein*. Gene, 1992. **111**(2): p. 229-33.
6. Chalfie, M., et al., *Green fluorescent protein as a marker for gene expression*. Science, 1994. **263**(5148): p. 802-5.
7. Inouye, S. and F.I. Tsuji, *Aequorea green fluorescent protein. Expression of the gene and fluorescence characteristics of the recombinant protein*. Febs Lett., 1994. **341**(2-3): p. 277-80.
8. Martin, C., *GREEN FLUORESCENT PROTEIN*. Photochemistry and Photobiology, 1995. **62**(4): p. 651-656.
9. Ormo, M., et al., *Crystal Structure of the Aequorea victoria Green Fluorescent Protein*. Science, 1996. **273**(5280): p. 1392-1395.
10. Yang, F., L.G. Moss, and G.N. Phillips, Jr., *The molecular structure of green fluorescent protein*. Nat. Biotechnol., 1996. **14**(10): p. 1246-51.
11. Cody, C.W., et al., *Chemical structure of the hexapeptide chromophore of the Aequorea green-fluorescent protein*. Biochemistry, 1993. **32**(5): p. 1212-1218.

12. Brejc, K., et al., *Structural Basis for Dual Excitation and Photoisomerization of the Aequorea victoria Green Fluorescent Protein*. Proc. Natl. Acad. Sci. U.S.A., 1997. **94**(6): p. 2306-2311.
13. Jaye, A.A., et al., *Time Resolved Emission Spectra of Green Fluorescent Protein*. Photochem. Photobiol., 2006 **82**(2): p. 373-379.
14. Lossau, H., et al., *Time-resolved spectroscopy of wild-type and mutant Green Fluorescent Proteins reveals excited state deprotonation consistent with fluorophore-protein interactions*. Chemical Physics, 1996. **213**(1-3): p. 1-16.
15. Chatteraj, M., et al., *Ultra-fast excited state dynamics in green fluorescent protein: multiple states and proton transfer*. Proc. Natl. Acad. Sci. U.S. A., 1996. **93**(16): p. 8362-7.
16. Stoner-Ma, D., et al., *Proton relay reaction in green fluorescent protein (GFP): Polarization-resolved ultrafast vibrational spectroscopy of isotopically edited GFP*. J. Phys. Chem. B., 2006. **110**(43): p. 22009-18.
17. Ward, W., W., et al., *Spectral Perturbations of the Aequorea Green-Fluorescent Protein*. Photochem. Photobiol., 1982. **35**(6): p. 803-808.
18. Bizzarri, R., et al., *Development of a novel GFP-based ratiometric excitation and emission pH indicator for intracellular studies*. Biophys. J., 2006. **90**(9): p. 3300-14.
19. Ward, W.W. and S.H. Bokman, *Reversible Denaturation of Aequorea Green-Fluorescent Protein: Physical Separation and Characterization of the Renatured Protein*. Biochemistry, 1982. **21**: p. 4535 - 4540.
20. Pal, P.P., et al., *Structural and Spectral Response of Aequorea victoria Green Fluorescent Proteins to Chromophore Fluorination*. Biochemistry, 2005. **44**(10): p. 3663-3672.
21. Cheng, L., et al., *Use of green fluorescent protein variants to monitor gene transfer and expression in mammalian cells*. Nat. Biotechnol., 1996. **14**(5): p. 606-9.
22. Cormack, B.P., R.H. Valdivia, and S. Falkow, *FACS-optimized mutants of the green fluorescent protein (GFP)*. Gene, 1996. **173**(1 Spec No): p. 33-8.
23. Delagrave, S., et al., *Red-shifted excitation mutants of the green fluorescent protein*. Biotechnology (NY), 1995. **13**(2): p. 151-4.

24. Reid, B.G. and G.C. Flynn, *Chromophore formation in green fluorescent protein*. *Biochemistry*, 1997. **36**(22): p. 6786-91.
25. Ehrig, T., D.J. O'Kane, and F.G. Prendergast, *Green-fluorescent protein mutants with altered fluorescence excitation spectra*. *Febs Lett.*, 1995. **367**(2): p. 163-6.
26. 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. 12501-12504.
27. Miyawaki, A., et al., *Dynamic and quantitative Ca²⁺ measurements using improved cameleons*. *Proc. Natl. Acad. Sci. U.S.A.*, 1999. **96**: p. 2135 - 2140.
28. Miyawaki, A., et al., *Fluorescent indicators for Ca²⁺ based on green fluorescent proteins and calmodulin*. *Nature*, 1997. **388**(6645): p. 882-7.
29. Miyawaki, A., *Fluorescent proteins in a new light*. *Nat. Biotechnol.*, 2004. **22**(11): p. 1374-6.
30. Cubitt, A.B., et al., *Understanding, improving and using green fluorescent proteins*. *Trends Biochem. Sci.*, 1995. **20**(11): p. 448-455.
31. Pollok, B.A. and R. Heim, *Using GFP in FRET-based applications*. *Trends Cell Biol.*, 1999. **9**(2): p. 57-60.
32. Barondeau, D.P., et al., *Understanding GFP Chromophore Biosynthesis: Controlling Backbone Cyclization and Modifying Post-translational Chemistry*. *Biochemistry*, 2005. **44**: p. 1960-1970.
33. Barondeau, D.P., et al., *Mechanism and Energetics of Green Fluorescent Protein Chromophore Synthesis Revealed by Trapped Intermediate Structures*. *Proc. Natl. Acad. Sci. U.S.A.*, 2003. **100**(21): p. 12111-12116.
34. Rosenow, M.A., et al., *The crystal structure of the Y66L variant of green fluorescent protein supports a cyclization-oxidation-dehydration mechanism for chromophore maturation*. *Biochemistry*, 2004. **43**(15): p. 4464-72.
35. Pouwels, L.J., et al., *Kinetic isotope effect studies on the de novo rate of chromophore formation in fast- and slow-maturing GFP variants*. *Biochemistry*, 2008. **47**(38): p. 10111-22.
36. Xie, J. and P.G. Schultz, *Adding amino acids to the genetic repertoire*. *Curr. Opin. Chem. Biol.*, 2005. **9**(6): p. 548-54.

37. Xie, J. and P.G. Schultz, *An expanding genetic code*. *Methods*, 2005. **36**(3): p. 227-38.
38. Liu, W., et al., *Genetic incorporation of unnatural amino acids into proteins in mammalian cells*. *Nat. Meth.*, 2007. **4**(3): p. 239-44.
39. Wang, L., et al., *Unnatural amino acid mutagenesis of green fluorescent protein*. *J. Org. Chem.*, 2003. **68**(1): p. 174-6.
40. Muralidharan, V. and T.W. Muir, *Protein ligation: an enabling technology for the biophysical analysis of proteins*. *Nat. Meth.*, 2006. **3**(6): p. 429-438.
41. Khan, F., et al., *¹⁹F NMR Studies of the Native and Denatured States of Green Fluorescent Protein*. *J. Am. Chem. Soc.*, 2006. **128**(33): p. 10729-10737.
42. Kim, K. and P.A. Cole, *Measurement of a Bronsted Nucleophile Coefficient and Insights into the Transition State for a Protein Tyrosine Kinase*. *J. Am. Chem. Soc.*, 1997. **119**(45): p. 11096-11097.

Chapter 2

Generation of GFP Circular Permutants and Characterization of Variants

Introduction.....	26
Methods.....	28
Results and Discussion.....	35
Summary.....	53
References.....	55

Introduction to approach and limitations of GFP circular permutants

To use a semi synthetic approach incorporating unnatural amino acids by native chemical ligation of a small peptide thioester to the GFP protein, we must first find and validate a circular permutant with the critical chromophore forming residues near the N-terminus that behaves as wild type GFP does. The chromophore forming residues of GFP are well within the middle of the protein sequence at residues 65 through 67. While direct truncation of a bacterially expressed portion of GFP is possible, generation of a thioester peptide longer than 50 amino-acids is technically difficult. In addition yields of purified peptide would be unacceptably low for the ligation reaction. Since a shorter peptide is required (less than 30 amino-acids), we will need to modify GFP so that the N-terminus is within an acceptable peptide length from the chromophore forming residues. Generating a new N-terminus in the middle of the protein sequence means linking the existing N- and C- termini so that a single amino-acid chain is maintained.

The generation of new N- and C- termini while linking the existing N- and C- termini, known as circular permutation, has already been accomplished for GFP with mixed results [1, 2]. Of the circular permutants generated in these studies, those close to the flexible loops at the ends of the GFP barrel were the most fluorescent in cells and soluble cell lysates [2]. These studies used a short, flexible hexapeptide linker

(GSGGTG and GGTGGS) that is just long enough to join the wild type N- and C-termini (distance from N to C in 1GFL structure wt is 22Å, hexapeptide sequence is 6 x 4.31Å or 25.86Å) [2]. Based on these results we chose to use a longer linker containing a hexahistidine purification tag between the wild type N- and C-termini and examine several sites close to the chromophore residues within the loops of GFP for new N- and C-termini.

Plasmids for circular permutants were constructed and the expressed protein was studied by UV/visible absorbance, fluorescence, Raman spectroscopy, and mass spectroscopy. The native chemical ligation step would most likely be carried out under strongly denaturing conditions and the ligation product would need to be folded. Therefore, the most promising circular permutant was expressed insolubly in cells, the insoluble inclusion body fraction purified, solublized in urea denaturant, and *de novo* folded [3]. Folding was monitored by rise in fluorescence from the *p*-hydroxybenzylidene chromophore as it forms within the properly folded protein.

Methods

Cloning and Mutagenesis

The *gfp* gene bearing the cycle 3 mutations (F99S, M153T, V163A) was obtained in the plasmid pIWT5563his (kindly provided by Dr. Andreas Plückthun, University of Zürich)[4]. The wild-type *gfp* gene with the superfluous Q80R mutation from the initial cloning out of *A. victoria* genomic DNA into the expression plasmid pRSETb was provided by Dr. Rebecca Wachter, Arizona State University (ASU). Restriction endonucleases were obtained from New England Biolabs and Invitrogen. DNA purification kits were purchased from Qiagen. The expression plasmid pET23b was obtained from EMD/Novagen. The NdeI restriction site contained within the *gfp* gene in the pIWT5563his plasmid was eliminated by silently mutating the codon for H77 from CAT to CAC using the primers 5' CCGTTATCCGGATCACATGAAACGGCATGAC 3' and 5' GTCATGCCGTTTCATGTGATCCGGATAACGG 3' via Quickchange mutagenesis (Stratagene) creating the plasmid pIWT5563his-NdeI. The GFP fragment (fragment 1) corresponding to amino acids 50 to 228 and the first 7 amino acids of the designed linker from the pIWT5563his-NdeI plasmid was amplified by polymerase chain reaction (PCR) from pIWT5563his by primers 5' GGAATTCCATATGACTGGAAAACCTACCTGTTCC 3',

5' CATGGTACCCGTTCCGCGTGGCACCAA 3'. Another GFP fragment (fragment 2) corresponding to amino acids 3 to 49 and the remaining 12 amino acids from the pIWT5563his plasmid was amplified by PCR from pIWT5563his-NdeI using the primers 5' GTAGGTACCGGATCTAGACATCATCACCAC 3', 5' GACGAGCTCTTATTAAGTGCAAATAAATTTAAG 3', both using pure Taq Ready-To-Go PCR Beads (GE Healthcare). The resulting dsDNA fragments were purified and digested with NdeI (New England Biolabs) and KpnI (New England Biolabs) for fragment 1 and KpnI and SacI (New England Biolabs) for fragment 2. The digested and purified fragments were ligated to each other for 10 minutes at 16°C using T4 DNA ligase (New England Biolabs) and then ligated into NdeI and SacI digested and purified pET23b overnight at 4°C. Ligation products were transformed into *Escherichia coli* strain XL1-Blue cells (Stratagene) and plated on ampicillin selective media 200 µg/mL. Resulting colonies were grown in liquid media (Luria bertani (LB) ampicillin 200 µg/mL) and plasmid DNA was isolated (Wizard Plus SV Miniprep system, Promega). The resultant plasmid was sequenced for the correct insert using a T7 promoter primer (ABI systems); yielding the plasmid p50-GFP. Similarly, the plasmids p48-GFP, p72-GFP, p82-GFP, and p88-GFP were constructed with corresponding primers (Table 2.1). Modifications to the linker region, on the periphery of the *gfp* gene in the pIWT5563his-NdeI plasmid were also generated using this process creating the plasmid p50-GFP α . The A206K mutation was created by Quickchange mutagenesis (Stratagene) using the primers 5' CTGTCGACACAATCTAAACTTTCGAAAGATCCC 3' and

5' GGGATCTTTCGAAAGTTTAGATTGTGTCGACAG 3' into the p50-GFP α plasmid after assessing the stability of this circular permutant.

Table 2.1. Primers used to generate circular permutants. Plasmids were constructed by generating two fragments of DNA from the piWT5563-his plasmid; Fragment 1 consisting of residues starting from the amino-acid listed in the circular permutant name (i.e. residue 48 for 48-GFP) through to the C terminus and half of the linker sequence, Fragment 2 consisting of residues starting with the rest of the linker through the beginning of the GFP sequence ending at the residue just before where Fragment 1 starts. Fragments 1 and 2 are cut with restriction endonucleases, ligated together, and then ligated into restriction endonuclease cut pET23b expression vector.

Circular Permutant	Fragment 1 Forward Primer
48-GFP	5 ' TGTCATATGTGCACTACTGGAAAACCTACCTGTTCCA3 '
50-GFP	5 ' GGAATTCCATATGACTGGAAAACCTACCTGTTCC3 '
50-GFP α	5 ' GGAATTCCATATGACTGGAAAACCTACCTGTTCC3 '
72-GFP	5 ' TGTCATATGTCCCGTTATCCGGATCACATGAAACGG3 '
82-GFP	5 ' TGTCATATGGACTTTTTCAAGAGTGCCATGCCCGAA3 '
88-GFP	5 ' TGTCATATGCCCGAAGTTATGTACAGGAACGCACT3 '
	Fragment 1 Reverse Primer
	5 ' CATGGTACCCGTTCCGCGTGGCACCAA3 '
	Fragment 2 Forward Primer
	5 ' GTAGGTACCGATCTAGACATCATCACCAC3 '
Circular Permutant	Fragment 2 Reverse Primer
48-GFP	5 ' GTGGAGCTCTTATTAAATAAATTTAAGGGTAAGCTT3 '
50-GFP	5 ' GACGAGCTCTTATTAAAGTGCAAATAAATTTAAG3 '
50-GFP α	5 ' CGGGGTACCAATCCCAGCAGCTGTTACAAACTC3 '
72-GFP	5 ' GTGGAGCTCTTATTAAAGCATTGAACACCATAAGA3 '
82-GFP	5 ' GTGGAGCTCTTATTAAATGCCGTTTCATGTGATCCGG3 '
88-GFP	5 ' GTGGAGCTCTTATTAGGCACTCTTGAAAAAGTCATG3 '

Protein Expression and Purification

Soluble protein expression was carried out by transforming plasmid into *E. coli* BL21(DE3)-pLysS cells (Stratagene) and growing in 500 mL of LB media shaking 250 RPM at 37°C supplemented with 200 µg/mL of ampicillin and 30 µg/ml of chloramphenicol to an OD₆₀₀ of ~0.7. After reducing the temperature of the media to 25°C, isopropyl-β-D-thiogalactopyranoside (IPTG) (Gold Biotechnology USA) was added to a final concentration of 0.8mM to induce cells to express protein overnight. Cells were harvested by centrifugation and stored at -20°C. After thawing, cells were resuspended in 40 mL per L cell culture 20mM sodium phosphate pH 7.9, 300mM NaCl (PBS buffer) supplemented with 10mM imidazole and lysed by passing through a french press 5 times. Lysate was cleared by centrifugation at 125,000 x g, passed through a 0.22 µm filter, and bound to a 5 mL bed volume of Ni²⁺-NTA resin (Qiagen). Column bound protein was washed and subsequently eluted with PBS buffer containing 20mM and 100mM imidazole, respectively. Protein fractions showing green fluorescence were pooled and dialyzed against PBS buffer to remove imidazole. Protein was concentrated using YM-10 centricon (Millipore) and characterized by UV/visible spectroscopy on a Cary-Bio 300 spectrophotometer (Varian) and buffer baseline corrected. Fluorescence measurements at 70nM protein were carried out on a Fluorolog-3 fluorimeter (Jobin Yvon-Spex) and buffer baseline corrected. Protein prepared for X-ray crystallography was further purified through a Q-sepharose column and concentrated to 15 mg/mL,

flash frozen in 60 μL aliquots, and sent to Dr. Rebecca Wachter (ASU) for crystallization and structural analysis. Raman spectroscopy was performed as detailed in Bell *et al*, 2000 [5]. Briefly, 70 μL of 200 μM protein was placed in 2 x 2 mm cell and scanned 80 times from 2400 cm^{-1} to 400 cm^{-1} within 8 seconds. An equal amount of buffer was then placed in the cell and scanning repeated. Using Grams/AI software (Thermo Scientific), the buffer spectrum was subtract from the protein spectra and the resulting spectra was baseline corrected.

Insoluble Protein Expression and *De Novo* Folding

Protein was expressed in inclusion bodies to high yield using protocols described by (Rosenow *et al.*) [6]. Briefly, plasmid was transformed into *E. coli* BL21(DE3) cells (Stratagene) and grown to mid-log phase (2 1/2 hrs). The incubation temperature was then raised to 42°C and IPTG was added to 0.8mM final. Cells were harvested after 3 hours by centrifugation at 5,000 x g for 20 minutes. Cells were resuspended and lysed by french press in 40 mL 50mM HEPES pH7.9, 300mM NaCl, 5mM β -mercaptoethanol per liter culture and centrifuged for one hour at 125,000 x g. The pelleted crude inclusion body preparation was successively resuspended and pelleted in 100mM Tris•HCl pH7.9, 500mM NaCl, 15mM EDTA, 5mM DTT four times, with 2% Triton X-100 present in the first resuspension. After a fifth resuspension in 50mM HEPES pH7.9, 50mM NaCl, 1mM DTT, the final inclusion body homogenate was flash frozen and stored as 1.5 mL aliquots at -80°C.

De novo protein folding was carried out by resuspending a thawed inclusion body preparation of 50GFP α A206K in 25mM MES pH 8.5, 8M Urea, 0.1mM DTT for one hour and rapidly diluting dropwise into 100 fold 50mM Tris•HCl pH 8.5, 500mM NaCl, 1mM DTT [7]. Dilute *de novo* folded protein was concentrated by Amcicon 10K MWCO pressure cell and/or Cetricon centrifugal 10K MWCO filters, filtered through 0.22 μ M PDVF syringe driven filters (Millipore), and dialyzed against PBS buffer. Protein was further purified by binding to Ni²⁺-NTA resin, washing with PBS buffer, eluting with 100mM imidazole in PBS, dialyzing against PBS, and concentrating by Centricon 10K centrifugal MWCO filters to needed concentrations. Alternatively, direct application of the dilute *de novo* folded sample to a Ni²⁺-NTA column, washing with PBS buffer, and elution with PBS buffer supplemented with 250mM imidazole yielded protein for characterization as well. Again, protein is dialyzed against PBS buffer and concentrated with centrifugal MWCO filters to needed concentrations or volumes. Chromophore formation was monitored by fluorescence at 508 nm using a Fluorolog-3 fluorimeter and absorption measurements were carried out on a Cary-Bio 300 spectrophotometer.

SDS-PAGE and Western Blotting

Protein samples were diluted in equal volume of 2x SDS-PAGE loading buffer (100mM Tris base pH 6.8, 4% SDS, 0.2% bromopheol blue, 200mM dithiotheritol) and boiled at 95°C for 10 minutes. Samples were then centrifuged at 14,000 RPM for 10 minutes and loaded into wells of a 15% polyacrylamide gel. Gels

were typically run for 2.5 hours at 100V and protein bands were visualized by rocking overnight in Coomassie brilliant blue R-250 stain solution. Excess Coomassie stain was removed with destain solution (50% methanol, 40% water, 10% acetic acid) and the gel digitally scanned.

Gels for western blots were transferred to a nitrocellulose membrane by 30V electrophoresis overnight in 25mM Tris base, 192mM glycine, 20% methanol, pH 8.3. Total protein loading was visualized by incubating in Ponceau staining solution (0.2% w/v Ponceau red stain in 3%w/v trichloroacetic acid) for 5 minutes. Excess stain was removed with washes of distilled deionized water and blocked by incubating in 5% non-fat dry milk in 1x TBS-T buffer (40mM Tris pH 7.6, 275mM NaCl, 0.1% Tween-20). Primary antibody anti-6XHis IgG (Sigma-Aldrich) was added in 1:5000 dilution and incubated at room temperature for 2 hours. The blot was washed with 3 rounds of 5% milk in TBS-T for 15, 5, and 5 minutes. The secondary alkaline phosphatase conjugated anti IgG antibody (Sigma-Aldrich) was incubated with the blot for 2 hours and washing was repeated as earlier. The blot was developed by adding 20 mL of 5-Bromo-4-Chloro-3'-Indolyphosphate p-Toluidine Salt/Nitro-Blue Tetrazolium Chloride (BCIP/NBT) (Thermo-Pierce Scientific) until bands developed (15 minutes), rinsed with distilled deionized water, and digitally scanned.

Results and Discussion

Expression and Evaluation of Circular Permutants

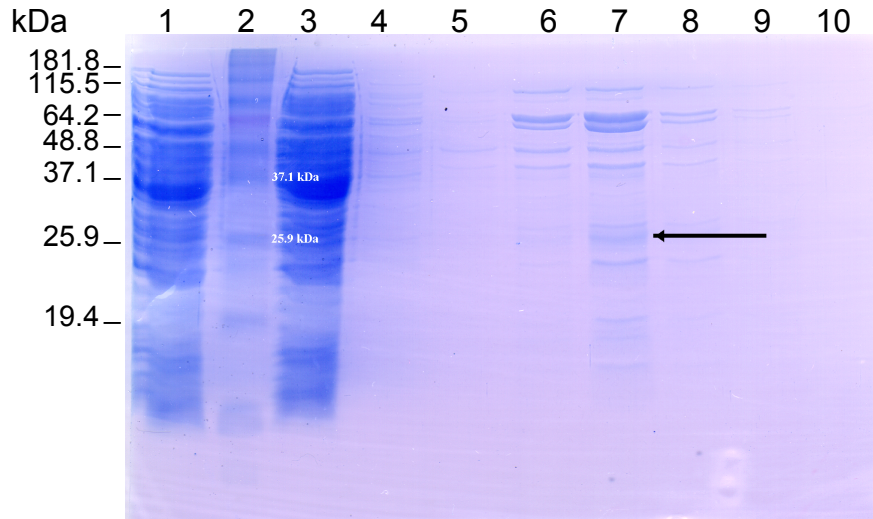
After construction of the circular permutant vectors, soluble expression of the resulting protein was preliminarily evaluated by visual inspection. To assess fluorescence of each permutant, each cell pellet was visually inspected while illuminated with shortwave ultraviolet light. Cell pellets exhibiting the characteristic green fluorescence of a fully formed *p*-hydroxybenzylidene imidazolinone chromophore within a protein barrel were purified; 48-GFP and 50-GFP. The other circular permutant constructs (72-GFP, 82-GFP, 88-GFP) did not express protein exhibiting green fluorescence and were excluded as circular permutant candidates. Both fluorescent circular permutants (48-GFP primary sequence beginning at Cys48 and ending at Ile47, 50-GFP primary sequence beginning at Thr50 and ending at Thr49; both permutants creating new termini in the loop that leads from one of the β -barrel strands into the central α helix) were then evaluated for expression, solubility, stability, and *de novo* folding efficiency to determine which would most suitable for subsequent experiments.

Initial experiments analyzing candidate expression and solubility were conducted by either Coomassie stained SDS-PAGE gel of whole cell lysate or soluble cell lysate, thereby evaluating both the overall expression efficiency and solubility of

the expressed permutant. Soluble cell lysates were then purified by Ni²⁺-NTA affinity chromatography to determine overall soluble expression and characterized by UV visible absorption. Purification of the first candidate, 48-GFP, showed only a minor absorption peak at 397 nm relative to the absorbance signal from protein (A₂₈₀) and co-purified with high molecular weight protein chaperones (Figure 2.1A and B). These results indicated protein instability. Attempts to improve protein folding and yield by induction at a lower temperature (16°C), longer inductions (48 hours), and adding additional cell nutrients (2% glucose) did not result in protein expression with improved chromophore formation (Figure 2.1A). Although 48-GFP was expressed in high yield and could be isolated as an inclusion body preparation, attempts to *de novo* fold protein from this cellular fraction did not produce matured chromophore. The inability of this permutant to produce large amounts of matured chromophore, either natively expressed or *de novo* folded, eliminated this permutant as a candidate for further experimental work.

The other fluorescing permutant, 50-GFP, with termini two amino acids further into the GFP primary sequence and directly bisecting the loop, expressed as well in cells as 48-GFP, but produced significantly more green colored protein in the soluble cell lysate fraction. Coomassie stained SDS-PAGE of a sample of the soluble cell lysate confirmed this soluble expression at the expected molecular weight. A Ni²⁺-NTA purified sample of 50-GFP showed an absorption spectrum that is similar to wtGFP, with only a slight variation in the absorption maximum of the anionic peak (Figure 2.2B). This discrepancy between absorption measurements is

A.



B.

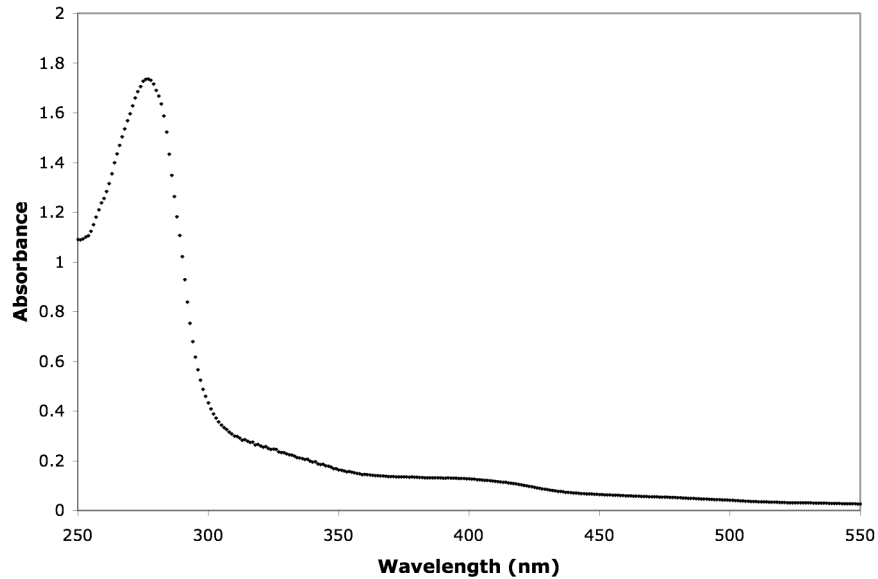


Figure 2.1. Purification Profile and Absorption Spectra of 48-GFP.

A. Ni²⁺-NTA purification of 48-GFP; (1) whole cell lysate supernatant, (2) Invitrogen Benchmark Ladder, (3) Ni²⁺-NTA column flowthrough, (4) 10mM imidazole buffer wash, (5) 20mM imidazole buffer wash, (6) to (10) elution with 100mM imidazole buffer. **B.** Absorption spectra of the purified 48-GFP shows a weak absorption point at 397 nm and a very large 280 nm peak indicative of very little chromophore formation relative to total protein within the sample.

matched in the fluorescence spectrum emission when irradiated at the anionic peak (475 nm) (Figure 2.2). The 50-GFP candidate had absorption spectra matching wtGFP and was expressed in soluble yields similar to wtGFP. To determine the orientation of the chromophore and the overall structure of 50-GFP as a circular permutant, large-scale samples were produced and sent to Dr. Rebecca Wachter at Arizona State University for protein crystallization and structural determination. These samples, upon shipment and storage at 4°C, degraded into two fragments, of approximately 20 kDa and 7 kDa. Anti-6xHis western blotting of these degradation fragments shows that the smaller fragment contains the hexahistidine tag, suggesting that proteolysis is occurring at the thrombin site adjacent to the histidine tag within the designed linker sequence (Figure 2.3).

Redesign of the linker region to eliminate this thrombin site, thereby creating the construct 50-GFP α , significantly improved long term storage stability while maintaining a high level of protein expression indicating that the shortened linker region is still long enough to bridge the 22Å span at the end of the GFP barrel. As detailed in Zacharias *et al.*, introduction of an A206K mutation forces the protein beta-barrel into a monomeric state by interrupting dimer binding at a hydrophobic patch even at high protein concentrations at which dimers would usually form [8]. Since the native chemical ligation reaction occurs in completely denaturing solution and the ligation product must be subsequently *de novo* folded, such an A206K

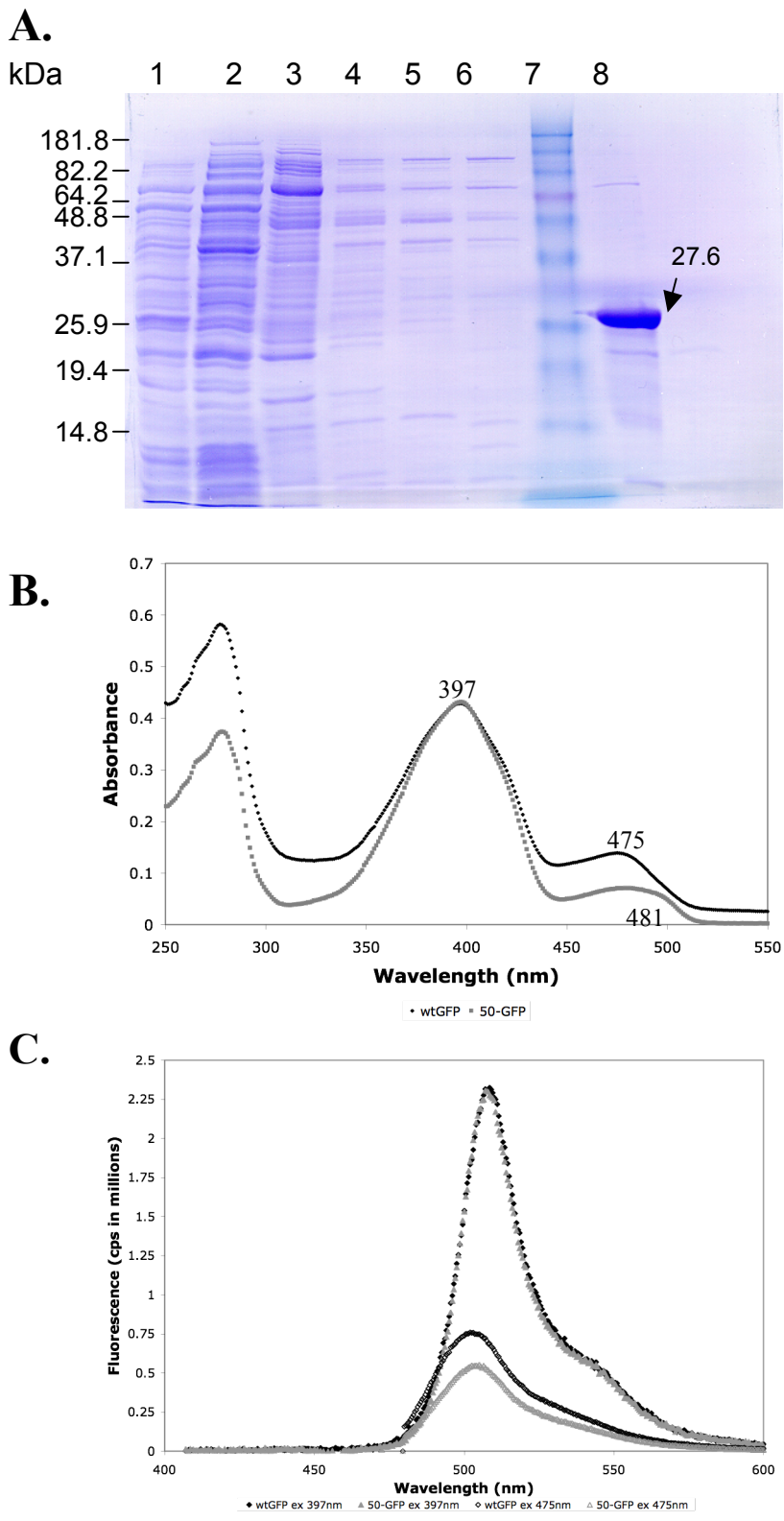


Figure 2.2.
Purification,
Absorption, and
Fluorescence of
50-GFP.

A. Purification profile of 50-GFP; (1) whole cell soluble, (2) Ni-NTA column flowthrough, (3) wash 0mM imidazole, (4) wash 10mM imidazole, (5) wash 15mM imidazole, (6) wash 20mM imidazole, (7) Benchmark Ladder, (8) 100mM imidazole elute.

B. Absorption spectra of wtGFP (black line) and 50-GFP (grey line), normalized to the 397nm peak.

C. Fluorescence emission spectra of wtGFP (black lines) and 50-GFP (grey lines) when excited into either the 397nm band (large peaks) for the neutral form of the chromophore or the 475nm band for the anionic form (smaller peaks) of the chromophore.

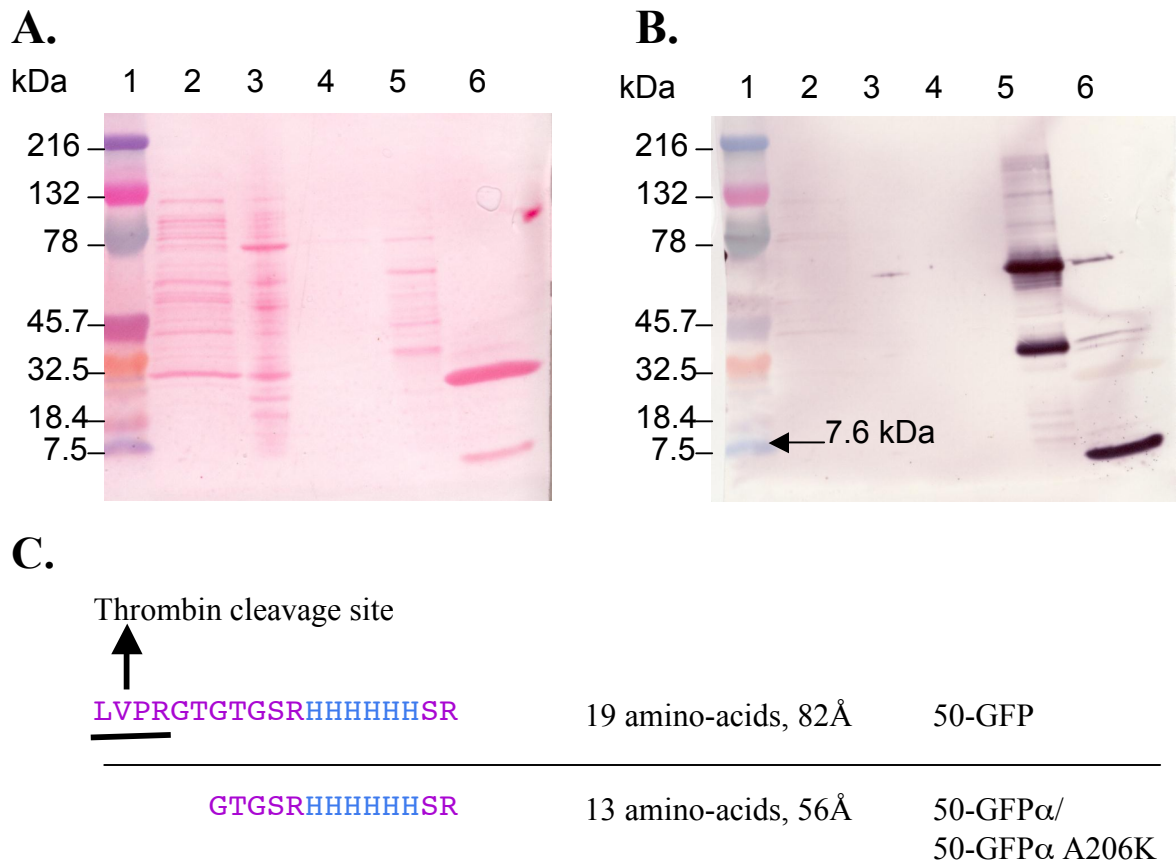


Figure 2.3. Western Blot analysis of 50-GFP degradation products and redesign of the linker region. **A.** Ponceau staining of nitrocellulose transfer of SDS-PAGE gel; a large fragment of GFP is seen just below the 32.5 kDa marker with a minor cleavage fragment seen at the 7.6 kDa marker (6). Other lanes in the gel are from an unrelated experiment **B.** Anti-6xHis antibody staining of this blot shows the 6xHis tag, part of the designed N/C termini linker, within the smaller cleaved fragment. **C.** sequence of the linker region containing the thrombin cleavage site encoded from the original pIWT5563his plasmid DNA and redesigned linker eliminating this thrombin site.

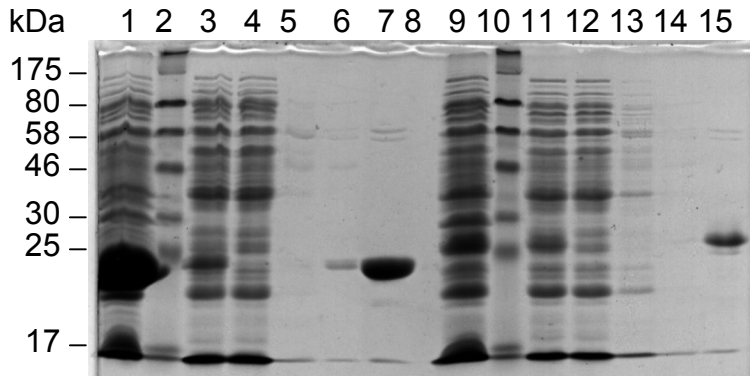
mutation may aid in preventing protein-protein interactions and aggregation while the protein folds and forms chromophore.

50-GFP α A206K exhibits absorption and fluorescent emission similar to wtGFP.

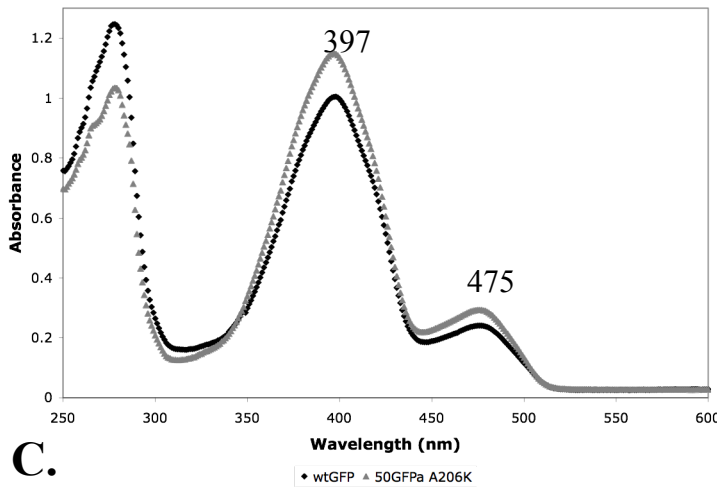
Redesign of linker region and the introduction of the A206K monomer-promoting mutation did not have any deleterious effect on both the cellular expression of the protein or on the solubility of the expressed construct, as evaluated by the method listed above. While this construct, termed 50-GFP α A206K, matches the absorption spectra of wtGFP closely, the fluorescence spectrum of this variant excited at 475 nm shows a slight decrease in emission at 503 nm relative to the same excitation of wtGFP (Figure 2.4). These differences in absorption and fluorescence properties suggest there may be differences in the structure of 50-GFP α A206K and wtGFP that may interfere with the normal kinetics of chromophore formation, potentially limiting our ability to use subtle changes within the chromophore structure to gain insight into wtGFP formation. To determine whether the slight difference in spectroscopy could hinder further investigation, we performed a detailed analysis of 50-GFP α A206K structure to compare the to previously published structure of wtGFP.

Intense irradiation of the chromophore with shortwave light (254 nm or 280 nm light) causes Glu222 decarboxylation via a Kolbe type mechanism that is seen within the wtGFP in the presence of a full hydrogen-bonding network [9, 10]. Therefore, we demonstrated decarboxylation in the circular permutant matches the decarboxylation of wtGFP, indicative of a chromophore within 50-GFP α A206K that is in a similar hydrogen bonding network to wtGFP (Figure 2.5).

A.



B.



C.

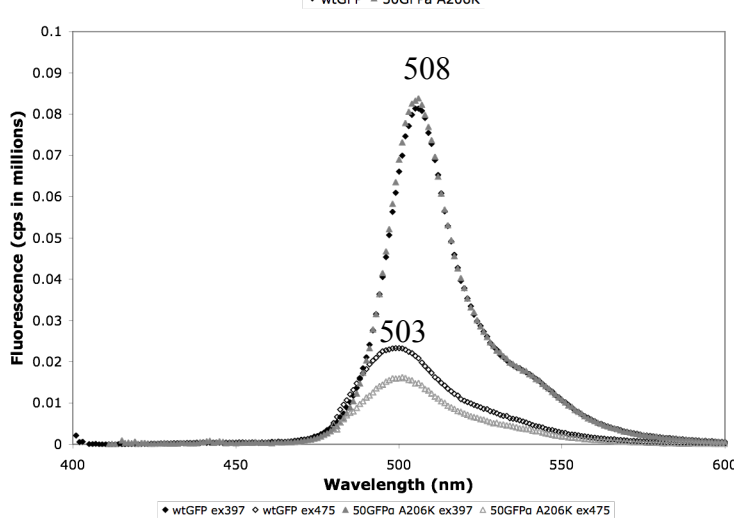


Figure 2.4.
Characterization of wtGFP and 50-GFPα A206K; purification, absorption, and fluorescence spectra.

A. SDS-PAGE gel of GFP purification, (1) 50-GFPα A206K whole cell lysate, (2) broad range ladder, (3) 50-GFPα A206K whole cell lysate, (4) Ni²⁺-NTA flowthrough, (5&6) low imidazole buffer washes, (7) 50-GFPα A206K elutant, (8) empty, 9 to (15) same as lanes 1 to 7, but for wtGFP. **B.** As with wtGFP (black line), the circular permutant 50-GFPα A206K (grey line) exhibits a peak at 397 nm and 475 nm, corresponding to the neutral and anionic forms of the GFP chromophore.

C. Fluorescence spectra of 70 nM of each protein excited at 397 nm and 475 nm corresponding to the absorption maxima in **B** (wtGFP in black, 50GFPα A206K in grey).

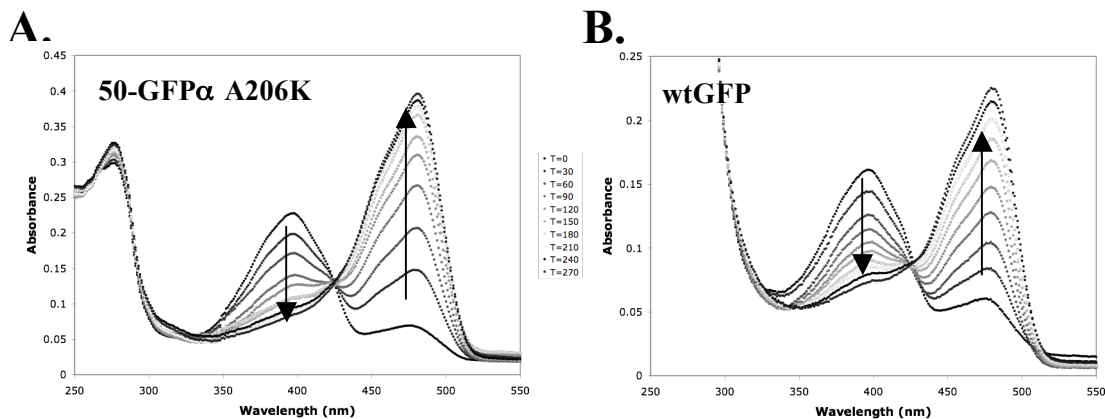


Figure 2.5. Irreversible photoconversion of 50-GFP α A206K and wtGFP due to UV irradiation induced decarboxylation at Glu222. Samples of wtGFP and 50-GFP α A206K were irradiated at 30 second intervals and the absorption spectrum scanned. A rise in the 475 nm absorption band indicates a conversion of the neutral form of the chromophore to an anion by disruption of the hydrogen bonding network in each sample caused by decarboxylation at Glu222 that permits the phenol of the chromophore to adopt an anionic species.

Circular permutant 50-GFP α A206K is structurally similar to wtGFP.

The overall structure of 50-GFP α A206K, derived by X-ray scattering crystallography, is highly similar to wtGFP with all eleven strands of the beta-barrel folding well around the central strained helix (Figure 2.6A). The 6xHis tag linker between the original N- and C- termini of the protein is visible within the structure and has distinct electronic density suggesting a well ordered and stable linker at the top of the protein barrel. In the cyclized version of GFP, expressed in the pIWT5563His construct published by Iwai *et al.*, the linker that joins the N- and C-termini is not visible in the electronic density map and was modeled into the structure

as methionines [11]. This result suggests a very flexible linker loop that does not form one predominant conformation within the crystal structure and may be subject to undesired thrombin proteolysis. This result confirms the need to redesign of the linker to a shorter sequence and eliminate the proteolytic site to improve overall protein stability. The newly generated N- and C- termini at residues 50 and 49, respectively, overlay well with the wtGFP structure (PDB ID: 1GFL), although the electronic density of residues 50 through 52 is missing, indicating that the break at this position causes the remaining loop at the end of the barrel to adopt multiple conformations (Figure 2.6A).

The chromophore of 50-GFP α A206K is well ordered within the protein barrel and the residues involved in the hydrogen bonding network, essential for fluorescence, are intact and in nearly identical conformations to wtGFP (Figure 2.6B). The only variation in this structure is the flipping of the imidazole ring of H148 closer to the chromophore phenol relative to this residue in wtGFP. The closer proximity of the imidazole ring to the chromophore may be responsible for the decreased emission of the anionic mode (i.e. 503 nm emission when irradiated at 475 nm) of the chromophore by promoting hydrogen bonding to the imidazole nitrogen upon excitation (Figure 2.6B).

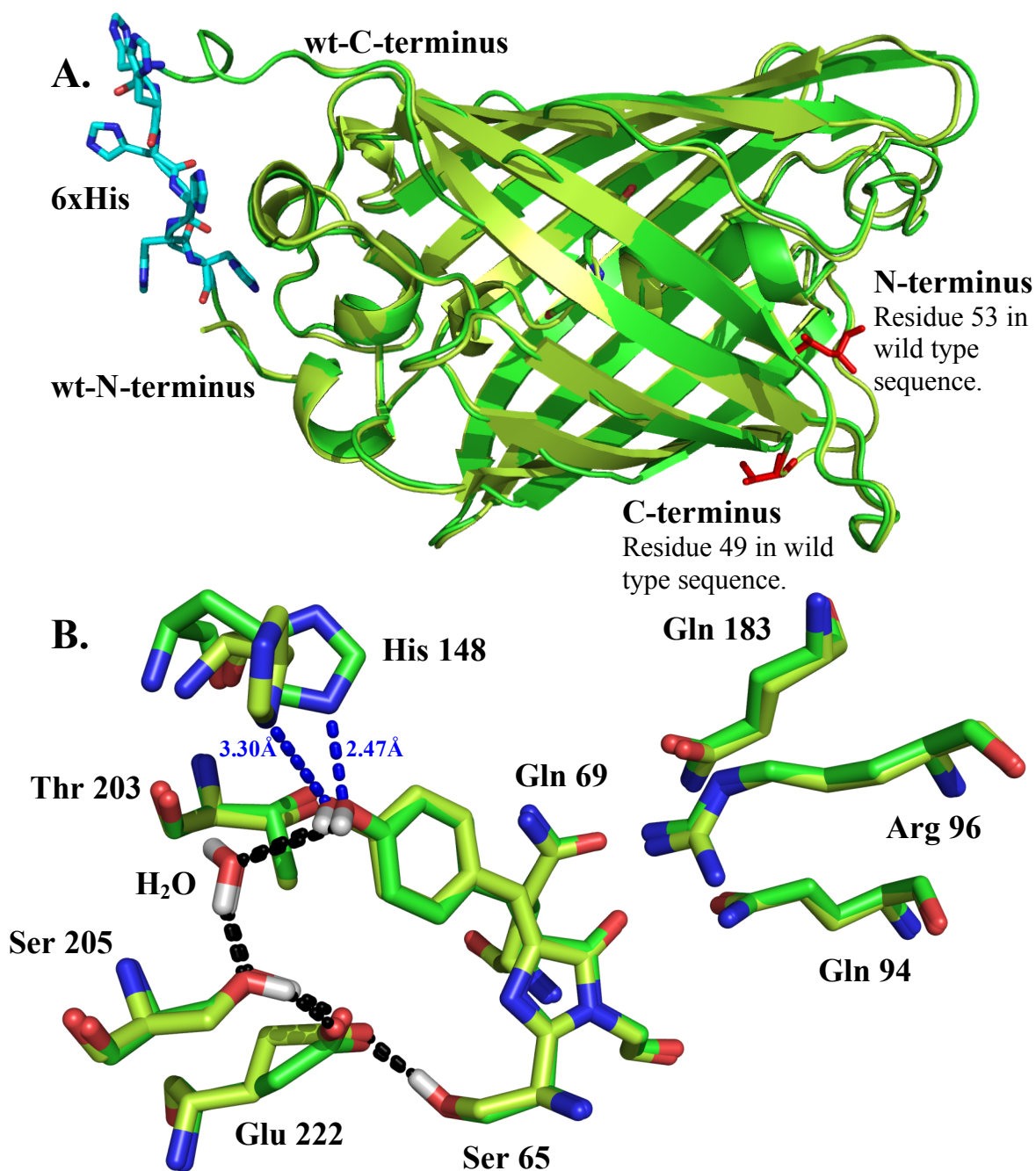


Figure 2.6. Structural Overlay of wtGFP (PDB ID: 1GFL) and 50-GFP α A206K.
A. wtGFP is in light green, 50-GFP α A206K is in green. The linker containing the 6xHis tag joining the N and C termini of the circular permutant is shown in light blue and the new N and C termini are shown in red. **B.** Overlay of residues surrounding the chromophore in wtGFP and 50-GFP α A206K. The hydrogen-bonding network in GFP is shown in black. Potential interaction and distances of the imidazole nitrogen of His148 in the two molecules to the chromophore phenol is shown in blue.

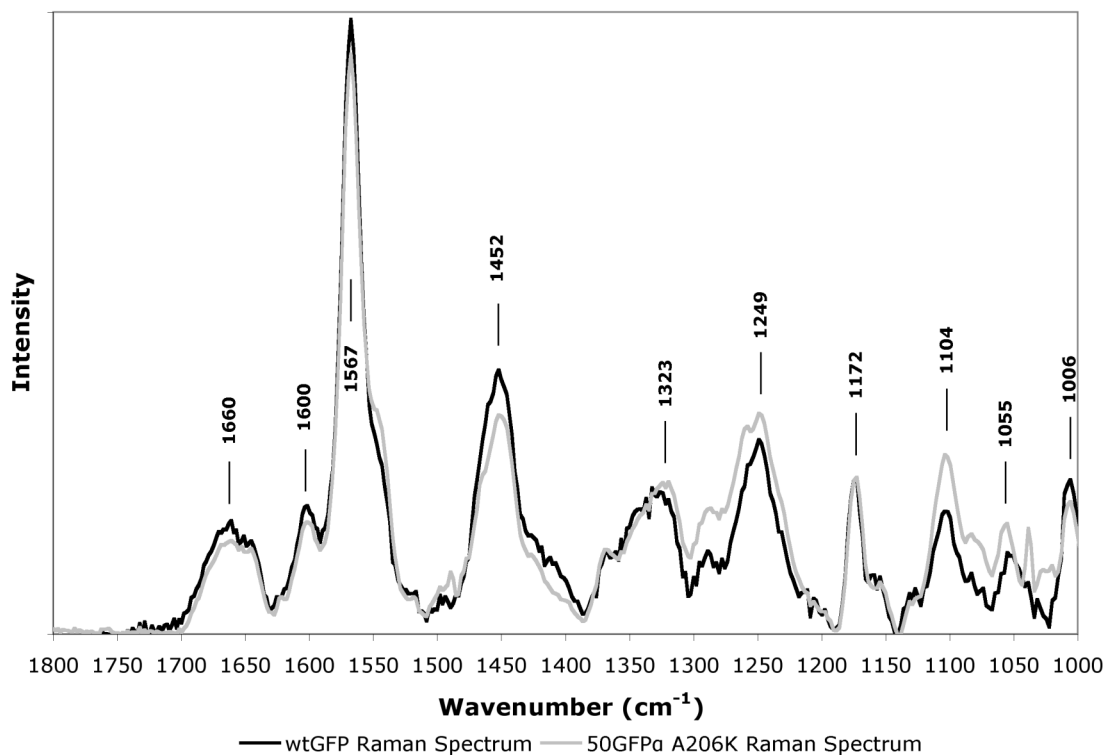


Figure 2.7. Raman spectrum of wtGFP and 50-GFP α A206K. Peaks corresponding to vibrational modes of the chromophore are indicated with their corresponding wavenumber, wtGFP in black and 50-GFP α A206K in grey.

A ground state vibrational analysis of the chromophore by Raman spectroscopy reveals a spectrum that closely resembles the spectrum of the wtGFP chromophore (Figure 2.7). Bands that are present in both wtGFP and 50-GFP α A206K include the phenol modes of the chromophore (1567 cm^{-1} , 1323 cm^{-1} , and 1249 cm^{-1}), carbon-hydrogen bond deformation within the phenol ring and in the remainder of the chromophore (1172 cm^{-1} and 1452 cm^{-1} , respectively), carbonyl stretching (1660 cm^{-1}), and modes from carbon-carbon stretching of the imidazolinone ring (1055 cm^{-1} and 1006 cm^{-1}) [5, 12]. With no extraneous bands or drastic peak

movements in the spectrum relative to wtGFP, the chromophore in 50-GFP α A206K appears to be in a protein environment that very closely matches wtGFP. This potentially indicates that proximity of the side chain of His148 to the phenol of the chromophore has only a minimal effect on the ground state of the chromophore. Were the phenol of the chromophore in 50-GFP α A206K affected by the proximity of His148, thereby modulating the ionization state, the major phenol band at 1567 cm⁻¹ would be altered, as seen in the ground state Raman spectrum of S65T GFP at pH 8.0 (Figure 7c in Bell *et al.*, 2000) [5]. Time-resolved infrared and Raman studies of 50-GFP α A206K and wtGFP may reveal further similarities or subtle differences in the excited state behavior of the chromophore.

Mass spectral analysis of the expressed 50-GFP α A206K variant is consistent with the expected mass for protein with a mature dehydrated and oxidized chromophore. With no mass spectral peaks corresponding to completely or partially immature chromophore, it appears that the expressed circular permutant efficiently forms chromophore post-translationally within the protein barrel (Figure 2.8).

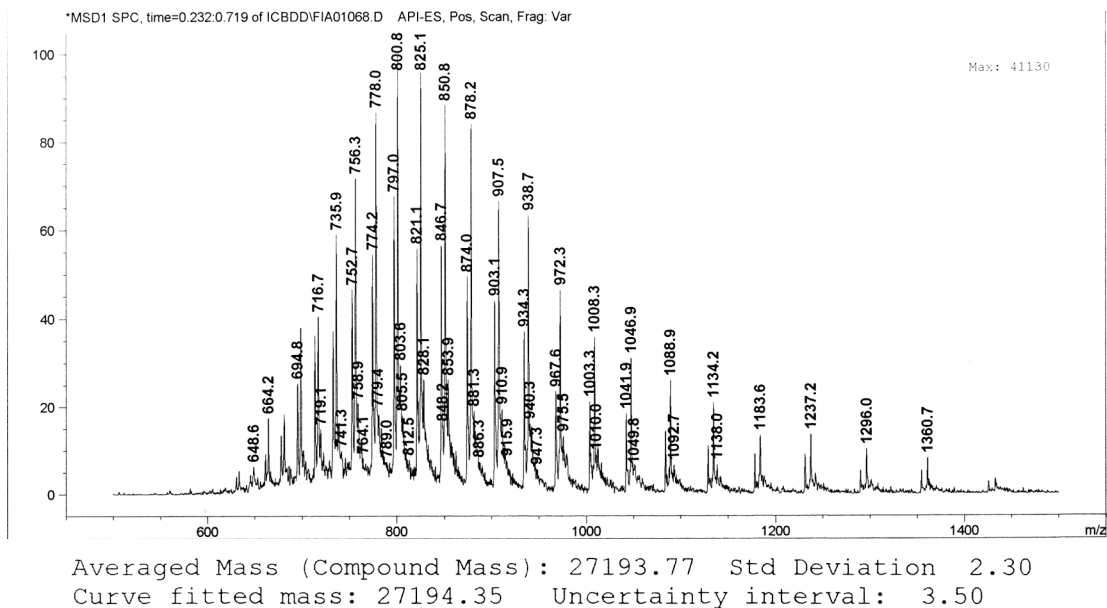


Figure 2.8. Flow injection ESI-MS of purified 50GFP α A206K in positive mode. Multiple ionization peaks were deconvoluted and compute to a mass of 27194.4 ± 3.5 , the expected mass of the mature chromophore protein is 27195.5.

The timescale of this maturation, monitored by *de novo* folding of urea solubilized 50-GFP α A206K inclusion bodies via fluorescence emission at 508 nm over time, is on the same order of magnitude as previously published data for the *de novo* folding of the S65T mutant of wtGFP (Figure 2.9B) and yielded protein whose absorption spectrum matched that of the natively expressed protein (Figure 2.9A) [3].

Initial experiments of *de novo* folding at 100x dilution with denatured protein concentrations at approximately 1.75 mg/ml showed a high initial fluorescence signal, suggesting scattering of incident excitation light from improperly folded protein aggregates in solution. Additionally, yields of matured 50-GFP α A206K from *de novo* folding of inclusion bodies using this method were approximately 1%.

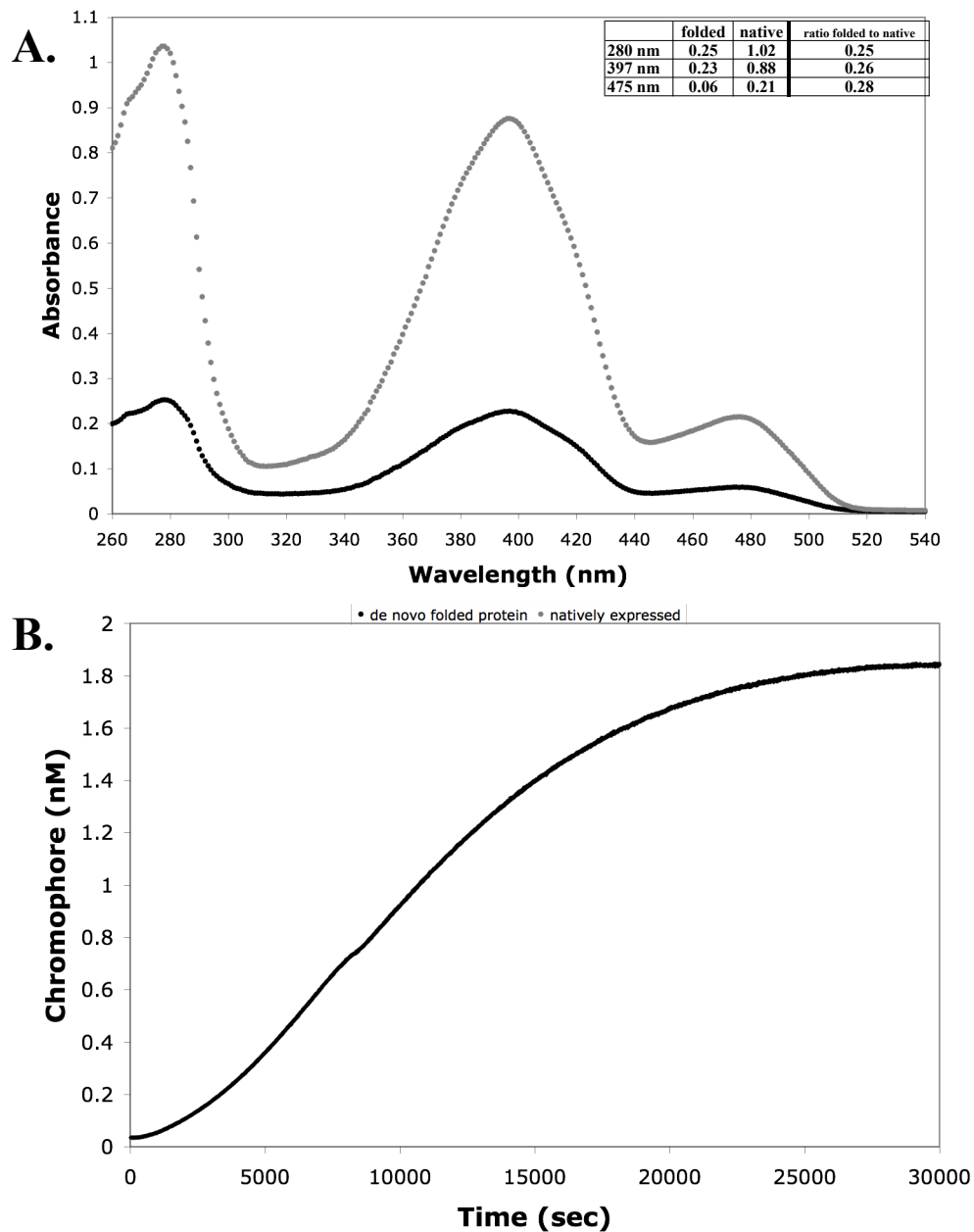


Figure 2.9. Comparison of *de novo* folded to natively expressed 50-GFP α A206K. **A.** Absorption spectra of 50-GFP α A206K of both the natively expressed and *de novo* folded protein, ratios the major absorption bands are shown in the inset table. **B.** Rate of *de novo* folding reveal that the rate of chromophore formation for 50-GFP α A206K from inclusion bodies is close to published rates of chromophore formation for S65T GFP.

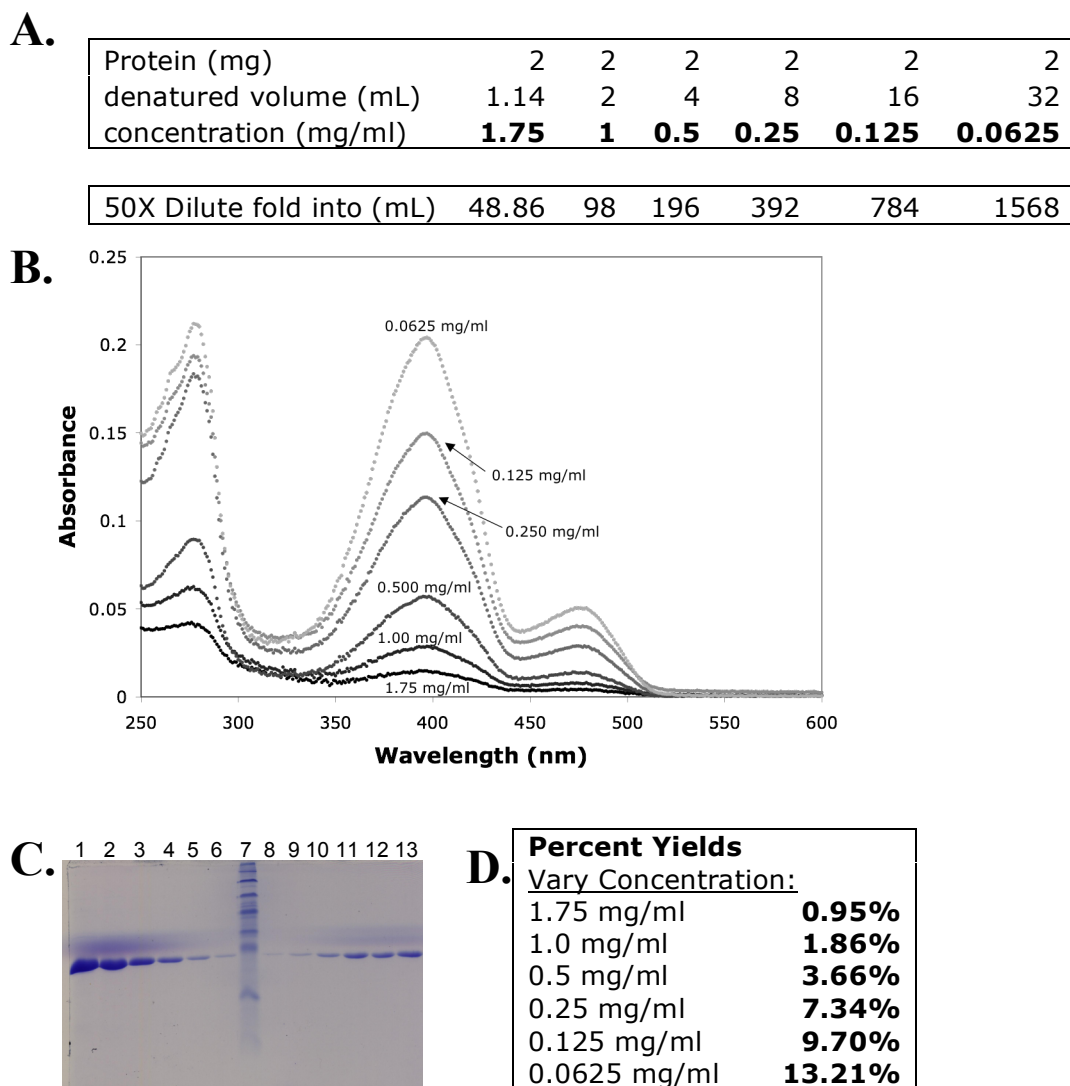
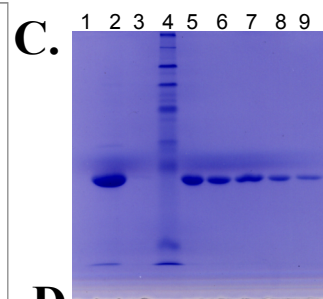
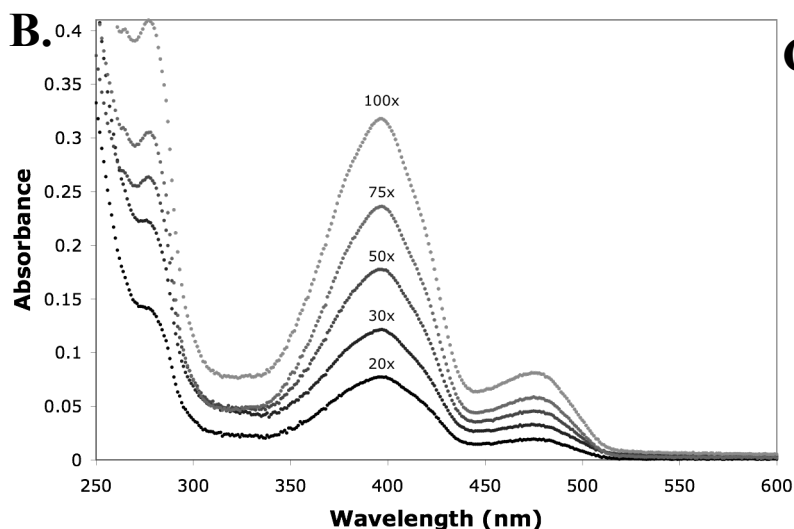


Figure 2.10. Optimizing *de novo* folding of the 50-GFP α A206K circular permutant by varying denatured protein concentration. Denatured 50-GFP α A206K inclusion bodies were diluted to the indicated protein concentration listed in bold in **A**. Protein was then folded by rapid dilution into 50 fold rapidly stirring folding buffer and concentrated/isolated via Ni²⁺-NTA purification. Resulting protein was concentrated to 1 mL exactly and quantified by UV/vis absorption **B**. Denatured protein samples at a each dilution and the resultant *de novo* folded protein is visualized in **C**. (1) denatured protein at 1.75 mg/mL, (2) 1.0 mg/mL, (3) 0.5 mg/mL, (4) 0.25 mg/mL, (5) 0.125 mg/mL, (6) 0.0625 mg/mL, (7) NEB Broad Range Ladder, folded samples at (8) 1.75 mg/mL, (9) 1.0 mg/mL, (10) 0.5 mg/mL, (11) 0.25 mg/mL, (12) 0.125 mg/mL, (13) 0.0625 mg/mL. **D**. Percent yield of mature folded protein recovered from *de novo* folding based on the quantity of chromophore from the fixed amount of denatured protein folded.

A.

protein (mg)	2	2	2	2	2
denatured volume (mL)	32	32	32	32	32
concentration (mg/mL)	0.0625	0.0625	0.0625	0.0625	0.0625

dilution factor	20	30	50	75	100
dilution volume (mL)	608	928	1568	2368	3168



D.

Percent Yields	
<u>Vary Dilution:</u>	
20X folding	5.02%
30X folding	7.89%
50X folding	11.53%
75X folding	15.36%
100X folding	20.65%

Figure 2.11. Optimization of *de novo* folding by modulating folding buffer dilution factor. Denatured 50-GFP α A206K inclusion bodies were diluted to 0.0625 mg/mL in denaturation buffer and folded by rapid dilution into folding buffer at the listed dilution factor in bold listed in **A**. Diluted protein was concentrated and isolated from folding buffer by flowing over Ni²⁺-NTA, then concentrating to exactly 1 mL and absorption spectra in **B**. were obtained. **C**. Protein samples were run on a 15% SDS-PAGE gel for total protein visualization, (1) empty, (2) concentrated inclusion body suspension (1.75 mg/mL), (3) inclusion bodies diluted to 0.0625 mg/mL, (4) NEB Broad Range protein ladder, (5) 100X dilution folding, (6) 75X dilution folding, (7) 50X dilution folding, (8) 30X dilution folding, (9) 20X dilution folding. **D**. Percentage yield of folded mature protein recovered from *de novo* folding based on the absorption of the neutral band on chromophore.

Optimizations of *de novo* folding at various denatured protein concentrations and several dilutions of folding factors were performed to decrease aggregate formation and improve folding yields. Yields of *de novo* protein folding improved significantly with a lower concentration of denatured protein in solution and with greater dilution factors (Figures 2.10 and 2.11), reaching a practical limit of 0.0625 mg/mL denatured protein in urea solution diluted 100 fold into rapidly stirring buffer. At these concentrations, the yield of *de novo* folding was improved to approximately 20%, sufficient for subsequent *de novo* folding of the ligation reaction products (Figure 2.11D).

Summary

To chemically access the chromophore forming residues within the middle of the amino-acid sequence of GFP using a synthetic method, we developed a circular permutant with an N-terminus only 15 amino-acids away from the chromophore forming residues, derived from a previously published fully circularized GFP construct [4]. This permutant would limit the size of the chemically expressed peptide to only 21 amino-acids, well within the range of SPPS techniques. The initial well-expressing permutant construct, 50-GFP, had a long flexible linker connecting the N- and C- termini that was subject to proteolysis upon storage. A redesign of this linker and the introduction of a monomer promoting A206K mutation, generating 50-GFP α A206K, produced a construct that expresses well and is stable.

Steady-state absorption and fluorescence spectroscopy on the 50-GFP α A206K circular permutant revealed a spectrum that matched wtGFP. The only notable deviation in these steady-state measurements is a slight decrease in the absorption and emission of the anionic form of the chromophore. This decrease may be due to stabilization of the anionic form of the chromophore by His148 that is 0.83 Å closer to the ionizable chromophore phenol. While there is a small deviation in structure, ground-state Raman spectra show that the chromophore exhibits no additional vibrational bands or significant changes in peak height relative to the

wtGFP chromophore. This result would suggest that the repositioning of His148 has a negligible effect on the photophysics of the chromophore in this permutant.

Lastly, this circular permutant was optimized for *de novo* folding, since the ligation reaction products of peptide to a truncated version of this protein will need to be *de novo* folded in completely denaturing conditions. With recovered yields of *de novo* folding approaching 20%, this circular permutant construct is a suitable for subsequent truncation and ligation work.

References

1. Baird, G.S., D.A. Zacharias, and R.Y. Tsien, *Circular permutation and receptor insertion within green fluorescent proteins*. Proc. Natl. Acad. Sci. U.S. A., 1999. **96**(20): p. 11241-6.
2. Topell, S., J. Hennecke, and R. Glockshuber, *Circularly permuted variants of the green fluorescent protein*. Febs Lett., 1999. **457**(2): p. 283-9.
3. Reid, B.G. and G.C. Flynn, *Chromophore formation in green fluorescent protein*. Biochemistry, 1997. **36**(22): p. 6786-91.
4. Iwai, H., A. Lingel, and A. Plückthun, *Cyclic green fluorescent protein produced in vivo using an artificially split PI-Pful intein from Pyrococcus furiosus*. J. Biol. Chem., 2001. **276**(19): p. 16548-54.
5. Bell, A.F., et al., *Probing the ground state structure of the green fluorescent protein chromophore using Raman spectroscopy*. Biochemistry, 2000. **39**(15): p. 4423-31.
6. Rosenow, M.A., et al., *The crystal structure of the Y66L variant of green fluorescent protein supports a cyclization-oxidation-dehydration mechanism for chromophore maturation*. Biochemistry, 2004. **43**(15): p. 4464-72.
7. Demidov, V.V. and N.E. Broude, *Profluorescent protein fragments for fast bimolecular fluorescence complementation in vitro*. Nat. Protoc., 2006. **1**(2): p. 714-9.
8. Zacharias, D.A., et al., *Partitioning of lipid-modified monomeric GFPs into membrane microdomains of live cells*. Science, 2002. **296**(5569): p. 913-6.
9. Bell, A.F., et al., *Light-driven decarboxylation of wild-type green fluorescent protein*. J. Am. Chem. Soc., 2003. **125**(23): p. 6919-26.
10. van Thor, J.J., et al., *Phototransformation of green fluorescent protein with UV and visible light leads to decarboxylation of glutamate 222*. Nat. Struct. Biol., 2002. **9**(1): p. 37-41.

11. Hofmann, A., et al., *Structure of cyclized green fluorescent protein*. Acta Cryst., 2002. **58**(Pt 9): p. 1400-6.
12. He, X., A.F. Bell, and P.J. Tonge, *Isotopic Labeling and Normal-Mode Analysis of a Model Green Fluorescent Protein Chromophore*. J. Phys. Chem. B, 2002. **106**(23): p. 6056-6066.

Chapter 3

Development of bacterially expressed truncation protein with an N-terminal cysteine, synthesis of a thioester peptide using Fmoc chemistry, native chemical ligation, *de novo* folding of ligation products, and characterization of ligated 50-GFP α A206K.

Introduction.....	58
Methods.....	70
Results and Discussion.....	80
Summary.....	98
References.....	102

Introduction

The unique chemistry of selectively joining one unprotected peptide or protein with a C-terminal thioester to another unprotected protein or peptide with an N-terminal cysteine has its biological origins in the intein reaction. Inteins were first discovered and described in yeast where the *vma1* gene product produced two distinct protein fragments from one gene [1]. The chemistry of the splicing reaction was detailed by Paulus and revealed that the intein promotes a series of chemical reactions as follows: N to S acyl shift of the peptide chain N-terminal creating a thioester, esterification of the resulting thioester to form a branched peptide intermediate with the Cys residue immediately C-terminal to the intein, and finally an S to N acyl shift forming a stable peptide bond [2]. The extein, the peptides just N- and C- terminal to the intein which the intein joins chemically, is then ejected from the intein complex generating two protein fragments from one translation product (Figure 3.1A). Introducing a mutation that prevents the splicing reaction (i.e. esterification to the C-terminal segment Cys) from completing but permitting thioester formation, generates an active thioester ready for ligation to a synthetically produced peptide with an N-terminal Cys residue.

Completing such a reaction, usually with the intein immobilized to chitin resin via an N-terminal chitin binding domain tag, ligates the synthetically generated peptide to the C-terminus of the bacterially expressed protein, in a reaction termed

Expressed Protein Ligation (EPL) [3] (Figure 3.1B). Joining two peptide fragments (one a C-terminal thioester and the other with an N-terminal Cys) in the absence of the intein subunit was pioneered by the Kent lab [4]. This intein-less reaction, termed Native Chemical Ligation (NCL), uses an organic thiol to promote the transesterification reaction (Figure 3.1C).

The NCL reaction permits the ligation of synthetically generated peptides, with the desired modifications to the chromophore forming residues, to a truncated version of 50-GFP α A206K, the circular permutant mimic of wtGFP. This strategy will allow for the incorporation of unnatural amino acid residues to probe either the formation or modulate the spectroscopy of the chromophore. Since the chromophore forming residues in the 50-GFP α A206K protein are close to the N-terminus in the primary sequence, the two peptides that would be needed are a short synthetic peptide thioester that includes the chromophore forming residues, and a truncated version of 50-GFP α A206K with an N-terminal cysteine. The closest cysteine past the chromophore residues, Cys70, will be used as the ligation point for thioester peptides.

Generation of this truncation protein, termed 50T70GFP α A206K, was attempted using several methodologies. While the traditional means of generating such a truncation construct using the above-described intein mediated approach was attempted, the truncated protein did not expressed solubly in necessary yields. Direct truncation of the circular permutant up to residue Cys70 leaves a start codon generated N-terminal methionine, thereby being ineffective as a ligation reactant. Therefore, endoproteases were used to produce an N-terminal cysteine protein.

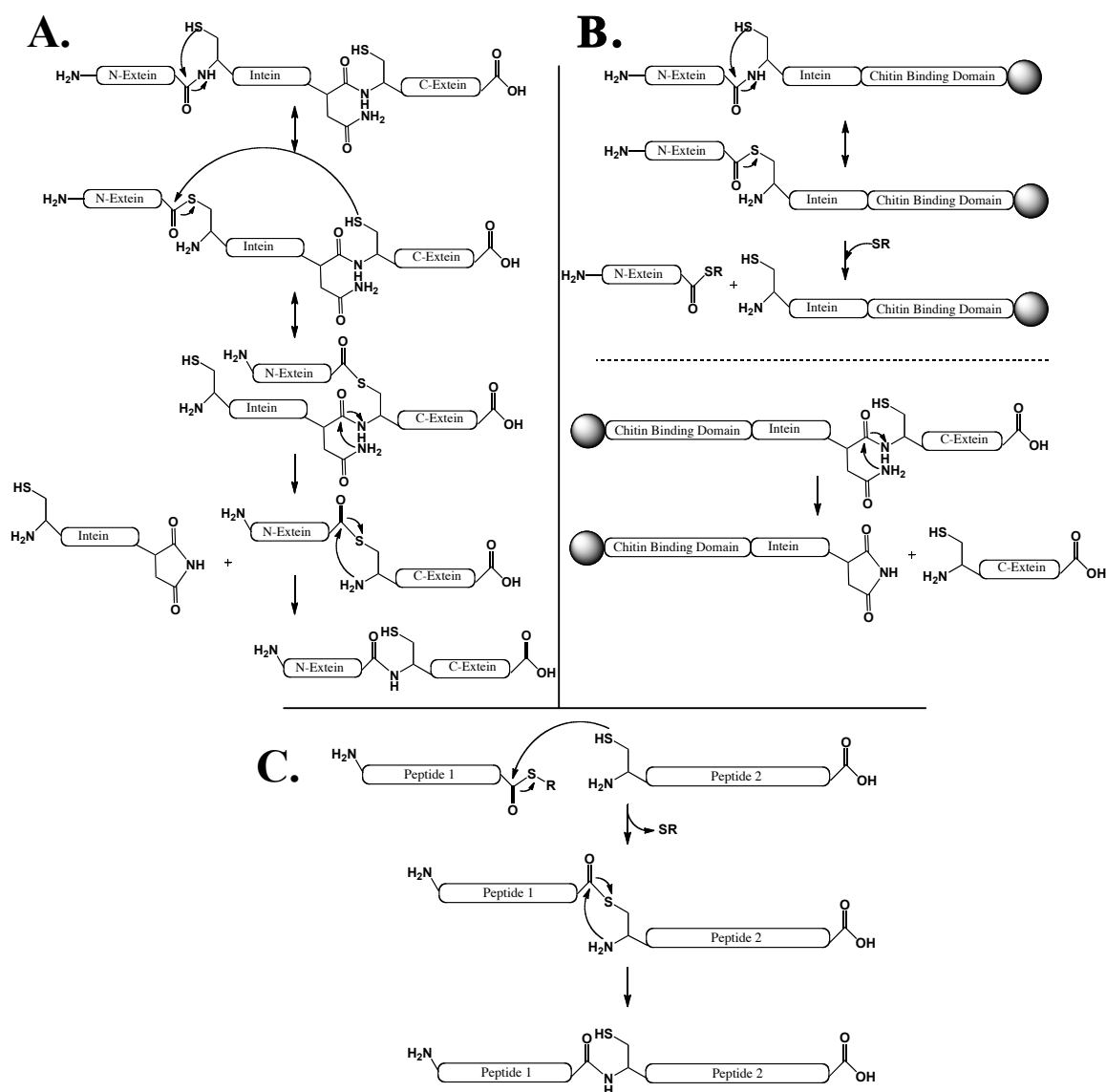


Figure 3.1. Mechanism of intein splicing and native chemical ligation. **A.** The N-terminal extein undergoes an N to S acyl shift and then esterifies to the cysteine side chain that begins the C-terminal extein. Formation of the asparagine residue of the intein cleaves the native peptide bond liberating the joined exteins. The exteins undergo a final S to N acyl shift generating a stable peptide bond. **B.** Splitting the intein and attaching it to a chitin binding domain allows isolation and purification of either an expressed thioester peptide or an N-terminal cysteine peptide. **C.** native chemical ligation requires only a C-terminal thioester peptide and an N-terminal cysteine peptide. Esterification joins the two peptide segments liberating a thiol side product. A final S to N acyl shift generates a stable peptide bond.

Since the initial truncated variants generated via the intein fusion method were not soluble, protein was expressed insolubly in inclusion bodies and solubilized in high molar urea solution. This denatured, but soluble protein preparation was reduced to low urea concentration by dialysis, and proteolysis attempted with either Factor Xa or with enterokinase. While Factor Xa proved to be intolerant of denaturing concentrations necessary to keep the tagged truncation construct soluble, enterokinase produced the desired truncated construct in large quantity. Protein was isolated from buffer by reverse phase high performance liquid chromatography (HPLC), yielding protein suitable for NLC and free of excess enterokinase and proteolyzed tag in high yield (approximately 20 mg per preparation). Our focus then turned to the synthesis of peptide thioester to ligate onto this truncated 50T70GFP α A206K protein.

Thioester Peptide Synthesis Strategy

The generation of peptide thioesters for use in native chemical ligation has proven challenging to date. While peptide thioesters may be synthesized using Boc-SPPS chemistry, these methods require the use of highly toxic HF as a final side-chain deprotection treatment that is not compatible with many laboratory environments [4-6]. For this reason, Fmoc based SPPS methods to generate a thioester peptide were explored. The problem with incorporation of a thioester tagged amino acid onto Fmoc-compatible solid phase resin is the ease with which thioesters may be converted to carboxylic acids under the basic conditions (20% piperidine) needed for Fmoc deprotection [7]. Recent work demonstrates several

methods for the generation of thioester peptides on Fmoc-SPPS compatible resin [8-10]. Two notable methods are the use of a backbone amide linker and a ‘safety-catch’ acylsulfonamide linker.

The backbone amide linker (BAL) method uses a resin bound trialkoxybenzyl handle to couple the backbone amine of an allyl esterified amino acid to the resin (Figure 3.2A). Synthesis of the peptide continues on the amine of the coupled allyl amino-acid. To generate the thioester, the allyl group is selectively removed leaving a free C-terminal carboxylic acid and a pre-formed thioester glycine is coupled to the available carboxylic acid [11]. TFA treatment of the peptide removes side-chain protecting groups and liberates the thioester peptide from resin. A derivative of this method uses the unprotected side chain of an allyl protected carboxylic acid glutamate residue as the attachment point to resin [12]. Again, deprotection of the allyl group leaves a C-terminal carboxylic acid available for coupling to a glycine thioester. As the side chain of the protected Glu residue is attached to an amide based resin, this amino-acid is converted to a glutamine residue.

The Kenner ‘safety-catch’ resin method allows coupling of the C-terminal amino-acid onto the amine of the acylsulfonamide linker and peptide synthesis proceeds using the PyBOP coupling reagent (Figure 3.2B). The sulfonyl group is activated to an alkylsulfonamide group by treatment with either trimethylsilyldiazomethane or iodoacetonitrile in *N,N*-diisopropylethylamine. The thioester is then liberated from resin by coupling to ethyl-3-mercaptopropionate using

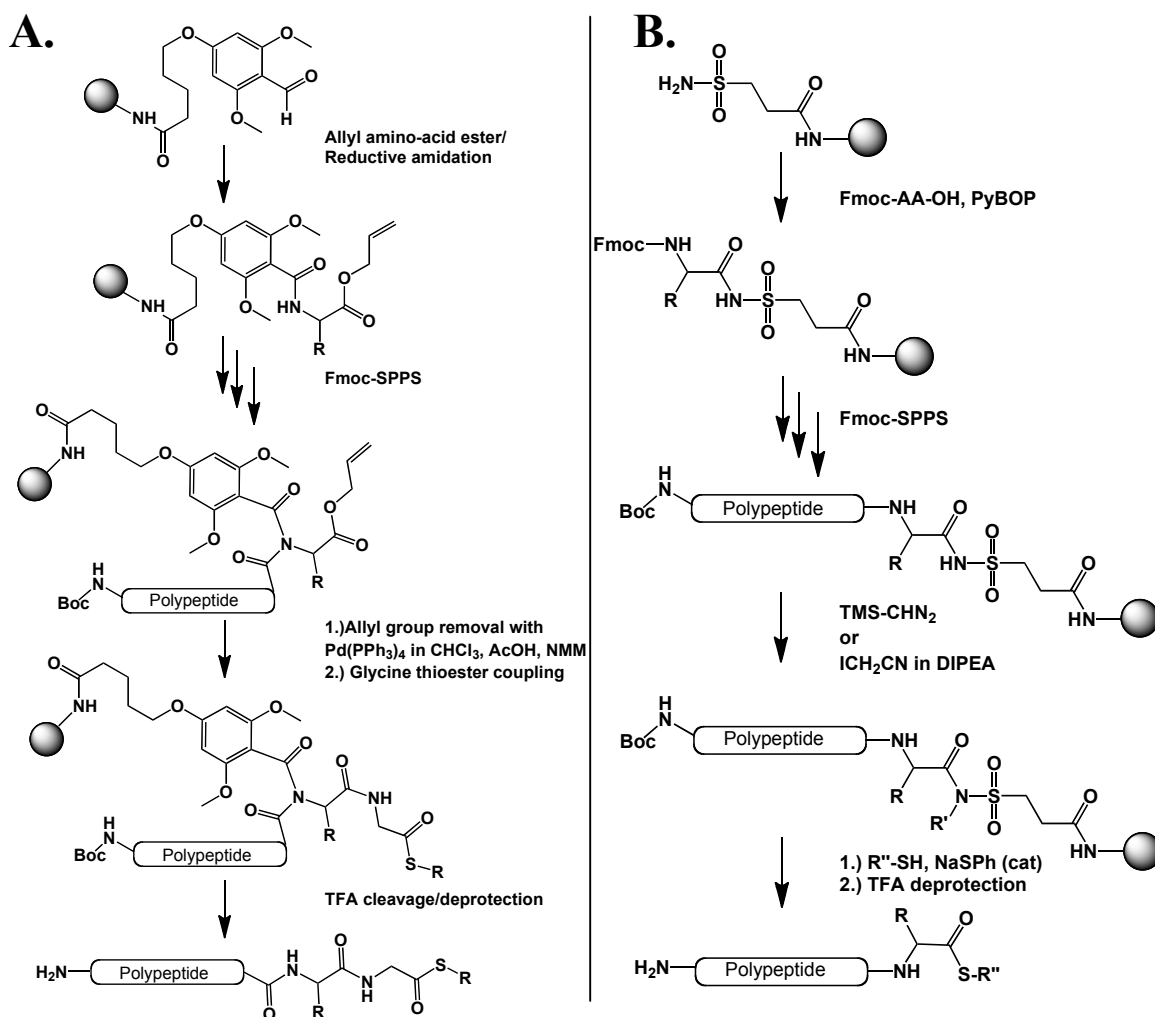


Figure 3.2. Two methods of thioester peptide generation using Fmoc-based solid phase synthesis.

A. The BAL method links an allyl protected amino-acid to the functionalized resin. Peptide synthesis is then completed and the allyl group is removed. Coupling a glycine thioester to the free carboxylic acid C-terminus generates the thioester peptide. TFA treatment liberates and side-chain deprotects the thioester peptide.

B. The Kenner ‘safety catch’ allows peptide synthesis to progress using standard coupling reagents. After activation of the sulfonamide group, thiophenolate-catalyzed thiolysis liberates the peptide from resin. Deprotection of side-chain protecting groups with TFA yields the thioester peptide.

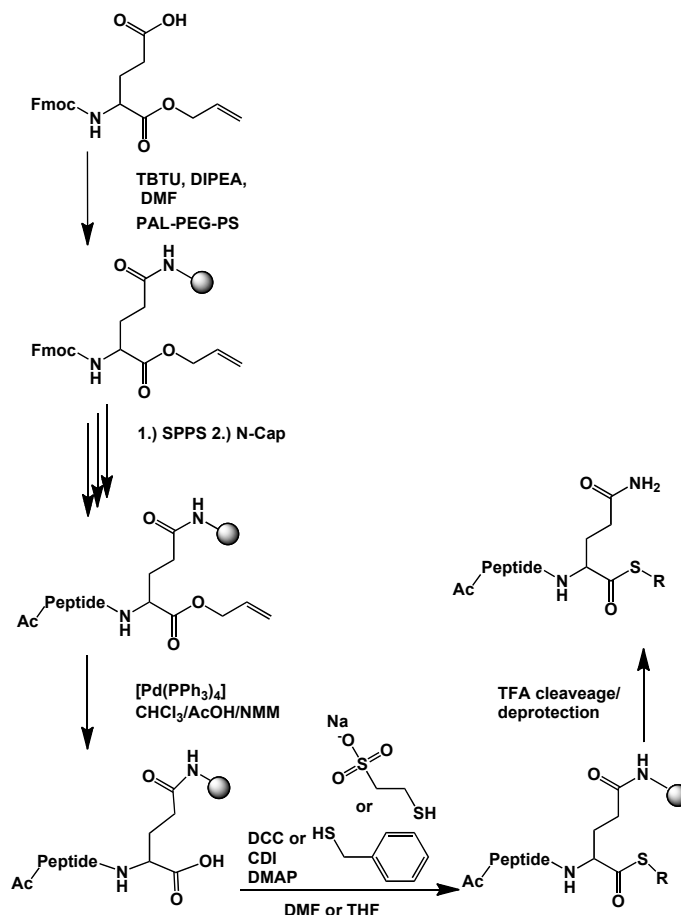


Figure 3.3. Attempted generation of a thioester peptide by using a side-chain anchored C-terminal protected amino acid scaffold. An allyl protected glutamate residue is loaded onto resin via the unprotected side chain. Solid-phase synthesis continues with standard coupling reagents on the backbone amine and the N-terminal amino-acid is protected with an acetic anhydride treatment. Palladium catalyzed hydrolysis of the allyl group using N-methylmorpholine in acetic acid/chloroform solvent generates a free carboxylic acid to which thiol coupling was attempted.

thiophenate as a catalyst and side-chain protecting groups liberated with TFA, thiolysis liberates the peptide from resin yielding the thioester peptide [13]. While this method has the benefit of not requiring a thioester glycine scaffold as a means of loading the thioester onto the C-terminus, it does suffer from premature liberation of

the peptide from resin during the alkylsulfonamide activation step and long reaction times when loading the C-terminal amino-acid [14, 15].

Our initial strategy to generate a peptide thioester was a modification of the method described in Dolphin, 2006 [12]. The needed sequence for ligation has a C-terminal glutamine, therefore, instead of coupling a glycine thioester to the C-terminal amino acid, direct esterification of the allyl deprotected glutamate residue was attempted (Figure 3.3). Using *N,N'*-dicyclohexylcarbodiimide (DCC) in 5% dimethylaminopyridine/DMF with differing thiols yielded only a peptide with a carboxylic acid C-terminus.

Suspecting limited chemical access of the C-terminal carboxylic acid group to coupling reagents while still attached to resin or premature hydrolysis of the thioester peptide during subsequent synthesis steps, a method to couple thiols to a resin free side-chain protected peptide was developed (Figure 3.4). Serendipitously, this method closely matched previously published methodology [16-21]. This method was therefore employed and generated the peptide thioester in yields sufficient for native chemical ligation.

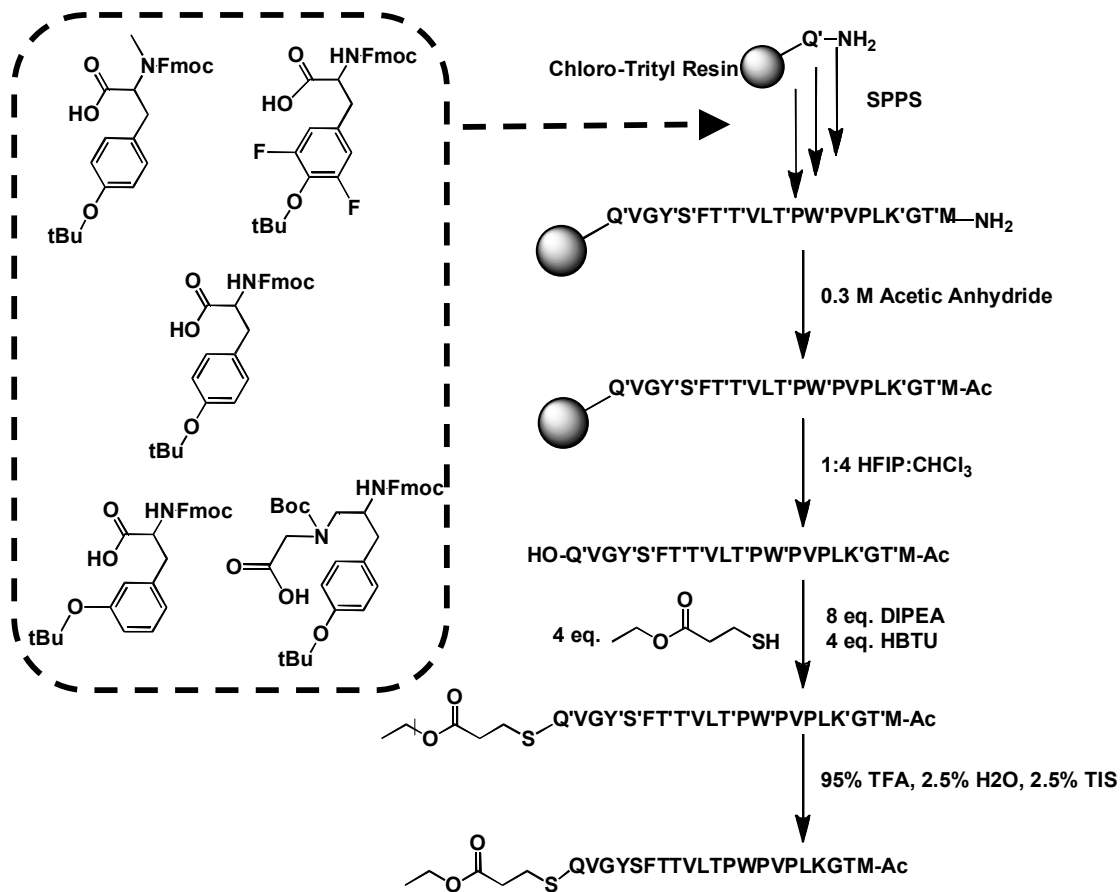


Figure 3.4. Synthesis of peptide thioester via a resin free protected peptide intermediate. Peptides were synthesized on chlorotriyl resin preloaded with trityl side-chain protected glutamine. Tyrosine(tbu)-OH or a designed unnatural amino-acid (outlined box) was inserted into the SPPS sequence at the appropriate position. The peptide was N-terminally protected by capping with acetic anhydride and liberated as a protected peptide with hexafluoroisopropanol in chloroform. Peptide was then esterified with an alkyl thiol and then side-chain deprotected with trifluoroacetic acid. Side chain protecting groups are listed as apostrophes as in; **X'**.

Native Chemical Ligation Strategy

Initial ligation conditions followed the extensively used protocol developed by Dawson [4, 8, 12, 13, 22]. N-terminal cysteine protein and C-terminal thioester peptides are suspended in denaturing solution at near neutral pH and the ligation reaction is initiated with the addition of 1-4% thiophenol and/or benzylmercaptan (Figure 3.5). The highly reactive added organic thiol displaces the resident thiol of the thioester and promotes transesterification to the N-terminal cysteine, thereby linking the two peptides together. A final intramolecular S-to-N acyl shift generates the native peptide bond, stably joining the two peptides.

Ligation of the enterokinase cut 50T70GFP α A206K with the 21-amino acid thioester peptide under these conditions caused the protein to precipitate out of solution and *de novo* folding of this reaction produced an insignificant amount of fluorescing protein. Since thiophenol is only sparingly soluble in aqueous solutions, excessive amounts of this catalyst are needed to drive a native chemical ligation reaction. Suspecting that the excess organic thiol (up to 4%) was causing protein precipitation in these initial ligation reactions, a water-soluble homologue of thiophenol, 4-mercaptophenyl acetic acid (MPAA), was used for further experiments [23]. This ligation catalyst was used in much lower concentration in solution and did not cause protein precipitation.

Ligation reactions using small peptides are usually combined at 1mM or more [8, 22, 24]. Since this reaction will ligate a 220 amino-acid GFP barrel with the 21

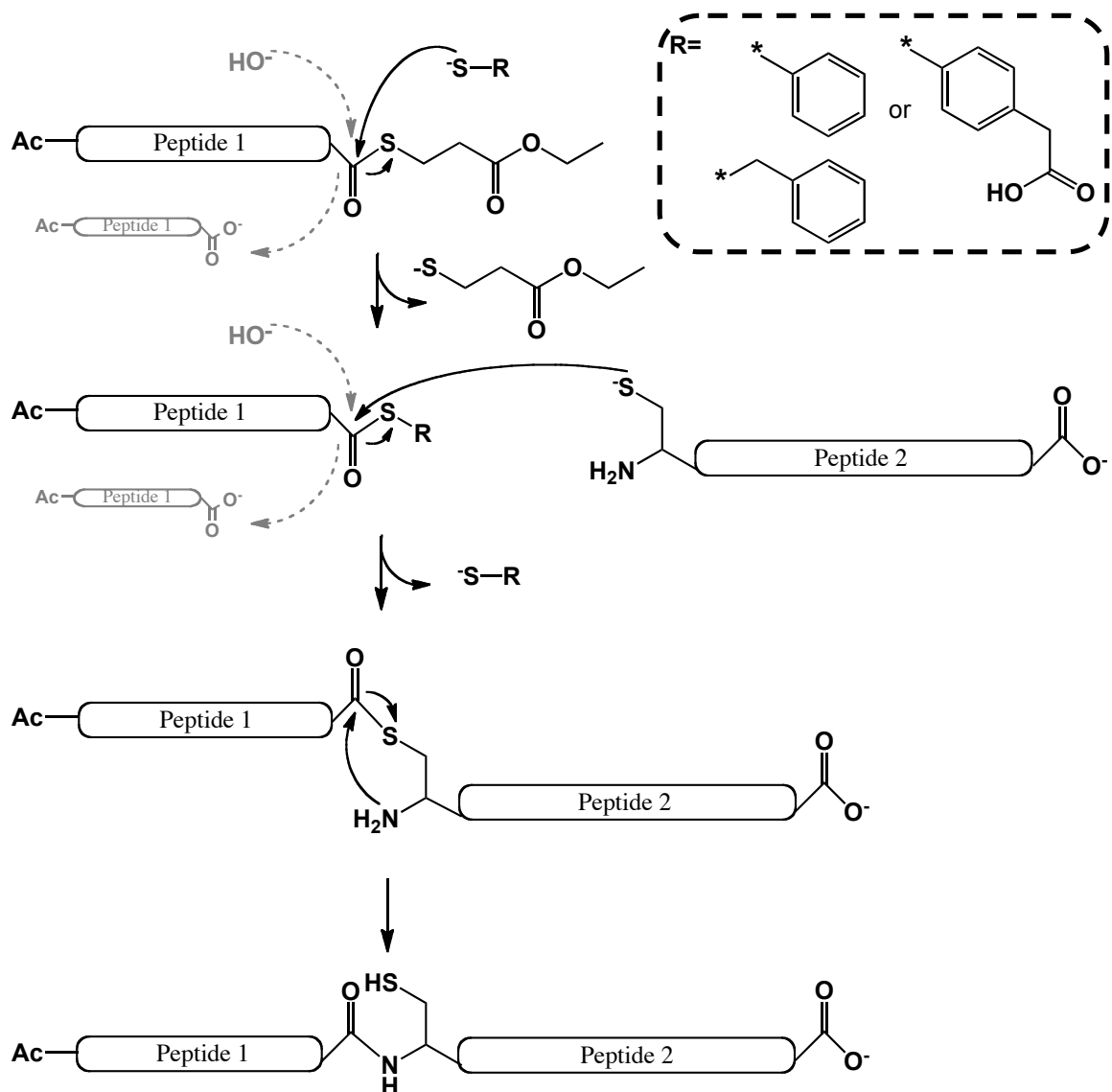


Figure 3.5. Native chemical ligation reaction promoted by *in situ* transesterification with reactive thiols. Alkyl thioesters on the peptide c-termini are displaced with a much more reactive benzylmercaptan or thiophenol. Mercaptophenyl acetic acid can be used as a water soluble alternative to thiophenol. These groups promote the transesterification of the thioester peptide onto the side chain of the N-terminal cysteine. An intramolecular S to N shift yields a peptide bond. Thiol and carboxylic acid groups shown ionized as they would be at pH 8. Potential base-catalyzed hydrolysis of thioester peptides shown in grey.

amino-acid synthetic barrel, a lower concentration of protein was used (0.25mM) to prevent protein tangling of 50T70GFP α A206K in denaturing solution. Additionally, as the N-terminal cysteine of 50T70GFP α A206K may not be as optimally accessible as a cysteine on a small peptide would be (i.e. may be partially buried in denatured protein conformations), peptide thioesters were used in 2 to 10 molar excess to promote the ligation reaction as the N-terminal cysteine becomes accessible. For this same reason, ligation reaction were carried out at pH 8 to ionize the N-terminal cysteine sulfur (pKa 8) and promote ligation [25]. While the N-terminal cysteine is partially deprotonated at pH 8, base-labile thioesters may hydrolyze to non-reactive carboxylic acids (Figure 3.5) [26]. Several molar excess of thioester would mitigate any errant hydrolysis.

After ligation reactions were optimized, products were diluted into high molar urea buffer and *de novo* folded into 100-fold buffer. Application of the folded protein solution to Ni²⁺-NTA affinity chromatography effectively concentrated and purified the protein. Ligated, *de novo* folded protein was characterized by UV visible absorption, fluorescence emission, and MALDI mass spectroscopy.

Methods

Cloning and Mutagenesis

Using the p50-GFP α A206K plasmid as template, the truncated DNA sequence encoding residues 70 through 228, the designed linker, and residues 3 through 49 on a single plasmid were amplified by PCR using the primers 5' GGTGGTGCTCTTCTAACTGCTTTTCCCGTTATCCGG 3' and 5' TTGGTTCTGCAGTTATTAAGTGCAAATAAATTTAAG 3'. The resulting amplified gene and pTWIN1 vector (New England Biolabs) were digested with PstI (New England Biolabs) and SapI (New England Biolabs). The gene was ligated into the cut pTWIN1 expression vector and transformed into *E. coli* XL1-Blue cells (Stratagene) on ampicillin selective media. Resulting colonies were grown in LB, 200 μ g/mL ampicillin, and plasmid DNA was isolated (Wizard Plus SV Miniprep system, Promega). The gene insertion, creating plasmid p50T70GFP α A206K-ptwin, was confirmed by Sanger sequencing using a T7 promoter primer (ABI systems).

The endoprotease sequence for factor Xa and enterokinase was inserted using two rounds of Quickchange mutagenesis (Stratagene). Modification of codons in the p50GFP α A206K plasmid ahead of Cys70 to the factor Xa sequence (Y66I/G67E/V68G/Q69R) used either primer set

5' GTCACTACTTTCTCT**ATTGA**AGTTCAATGCTTTTCC 3' and

5' GGAAAAGCATTGAACTTCAATAGAGAAAGTAGTGAC 3' or
5' ACTTTCTCTATTGAAGGTCGTTGCTTTTCCCGTTAT 3' and
5' ATAACGGGAAAAGCAACGACCTTCAATAGAGAAAGT 3'. Isolated
plasmids were sequenced with a T7 promoter primer, confirming the codon changes,
yielding the plasmid p50GFP α A206K-Factor Xa.

Similarly, the enterokinase recognition sequence was added into the
p50GFP α A206K-Factor Xa plasmid N-terminal to Cys70
(S65D/Y66D/G67D/V68D/Q69K) via Quickchange mtagenesis by using the two sets
of primers 5' TTGTCACTACTTTTCGATGATGAAGGTCGTTGCTTTTC 3' and
5' GAAAAGCAACGACCTTCATCATCGAAAGTAGTGACAA 3' or
5' ACTTTCGATGATGATGATAAATGCTTTTCCCGTTATC 3' and
5' GATAACGGGAAAAGCATTATCATCATCATCGAAAGT 3' yielding the
plasmid p50GFP α A206K-Entero. The plasmid was sequenced using a T7 promoter
primer.

Protein Expression and Purification

The constructed plasmid p50T70GFP α A206K-ptwin was transformed into
BL21(DE3)-pLysS (Stratagene) cells and grown in 500 mL LB media shaking at 250
RPM with 200 μ g/mL of ampicilin and 30 μ g/ml of chloramphenicol at 37°C for 2 ½
hours. The temperature was reduced 30 minutes prior to induction to either 25°C or
16°C and cultures induced with 0.3mM IPTG and then cultured overnight. Cells were

centrifuged at 5,000x g to pellet and stored at -20°C. Cell pellets were resuspended in 40 mL per L cell growth chitin binding buffer (10mM HEPES pH 8.5, 500mM NaCl, 1mM EDTA) and lysed via sonication at 18W in 45 second bursts in 1-minute intervals. Lysate was centrifuged at high speed (125,000 x g) for 90 minutes to pellet crude cell debris and filtered through a 0.22 µm filter. Supernatant was loaded onto 5 mL of chitin beads pre-exchanged into chitin binding buffer and washed with 10 column volumes of chitin binding buffer. Protein on column was washed again quickly with 3 volumes of chitin elution buffer (10mM HEPES pH 7.0, 500mM NaCl, 1mM EDTA), exchanged into chitin-Ni²⁺-NTA buffer (10mM HEPES pH 7.0, 500mM NaCl), and stored at room temperature overnight to permit intein self-cleavage of the truncated protein construct. Column flow through was saved and dialyzed against 3 L of 10mM HEPES pH 7.0, 500mM NaCl, 10mM imidazole prior to loading onto a Ni²⁺-NTA resin column. Column eluants were further analyzed and purified by binding to Ni²⁺-NTA resin, washing with chitin-Ni²⁺-NTA buffer supplemented with 10mM imidazole, and eluted with chitin-Ni²⁺-NTA buffer supplemented with 100mM imidazole. Eluting fractions were evaluated for truncated protein (termed 50T70GFPα A206K) by Coomassie stained SDS-PAGE gel.

Insoluble Expression of Protein and Endoproteolytic Cleavage

The plasmid p50-GFPα A206K-Factor Xa, was transformed into BL21(DE3)-pLysS cells and grown in 500 mL LB media shaking at 250 RPM complimented with 200 µg/mL of ampicilin and 30 µg/mL of chloramphenicol at 37°C until mid-log

phase. The temperature of the media was raised to 42°C and protein was induced with 0.8mM IPTG. Cells were grown for an additional 3 hours then pelleted by centrifugation at 5,000 x g. Cell pellets were resuspended in 40 mL per L cell growth 50mM HEPES pH 7.9, 300mM NaCl, 5mM β-mercaptoethanol, and 0.1mM phenylmethylsulfonyl fluoride. Cell suspension was lysed by passage through a French press four times and centrifuged at 125,000 x g for one hour. After removal of the soluble cell lysate fraction, the remaining insoluble inclusion body pellet was resuspended in 100mM Tris•HCl pH7.9, 500mM NaCl, 15mM EDTA, 5mM DTT, 2% Triton X-100 and centrifuged at 5,000 x g for 30 minutes. The pellet was washed by resuspending and centrifuging at 5,000 x g twice in 100mM Tris•HCl pH 7.9, 500mM NaCl, 15mM EDTA, and 5mM DTT. Washing was repeated twice with 50mM HEPES pH 7.9, 50mM NaCl, 1mM DTT, flash frozen in liquid N₂ as 1.5 mL aliquots, and stored at -80°C.

Aliquots of 1.5 mL of inclusion bodies were solubilized in 13.5 mL 8M Urea, 50mM Tris•HCl pH 8.0, 100mM NaCl, 1% Triton X-100, 5mM CaCl₂ and incubated for 1 hour at room temperature. Protein was then filtered through 0.22 μM filter and dialyzed against 1 L 50mM Tris•HCl pH 8.0, 100mM NaCl, 1% Triton X-100, 5mM CaCl₂, and again filtered through a 0.22 μM filter. Factor Xa (EMD Chemicals) was added to protein samples at one unit per 50 μg protein and cleavage proceeded overnight. The reaction was monitored by band shift in SDS-PAGE gel.

The 50-GFPα A206K plasmid with the enterokinase recognition tag encoded N-terminal to Cys70 (p50GFPα A206K-entero) was transformed into BL21(DE3)-

pLysS cells and grown in 500 mL cultures of LB media supplemented with 200ug/ml of ampicilin and 30 µg/mL of chloramphenicol at 37°C shaking 250RPM for 2 ½ hours. The temperature was raised to 42°C and protein expression was induced with 0.8mM IPTG for 3 hours. Cells were harvested by centrifugation at 5,000 x g and stored at -20°C. Cells were resuspended in 50mM HEPES pH 7.9, 300mM NaCl, 5mM β-mercaptoethanol and passed through a French press 4 times. Whole cell lysate was centrifuged for 90 minutes at 125,000 x g and the soluble lysate fraction removed. The remaining inclusion body pellet was resuspended in 100mM Tris•HCl pH7.9, 500mM NaCl, 15mM EDTA, 5mM DTT, 2% Triton X-100 and centrifuged again for 30 minutes at 5,000 x g. The inclusion bodies were washed (i.e. resuspended and centrifuged at 5,000 x g) twice with 100mM Tris•HCl pH7.9, 500mM NaCl, 15mM EDTA, 5mM DTT, then washed twice with 50mM HEPES pH 7.9, 50mM NaCl, 1mM DTT, and finally flash frozen in 1.5 mL aliquots in liquid N₂ and stored at -80°C.

Thawed aliquots were solubilized in 13.5 mL 20mM Tris•HCl pH 7.9, 50mM NaCl, 2mM CaCl₂, 20mM methylamine, 1mM DTT (enterokinase buffer) supplemented with 8M urea and incubated for one hour at room temperature. Protein was filtered through a 0.22 µM filter and dialyzed overnight to 1.5M urea in enterokinase buffer. The resulting protein was cleaved with the addition of 1:1000 U/ug recombinant enterokinase (Novagen) or 0.000075% w/w recombinant enterokinase (New England Biolabs) at 4°C for 24 hours, flash frozen, and lyophilized. The dried protein was suspended in 50/49.9/0.1 acetonitrile/water/TFA

and purified via HPLC on a Vydac protein and peptide C18 reverse phase column with a linear gradient from 95/4.9/0.1 water/acetonitrile/TFA to 95/4.9/0.1 acetonitrile/water/TFA over 47.5 minutes at 8 mL/min. Eluant that showed an intense 280 nm peak was collected, rotary evaporated, and lyophilized. This product was used subsequently for native chemical ligation.

Synthesis of peptide thioesters on a side-chain anchored, allyl protected amino acid

Using the general methodology outlined by Dolphin, 2006, the 21-amino acid peptide was generated by manual peptide synthesis on 0.5 g of PAL-PEG-PS resin (Applied Biosystems; 0.2 mmol/g substitution) in a 25 mL peptide synthesis reaction vessel [12]. Fmoc-amino acids and peptide coupling reagents, unless otherwise noted, were obtained from EMD Novabiochem and/or Advanced Chemtech. Resin was swollen in dimethylformamide (DMF; Fisher Scientific) for 45 minutes and drained. Fmoc-amino acids were coupled to the resin by pre-activating 4 equivalents of Fmoc-amino acid, 4 equivalents of 2-(1H-Benzotriazole-1-yl)-1,1,3,3-tetramethyluronium tetrafluoroborate (TBTU), and 8 equivalents of N,N'-diisopropylethylamine (DIPEA; Fisher Scientific) in DMF for 10 minutes and then reacting with resin for 60 minutes, shaking 300 RPM. After extensive washing with DMF, Fmoc protecting groups were removed with 20% piperidine (Sigma-Aldrich) in DMF and washed again liberally with DMF. Fmoc-Trp(Boc)-OH was doubly coupled after Kaiser ninhydrin testing revealed an incomplete coupling after a single

reaction [27]. Samples of resin were evaluated by MALDI to confirm completion of SPPS.

The N-terminus was capped by reacting the deprotected amino-acid with 0.3M acetic anhydride (Sigma-Aldrich) in DMF for 15 minutes. The allyl group on the C-terminal glutamine residue was removed by reacting the resin with tetrakis(triphenylphosphine)palladium(0) [Pd(PPh₃)₄] (3 equiv) in a mixture of dry chloroform, acetic acid, and N-methylmorpholine (17:2:1; 20 mL total) for 3 hours under N₂ atmosphere. The reaction was quenched and washed with 10 mL of 20mM diethyldithiocarbamic acid in DMF and 10 mL of 30mM DIPEA in DMF. The resin was then washed with DMF to remove any trace synthesis reagents, dried by N₂ stream, and stored under nitrogen at 4°C.

Samples of resin were esterified on a trial basis under the following conditions; 0.25 g resin (50 μmol), 20.52 mg (125 μmol) sodium 2-sulfanylethanesulfonate (MESNA), 0.305 mg (2.5 μmol) dimethylaminopyridine (DMAP), 11.34 mg (55 μmol) *N,N'*-dicyclohexylcarbodiimide (DCC) in dry DMF at 0°C and allowed to warm to room temperature for 3 hours. An alternate solvent attempted was dry THF, an alternate organic thiol attempted was benzylmercaptan at either 3 equivalents (150 μmol) or 6 equivalents (300 μmol), and an alternate coupling reagent attempted was carbonyldiimidazole (CDI). Samples of resin after reaction were treated with 92.5% TFA, 2.5% H₂O, 2.5% anisole, and 2.5% triisopropylsilane (TIS) to cleave and deprotect side-chain groups, precipitated with

cold ethyl ether, dried under N₂ stream, suspended in 50/49.9/0.1 acetonitrile/water/TFA and analyzed by MALDI mass spectrometry.

Synthesis of peptide thioesters via a protected peptide intermediate synthesized on chlorotrytil resin

The 21-amino acid peptides were generated by manual peptide synthesis on 0.75 g of 2-chlorotrytil resin pre-loaded with H-Gln(Trt) (Advanced Chemtech; 1mmol/g substitution) in a 25 mL peptide synthesis reaction vessel. Resin was swollen in dimethylformamide (DMF; Fisher Scientific) for 45 minutes and drained. Fmoc-amino acids were coupled to the resin by pre-activating 3 equivalents of Fmoc-amino acid, 3 equivalents of 2-(1H-benzotriazole-1-yl)-1,1,3,3-tetramethylammonium hexafluorophosphate (HBTU), and 6 equivalents of DIPEA in DMF for 10 minutes and then reacting with resin for 60 minutes, shaking 300 RPM. After extensive washing with DMF, Fmoc protecting groups were removed with 20% piperidine (Sigma-Aldrich) in DMF and washed again liberally with DMF. Fmoc-Trp(Boc)-OH was doubly coupled after ninhydrin testing revealed an incomplete coupling after a single reaction. The N-terminus was capped by reacting the deprotected amino-acid with 0.3M acetic anhydride (Sigma-Aldrich) in DMF. The resin was then successively washed with dichloromethane, DMF, isopropyl alcohol, methanol, and ethyl ether to remove any trace synthesis reagents and stored under nitrogen at 4°C.

The finished protected peptide was liberated from the resin by reacting with 1:4 hexafluoroisopropanol:chloroform for 1 hour. The suspension was filtered and

the resin was washed three times with chloroform. The flow-through was collected and evaporated to dryness. The C-terminal carboxylic acid was esterified following a modified protocol reported by the Imperiali group with 4 equivalents of 3-mercaptopropionic acid ethyl ester (MPA; TCI America), 4 equivalents of HBTU, and 8 equivalents of DIPEA in DMF for 16 hours [21]. The resulting product was precipitated with six volumes of diethyl ether and pelleted by centrifugation. The pellet was washed with diethyl ether, centrifuged again, and dried under a stream of nitrogen gas. Amino-acid side chain protecting groups were removed by resuspending the dried thioester product in 95% TFA, 2.5% triisopropyl silane, 2.5% water for one hour. Precipitation with diethyl ether and centrifugation afforded a white pellet. This pellet was again washed with diethyl ether and centrifuged before suspending in water with 0.1%TFA and lyophilizing. The crude thioester peptide was further purified by C18 RP-HPLC and mass was confirmed by ESI-MS.

Native chemical ligation and *de novo* folding of peptide thioester with 50T70GFP α A206K

Purified fragments of truncated 50T70GFP α A206K with an N-terminal cystein were combined with between 2 to 10 molar equivalent excess of purified 21-mer thioester peptides in 6M Guanidine•HCl, 0.1M sodium phosphate pH 8.0, 5mM 4-mercaptophenylacetic acid, 5mM tris-2-carboxyethyl phosphine (TCEP) to 0.25mM with respect to the truncated protein for 24 hours. Sampling 20 μ L of the reaction mixture at set intervals, isolating the protein fraction by TCA precipitation in

200 μ L 5% trichloroacetic acid, and evaluating resulting protein bands by SDS-PAGE, progress of the native chemical ligation was monitored. The ligation reaction mixture was diluted into 25mM MES pH 8.5, 8M Urea, 10mM EDTA, 0.1mM DTT so that the protein concentration was between 0.128 mg/mL and 0.0625 mg/mL and incubated at room temperature for one hour. The ligation reaction products were then folded by rapidly diluting dropwise into 100-fold 50mM Tris•HCl pH 8.5, 500mM NaCl, 1mM DTT. Folded protein was isolated from solution by passage over Ni²⁺-NTA resin, exchanging on resin into 20mM sodium phosphate pH 7.5, 300mM NaCl buffer, and then eluting with the same buffer supplemented with 250mM imidazole. The eluting protein was dialyzed against buffer without imidazole and concentrated to approximately 1.5 mL with centrifugal 3 kDa MWCO filters.

Fluorescence measurements were taken on a Quanta Master fluorimeter (Photon Technology International). Absorption measurements were taking on a BioCary 300 UV/visible spectrophotometer. Background absorption and fluorescence signals due to buffer were subtracted from spectra manually. Samples for MALDI mass spectrometry were prepared by C18 Zip-tip (Millipore).

Results and Discussion

N-terminal cysteine truncation protein generated by intein cleavage.

The generation of a protein construct suitable for native chemical ligation has traditionally been generated using an intein system that liberates a protein without the canonical start methionine. This system can therefore be used as a means of producing a protein with an N-terminal cysteine needed for native chemical ligation reactions. Using the pTwin1 system (New England Biolabs), a truncated version of the 50-GFP α A206K gene starting at Cys70 (50T70GFP α A206K) was introduced directly after the *Ssp* DnaB Intein in the pTwin1 vector. This construct was expressed under various conditions in an attempt to generate the desired truncation product. Expression overnight at 25°C was attempted initially to generate soluble truncation product. While the chitin binding domain-intein fusion did express and bind to chitin beads, it was bound as an already cleaved product (Figure 3.6B). Anticipating premature intein cleavage during cell culture and purification, the cell lysate passed through the chitin column was passed through a Ni²⁺-NTA column to rescue any already cleaved truncated protein product (Figure 3.6C). While there were several bands near the 25 kDa marker that could correspond to a truncation product, none proved to be of the correct molecular weight. Additionally, multiple protein bands

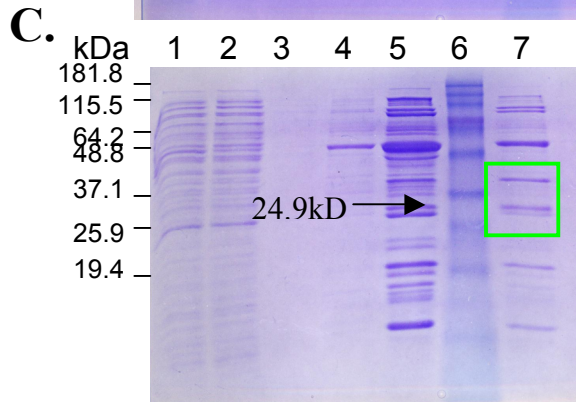
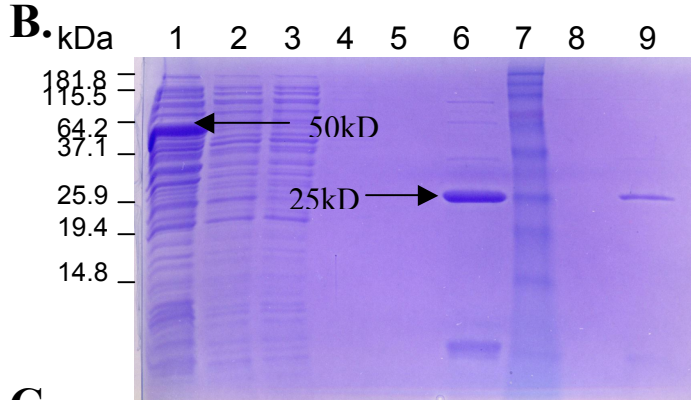
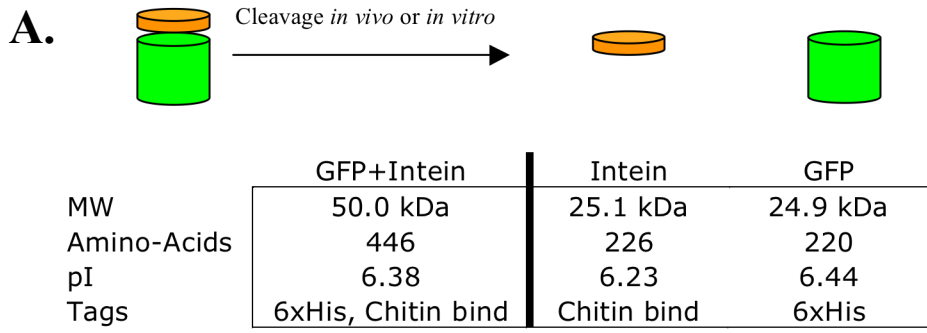


Figure 3.6.
Characterization and purification of 50T70GFP α A206K-*Ssp* DnaB intein fusion construct.

A. 50T70GFP α A206K can be liberated *in vitro* or *in vivo* from the *Ssp* DnaB intein fusion construct. Parameters show that the intein fragment is close to the same size and charge as 50T70GFP α A206K, leaving differing affinity tags as a means

of separating these two proteins **B.** chitin resin purification of the fusion product; (1) whole cell lysate cell lysate soluble, (2) chitin column flowthrough, (3) wash fraction pH 8.5, elution buffer wash fraction pH 7.0, (4-5) empty, (6) sample of protein bound chitin resin, (7) Invitrogen Benchmark ladder, (8) elution fraction after overnight incubation, (9) chitin resin sample after elution. The intein fusion construct and cleaved intein protein are marked by arrows. **C.** Ni²⁺-NTA purification of the chitin column flowthrough; (1) flowthrough from chitin column, Ni²⁺-NTA flowthrough, (2) wash 0mM Imidazole, (3-4) wash 20mM imidazole, (5) sample of protein bound to Ni²⁺-NTA, (6) Invitrogen Benchmark Ladder, (7) 150mM resin elution. Potential 50T70GFP α A206K bands are highlighted in the green box.

eluted from the Ni²⁺-NTA resin, indicating protein impurity and requiring additional purification steps to isolate the desired protein.

Since the full fusion protein is expressed within the cell, as seen by the band at 50 kDa in the whole cell lysate fraction, protein expression was attempted at 16°C to prevent premature protein splicing and encourage soluble protein production. While induction at lower temperatures did produce fusion protein that bound to chitin resin, no detectible soluble protein eluted from the chitin resin after inducing intein splicing overnight. Again, loading of the chitin column flow-through onto Ni-NTA resin to rescue already cleaved 50T70GFP α A206K, yielded a mixture of proteins bound to resin with only a major band at the incorrect molecular weight of 18.3 kDa (Figure 3.7).

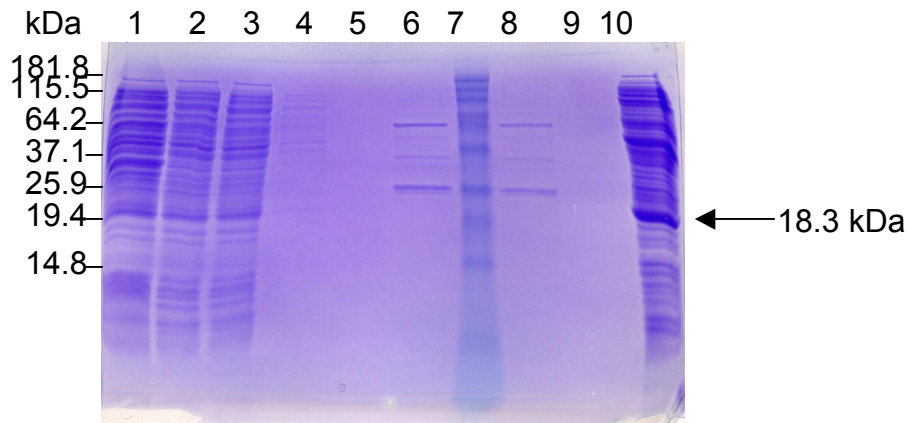


Figure 3.7. Chitin purification of the 50T70GFP α A206K-*Ssp* DnaB intein fusion construct induced at 16°C. Protein was expressed at 16°C and purified as described, (1) whole cell lysate, (2) soluble cell lysate, (3) chitin resin flowthrough, (4) wash buffer at pH 8.5, (5) wash buffer pH 7.0, (6) sample of chitin resin before inducing splicing, (7) Invitrogen Benchmark Ladder, (8) chitin resin sample after overnight incubation, (9) chitin resin elute, (10) sample of Ni²⁺-NTA resin with chitin resin flowthrough bound.

N-terminal cysteine truncation protein generated by endoproteolysis.

These results showed 50T70GFP α A206K construct is not soluble in solution, but expresses well as an insoluble protein that is shuttled to inclusion bodies. Isolation of this unfolded protein in inclusion bodies and solubilization in denaturing solution may provide the raw material needed for the native chemical ligation reaction. While inclusion body expression, isolation, and post-refolding splicing of the *Ssp* DnaB intein has been attempted before, the fusion protein 50T70GFP α A206K by itself has not been demonstrated as soluble product[28]. Any post folding spliced product would most likely precipitate out of solution. Insoluble expression of a direct truncation product leaves a start methionine residue at the N-terminus. Tagging the protein with an endoprotease sequence ahead of Cys70 and allowing endoproteolytic cutting of this tag generated the start cysteine needed for ligation. These proteolytic reactions were carried out in mildly denaturing conditions to maintain the 50T70GFP α A206K truncation product in soluble form.

Plasmid bearing the Factor Xa sequence just in front of the Cys70 in the 50-GFP α A206K construct (p50-GFP α A206K-Factor Xa) was expressed as inclusion bodies. These inclusion bodies were purified and solubilized in 8M urea then dialyzed to less than 100mM urea in buffer supplemented with 1% Triton X-100. Factor Xa was cut in 1% Triton-X 100 to maintain protein solubility. After the addition of Factor Xa and incubation at room temperature overnight, the protein precipitated out of solution into a gelatinous mass that was unsuitable for further

work. As Factor Xa is intolerant of urea concentrations above 250mM, this denaturant could not be used in place of 1% Triton-X100. Coomassie stained SDS-PAGE analysis of this reaction (both solution and precipitated protein) showed no shift in protein bands, indicating no proteolytic activity by the Factor Xa enzyme (Figure 3.8).

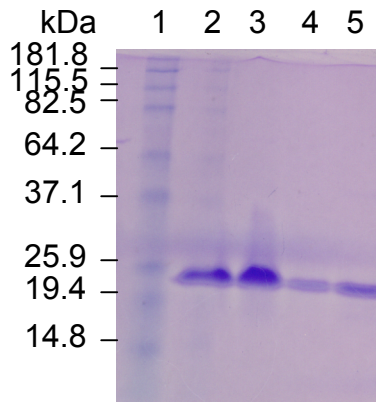


Figure 3.8. Factor Xa endoproteolysis of 50-GFP α A206K-factor Xa. Proteolysis of the inclusion body expressed and urea solubilized protein by factor Xa, from left to right; Invitrogen Benchmark Ladder, Urea solubilized inclusion bodies, post dialysis inclusion body sample, proteolysis overnight sample soluble fraction, proteolysis overnight sample insoluble fraction.

The enterokinase recognition sequence was encoded ahead of Cys70 in the p50-GFP α A206K construct. This plasmid was expressed in PLysS cells under conditions that produced inclusion bodies. Partially purified inclusion bodies were solubilized in urea and dialyzed overnight to 1.5M urea to maintain the truncation protein in solution while permitting enterokinase proteolysis (Figure 3.9). Pilot experiments with 1:2000 U/ μ g substrate protein showed that enterokinase cleaved

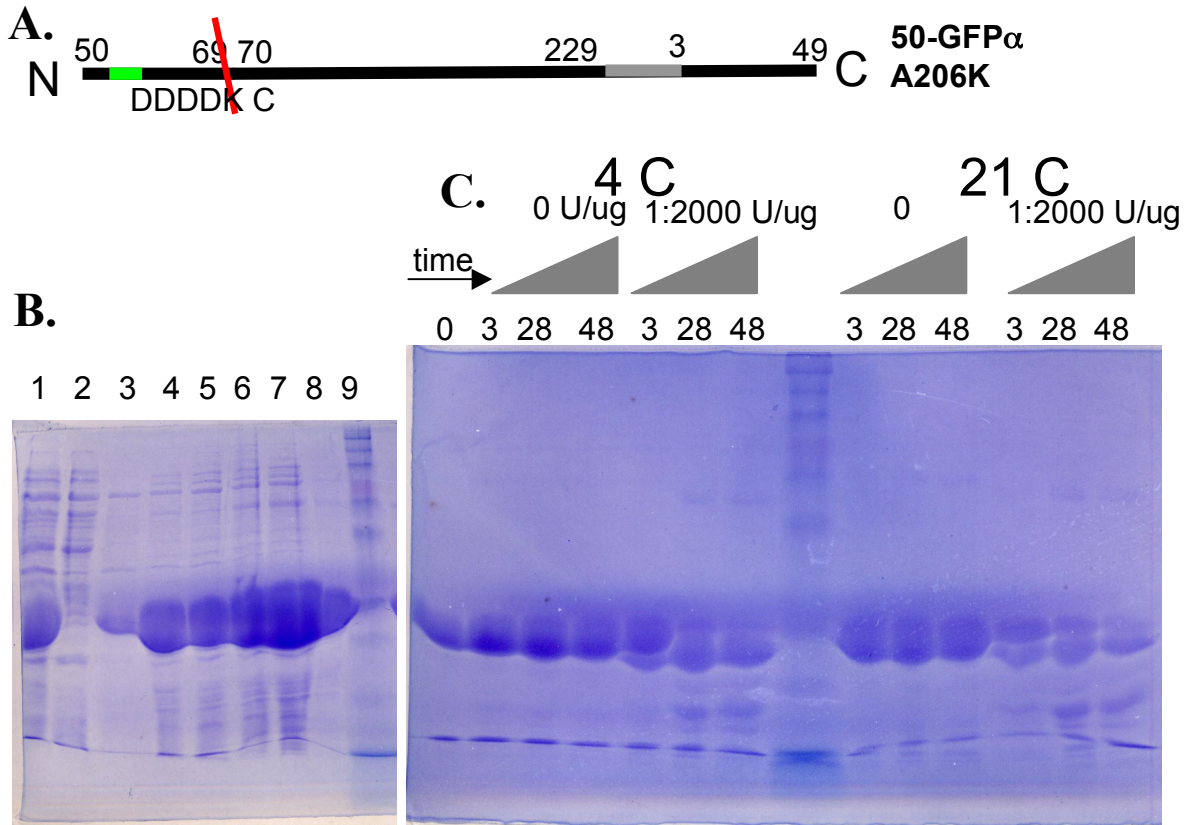


Figure 3.9. Enterokinase endoproteolysis of 50-GFP α A206K-entero. **A.** Schematic diagram of the sequence of 50-GFP α A206K-entero, residues that would form the chromophore are in green, linker joining the wild-type N and C termini in this circular permutant in grey, pertinent residue numbers listed above the diagram, and the enterokinase recognition sequence is denoted below the diagram. **B.** Inclusion body preparation of 50-GFP α A206K-entero, (1) whole cell lysis, (2) whole cell soluble fraction, inclusion bodies after 1st wash with Triton X-100, (3-5) 2nd wash, 3rd wash, 4th wash, (6) final inclusion body preparation, (7) urea solubilized inclusion bodies, (9) Invitrogen Benchmark Ladder. **C.** Trial enterokinase proteolysis assay, temperature of assay is denoted in large text, after a timepoint at the start of these reactions, three samples for each condition (control with no enterokinase, experimental with a 1:2000 U/ μ g ratio) were taken at 3, 28, and 48 hours.

most efficiently when the reaction was carried out at 4°C, therefore further process scale experiments were carried out at this temperature.

The enterokinase reaction was quenched by flash freezing the preparation in liquid nitrogen, and then lyophilizing the reaction. The suspension was dissolved in 50/49.9/0.1 acetonitrile/water/TFA and injected in 4 mL aliquots onto a preparative C18 protein & peptide column. The cleaved protein was isolated as a major peak along a linear gradient over 47.5 min between 24 to 26 minutes into the run (Figure 3.10). Protein yields were generally 20mg per 1.5mL of initial inclusion body preparation. Coomassie stained SDS-PAGE analysis and ESI-MS analysis (expected mass 24910.8, actual 24908) showed a cleanly proteolyzed protein that was lyophilized and stored at -20°C until needed.

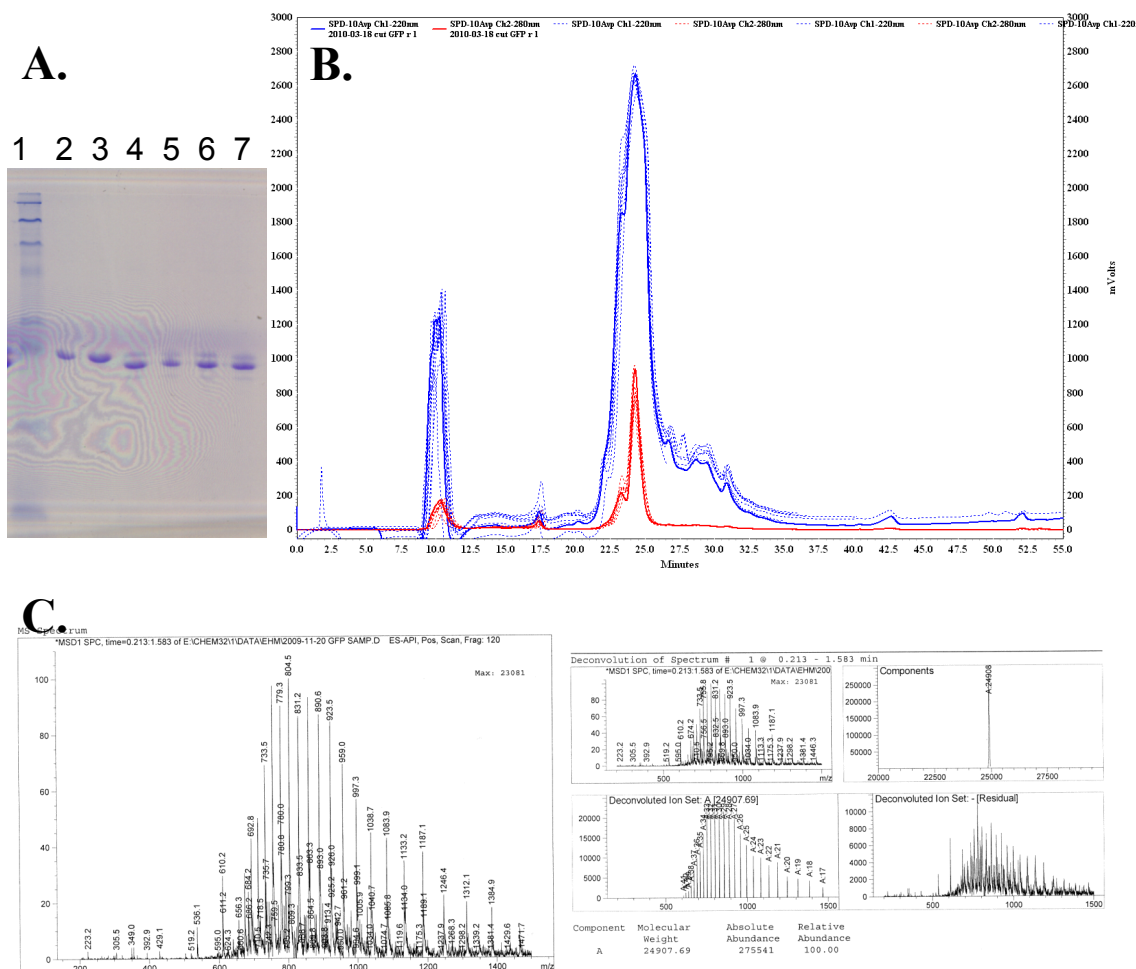


Figure 3.10. Large scale enterokinase endoproteolysis of 50-GFP α A206K-entero generating 50T70GFP α A206K. **A.** SDS-PAGE gel analysis of enterokinase cleaving of 50GFP α A206K-entero at 4°C over 24 hours, (1) Broad Range Protein Ladder (New England Biolabs), (2) control timepoint 0hrs, (3) control timepoint 24hrs, (4-7) reaction 1 through 4 experimental timepoint 24 hours. **B.** HPLC purification of enterokinase reaction, 220 nm absorption is in blue, 280 nm absorption in red, the first run is shown as a bold trace with subsequent runs shown as dashed traces. The major peak from 24.5 minutes to 25.5 minutes was isolated and lyophilized. **C.** ESI flow-injection analysis of the purified 50T70GFP α A206K, whole spectrum and deconvolution analysis; expected mass 24910.8, found mass 24908.

Synthesis of peptide thioesters via O-Allyl protected side chain anchored glutamine

In an effort to incorporate unnatural amino-acids into a peptide thioester that is suitable for ligation to the truncated circular permutant 50T70GFP α A206K, two methods for thioester generation using Fmoc-SPPS were explored. Initial attempts to generate a thioester on resin using a derivative of the backbone amide linker (BAL) methodology involved the coupling of the side chain of Fmoc-Glu-O-Allyl to PAL-PEG-PS resin. Manual synthesis extended on this first amino-acid was completed to generate the 21 amino-acid peptide. A sample of the assembled peptide (prior to acetic anhydride N-capping) was evaluated by liberating from resin with TFA. MALDI mass returned a peak corresponding to the expected peptide (Figure 3.11). After capping the N-terminal amine with an acetyl group, the allyl group on the C-terminus of the peptide was deprotected using palladium to yield the resin bound carboxylic acid peptide product. While esterification of this carboxylic acid was attempted using two thiols of differing reactivities (MESNA or benzylmercaptan) with two differing coupling reagents known to react thiols onto carboxylic acids (DCC or CDI) under solvent systems regularly used in peptide synthesis (THF or DMF), no reaction condition was found that efficiently converted the carboxylic acid to a thioester (Figure 3.12A & B). It may be possible to generate this thioester product with the use of other coupling reagents, solvent systems, or larger excesses of thiol.

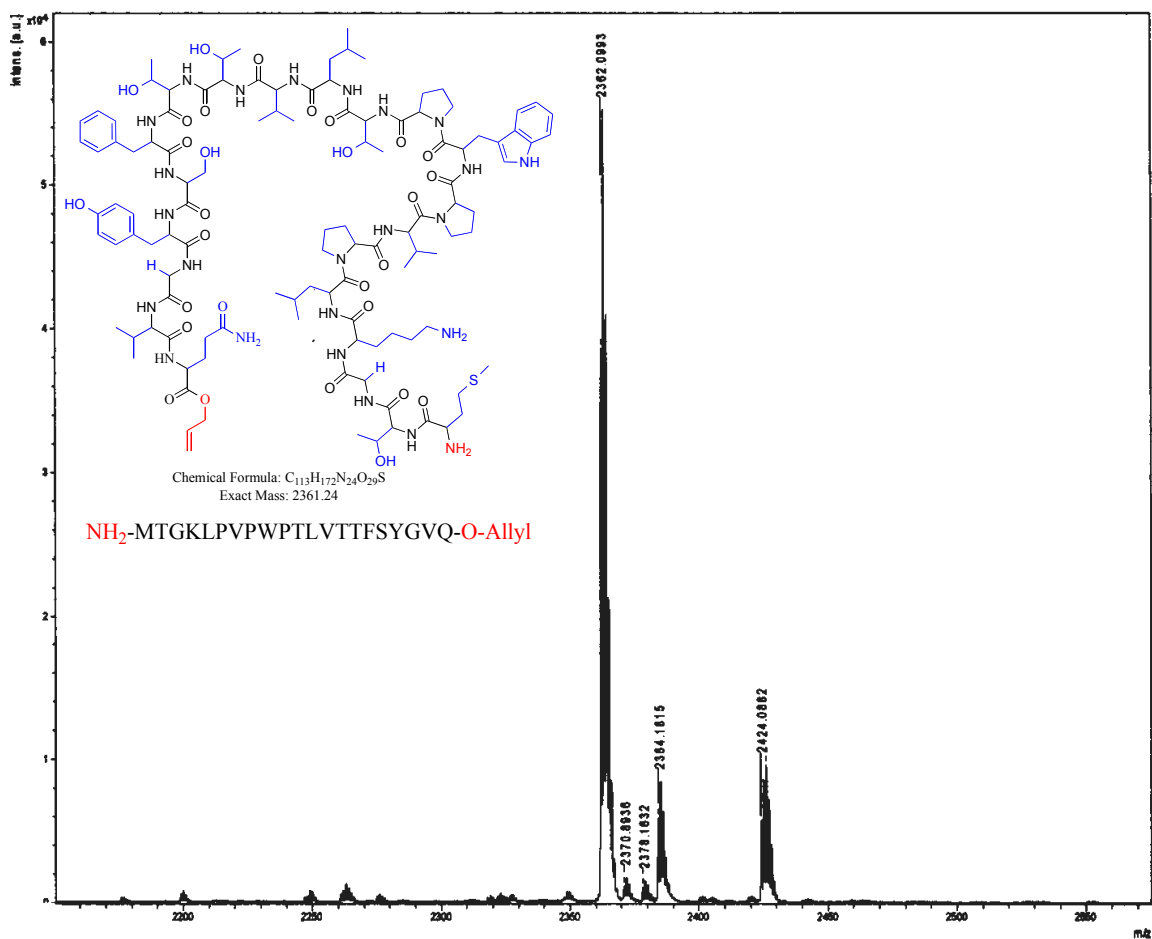


Figure 3.11. MALDI mass spectrometry of the 21-amino acid O-allyl protected peptide. A small sample of peptide was cleaved from resin & side-chain deprotected with TFA, precipitated with diethyl ether, and suspended in acetonitrile/water/0.1%TFA. The major peak in the spectrum, 2362.1 Da agrees with the expected mass of the allyl protected peptide of M+1; 2362.2 Da.

The proximity of the C-terminal carboxylic acid group to the resin matrix, with the majority of the peptide extending away from the resin, may be either sterically hindered from reacting with thiols as nucleophiles or a thioester at this position may form but be easily hydrolyzed back to a carboxylic acid. This may be why previous work with both the backbone amide linker method and this side

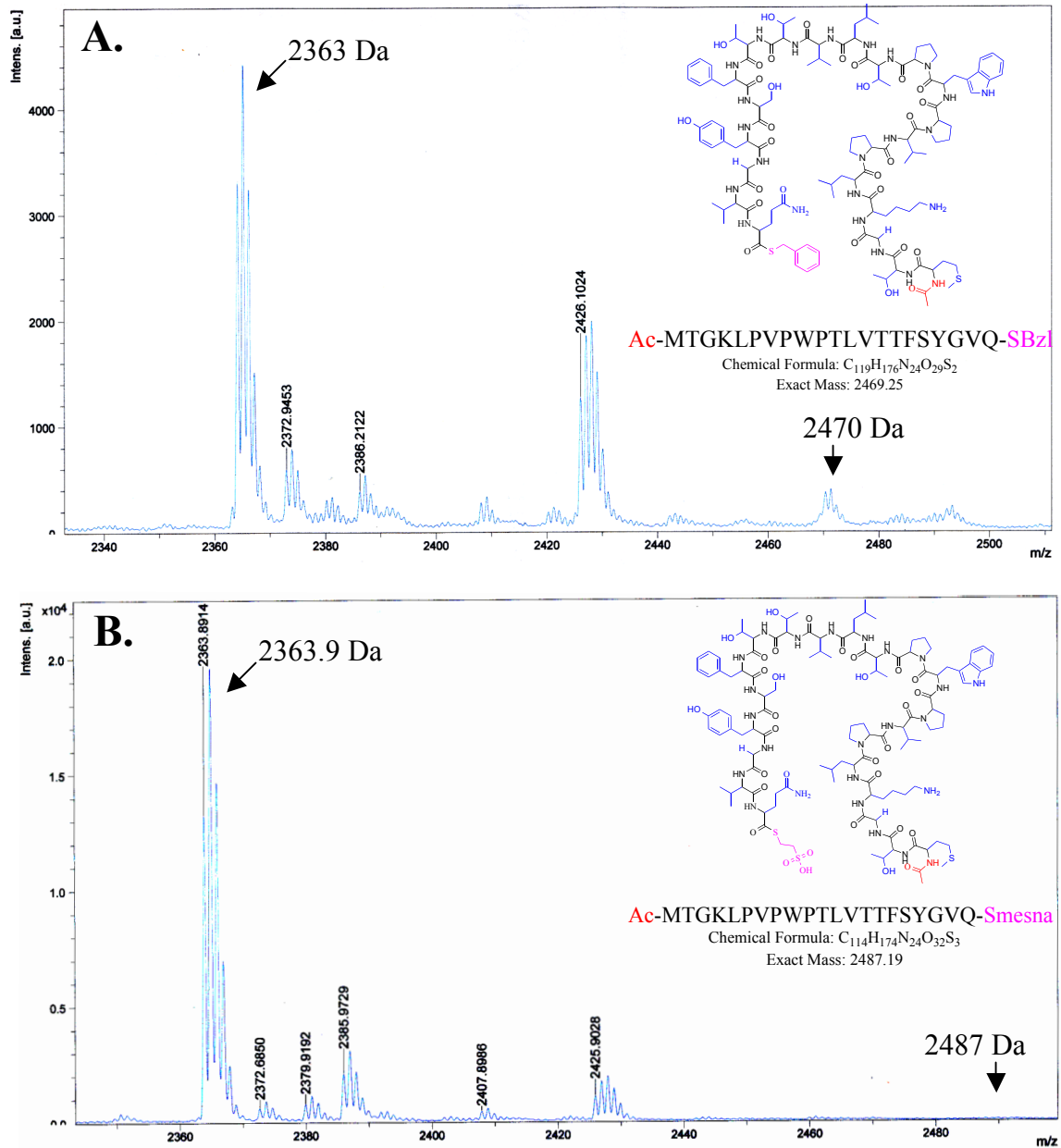


Figure 3.12. Unsuccessful esterification reactions on the carboxylic acid of the side chain anchored peptide. **A.** Esterification with benzyl mercaptan produced little usable product, only a minor peak at the expected mass of 2470 Da. **B.** Esterification with sodium 2-sulfanylethanesulfonate produced no thioester product, expected mass 2487 Da. Starting carboxylic acid peptide (calculated mass 2363.2 Da) is seen as the major product peak in these spectra.

chain anchored method use a preformed glycine thioester as a C-terminal coupling group as the peptide bond formed is much more stable under these conditions [11, 12]. Since the C-terminal amino acid in the needed thioester sequence is a glutamine residue, the addition of a glycine at the C-terminus of this sequence would most likely interfere with the proper positioning of this peptide within the GFP protein barrel.

Synthesis of peptide thioesters via a resin-free side chain protected peptide intermediate

Since the coupling of thiols to the C-terminus of a protected peptide while still anchored to synthesis resin was unsuccessful, liberation of the peptide from resin with side-chain protecting groups intact left the carboxylic acid at the C-terminus available in solution for thioesterification. Chlorotrityl resin permits the liberation of a side-chain protected peptide through the use of mildly acidic conditions (1:4 hexafluoroisopropanol:chloroform or 2:2:6 acetic acid:trifluoroethanol:dichloromethane) [29-34]. The 21-amino-acid peptide was synthesized manually on pre-loaded glutamine chlorotrityl resin. Efficiency of each coupling was monitored by absorbance at 300 nm of the deprotected Fmoc group (Figure 3.13A). The liberated protected peptide was reacted with an alkyl thiol (mercaptopropionic acid ethyl ester) whose structure mimics the thioester group left by Boc synthesis of thioesters on TAMPAL resin and has been used previously for thioester peptide generation (Figure 3.13C) [13, 35]. Side chain protecting groups were then removed with TFA treatment, the thioester peptide was purified via HPLC,

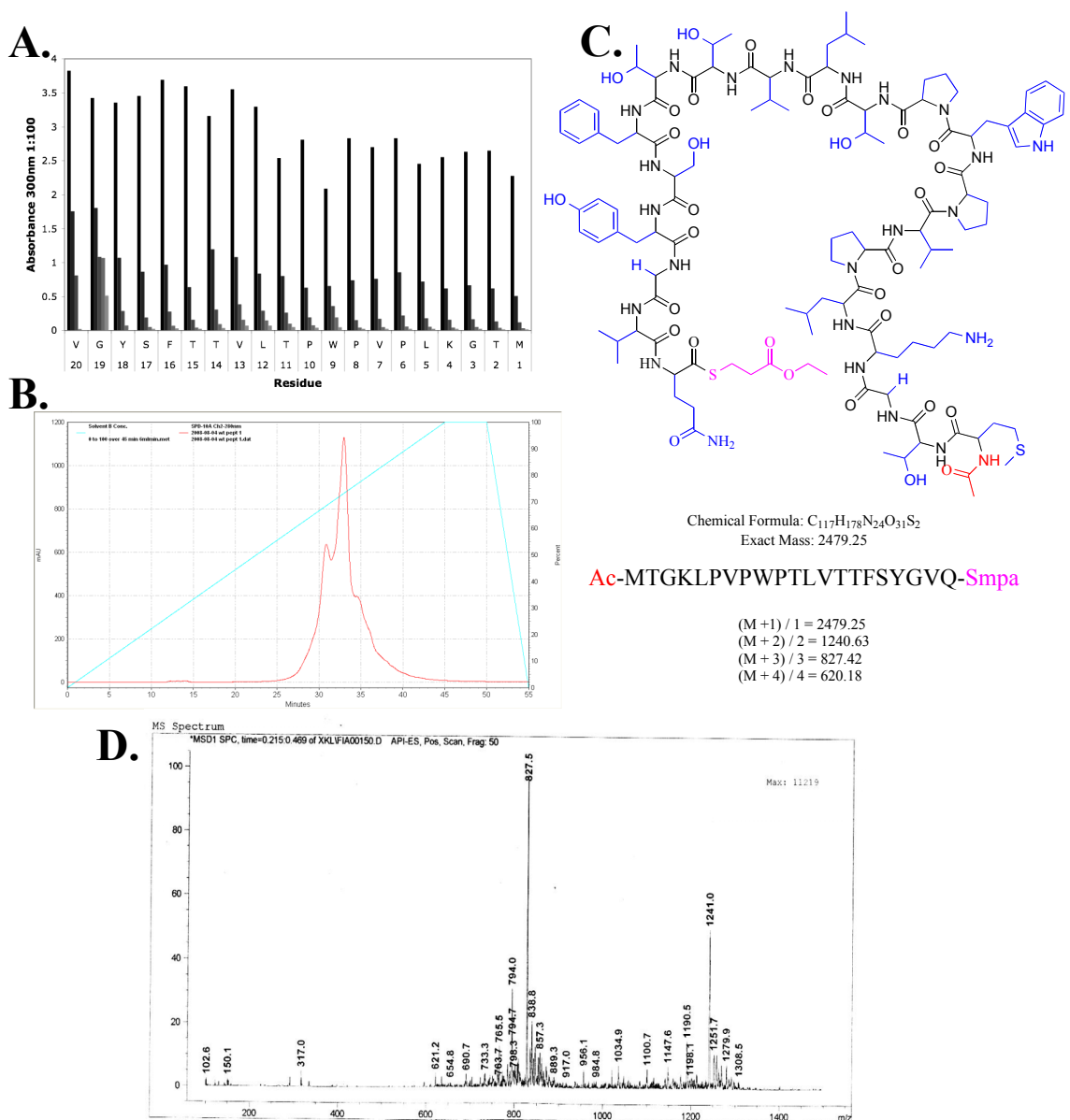


Figure 3.13. Characterization of the synthesis of thioester peptides on chlorotriyl resin. **A.** Absorbance measurement of each amino-acid deprotection step by release of the Fmoc adduct with 20% piperidine. Four subsequent washes of the resin with DMF are also shown. Samples have been diluted 1:100 in methanol to permit measurement. **B.** HPLC purification of the thioester peptide product; linear gradient shown in teal, absorbance at 280 nm in red. The major peak at 33 minutes was isolated and lyophilized. **C.** Structure of the thioester peptide with side chains in blue, N-terminal acetyl cap in red, and the C-terminal thioester in pink. Calculated molecular weights for the M through M+4 mass peaks are listed and correspond well to the M+1 and (M+2)/2 mass peaks seen in a sample of the 33 minute HPLC elute peak in **D.**

and the product confirmed by ESI mass spectrometry (Figure 3.13D). These purified peptide thioesters were then used as reactants in native chemical ligation.

Native chemical ligation and *de novo* folding

Native chemical ligation reactions were set-up as described and used the amounts of thioester peptide and N-terminal cysteine 50T70GFP α A206K in Table 3.1. Each protein and peptide thioester pair was dissolved in ligation buffer to 250 μ M and mercaptophenyl acetic acid added to initiate the ligation reaction. After an initial sampling, reactions were permitted to continue for 24 hours.

Table 3.1 Protein and peptide amounts used in native chemical ligation reaction.

Amino-Acid derivative at Y66	N-terminal Cys 50T70GFP α A206K in mg (nmol)	C-terminal peptide thioester in mg (nmol)	Molar ratio
Tyrosine	11.9 (478)	11.68 (4711)	9.85

Even with the use of a reactive water-soluble thiol catalyst, ligation reactions did not go to completion (Figure 3.14). Longer incubation times may drive the reaction further, but longer exposure to denaturants may also have deleterious effects on the protein. While ligation reactions of equal sized peptides are traditionally purified and isolated by HPLC, the ligation of thioester peptides to 50T70GFP α A206K produces only a small change in mass that was not separable from unligated starting materials [8, 22, 24]. For this reason, ligation reactions were directly diluted to approximately 0.128 mg/mL in urea denaturation buffer and incubated for 1 hour at room temperature to equilibrate the ligation product in this buffer. Protein was then

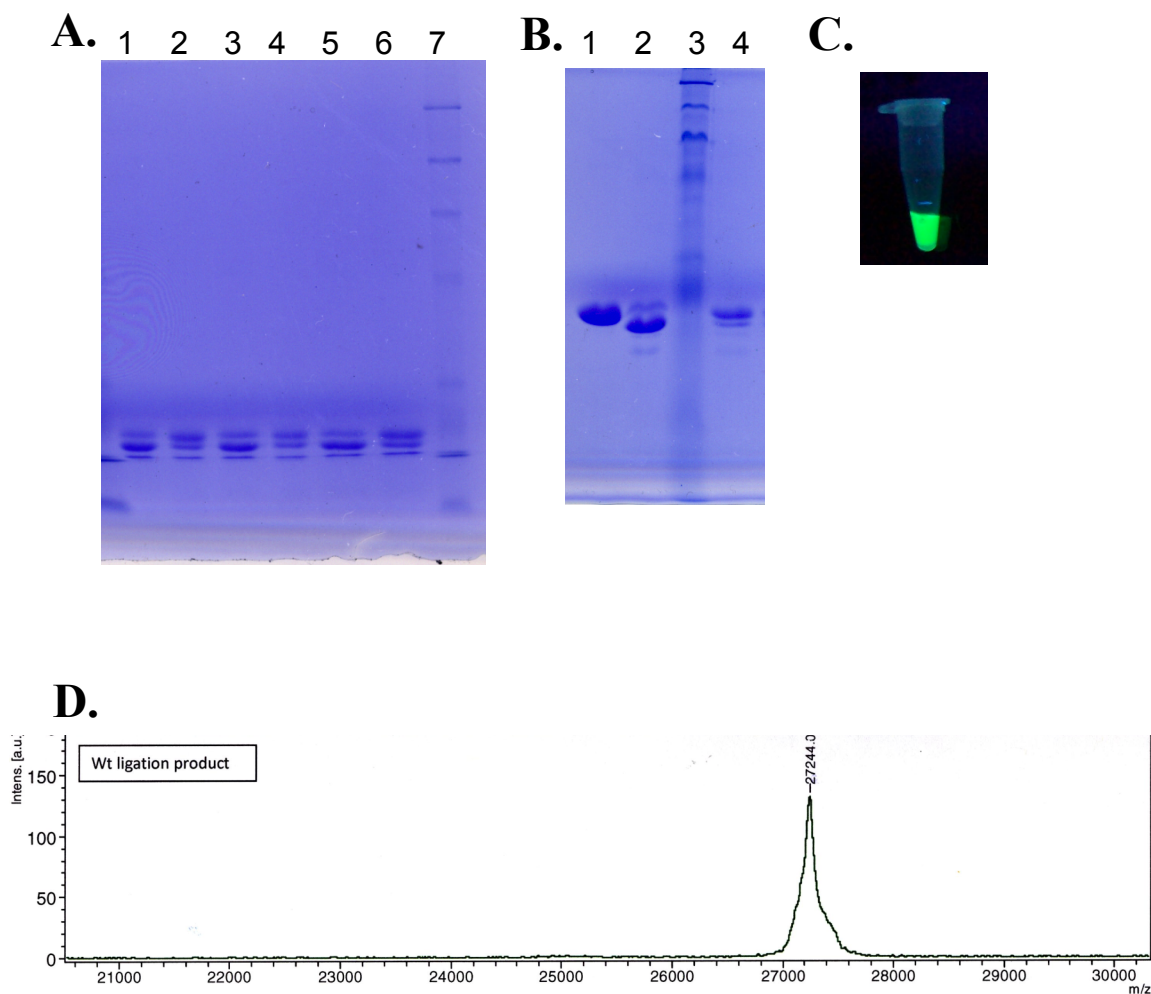


Figure 3.14. Native chemical ligation of the peptide thioester to 50T70GFP α A206K. **A.** Thioester peptides and the N-terminal cysteine protein 50T70GFP α A206K were ligated in denaturing solution using a water soluble thiol catalyst for 24 hours. Samples are taken when the reaction was initiated and terminated. Gel bands (1) Wild-type thioester ligation T=0, (2) T=24hrs., (3) N-methyl tyrosine ligation T=0, (4) T=24hrs., (5) Pseudodipeptide ligation T=0, (6) T=24hrs., (7) NEB Broad Range Ladder. **B.** ligation product was *de novo* folded and concentrated. Samples are as follows; (1) 50-GFP α A206K-entero, (2) 50T70GFP α A206K, (3) NEB Broad Range Ladder, (4) wt peptide thioester to 50T70GFP α A206K **C.** Visible evaluation of chromophore formation in the *de novo* folded ligation sample by exposure to 365 nm transillumination. **D.** MALDI mass analysis of *de novo* folded ligation product; expected mass 27238, found 27244

directly folded by dilution into rapidly stirring buffer and incubated overnight at 4°C to permit chromophore formation. The dilute folded buffer was applied to Ni²⁺-NTA resin to bind the ligated protein and effectively concentrated it. Washing and elution with imidazole yielded the ligated folded protein that was subsequently concentrated and evaluated by absorbance and fluorescence spectroscopy.

The folded wt ligation product shows visible fluorescence on transillumination and the characteristic absorption peaks for GFP, with absorption maxima at 397 nm and 475nm and fluorescence emission at the usual 503 nm and 508 nm, respectively (Figures 3.15A & B, 3.14 C). Although the ratio of chromophore absorption peaks to protein absorbance at 280 nm is low, indicating portions of ligated protein that does not have chromophore, the presence of any peaks indicates that the ligation product does form chromophore upon *de novo* folding and illustrates that this method is feasible. Additionally, the peaks within both the absorption and fluorescence emission scans match to that of the natively expressed 50-GFP α A206K and wtGFP, suggesting that the critical hydrogen bonding network important for fluorescence is maintained in the ligation product. PAGE analysis of the folded ligation product shows a single band at a mass that matches the uncut 50-GFP α A206K entero sample run as a standard for comparison and shows little to no protein at lower molecular weight relative to the ligation reaction mixture at 24hrs (Figure 3.14 B). These results indicate that the *de novo* folding and subsequent Ni²⁺-NTA affinity purification effectively eliminates some of the leftover unligated starting material. This assertion is supported by MALDI mass analysis of the ligation

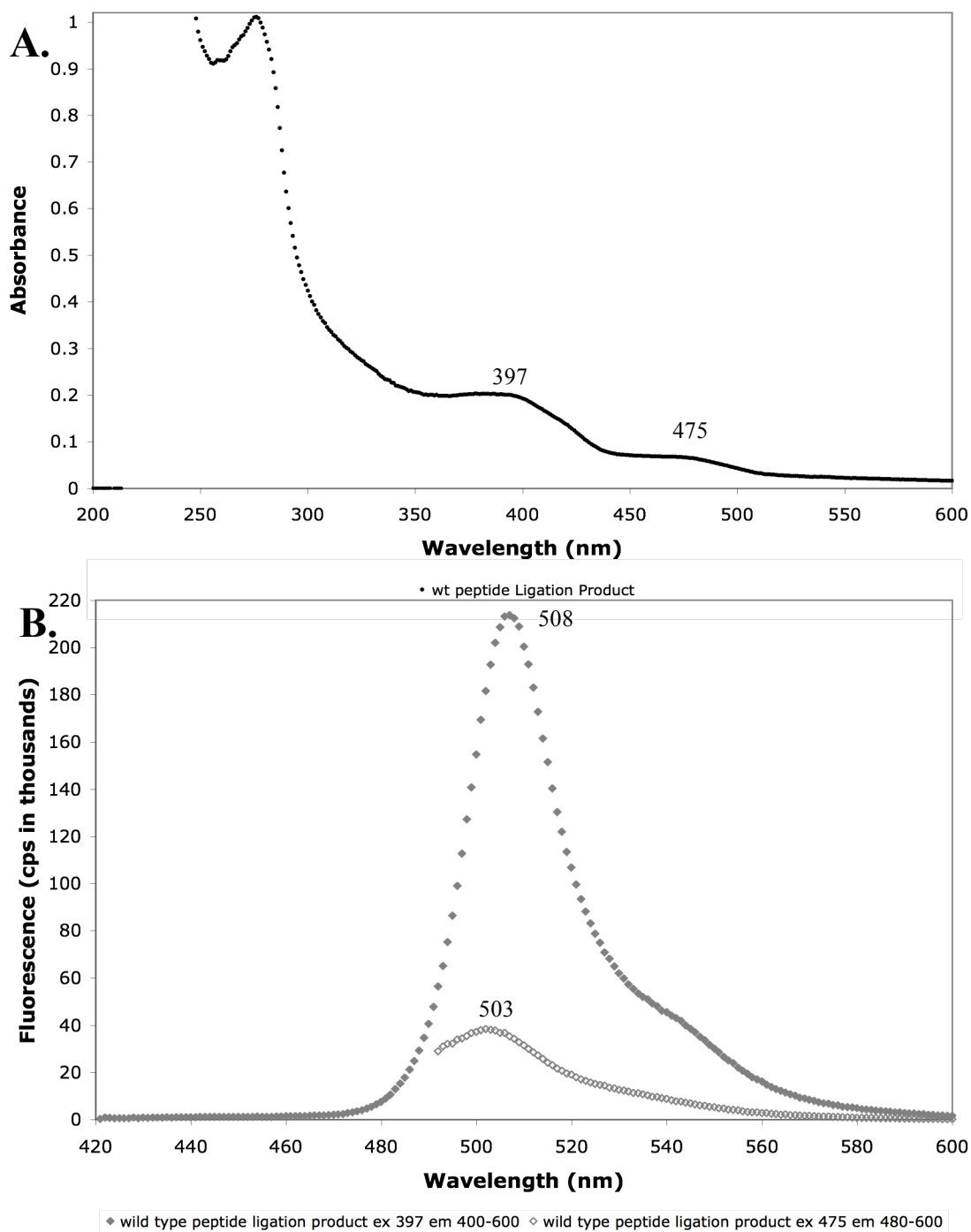


Figure 3.15. Absorption and fluorescence spectra of folded ligation product. *de novo* folded peptide ligation product was concentrated to approximately 1.5 mL; absorption spectrum in (A) and fluorescence emission in (B) with listed excitation (ex) and emission (em) wavelengths. Absorption and emission maxima are listed adjacent to each peak.

product, where only a single peak corresponding to the predicted mass + 6 Da is seen (Figure 3.14D). Yields, based on the initial cut 50T70GFP α A206K used in this reaction, protein 280 nm absorption (assuming all protein is ligated product), and chromophore absorption in the neutral band, indicates that of the 477.7 nmols reacted, 76.67 nmols are in folded into solution (16.0%), and 14.28 nmols show chromophore (3.0%). Further optimization of the ligation conditions and the *de novo* folding methods may improve yields, but the method established thus far suffices for preliminary studies.

Summary

The native chemical ligation reaction requires one peptide with an N-terminal cysteine residue and another with a C-terminal thioester. Several techniques were attempted to generate the N-terminal cysteine peptide (50T70GFP α A206K) in significant yields. Initially, the truncated protein was cloned into the pTwin1 system (New England Biolabs), in frame with the chitin binding domain tagged *Ssp* DnaB intein. While this construct did express fusion protein of 50 kDa within the cell, the full fusion protein had limited solubility, even with expression at reduced temperatures. Additionally, the fusion construct produced little soluble truncated protein. This result would suggest that the truncated protein itself is insoluble in solution and further work must be carried out under denaturing conditions. Since we are removing the central core helix of the GFP barrel, the instability in the 50T70GFP α A206K truncation protein may be due to a complete inability of the protein to fold and find a stable soluble conformation without the central helix.

Efforts to generate the 50T70GFP α A206K truncation protein were therefore carried out under mildly denaturing conditions to allow endoproteolytic digestion of the tag ahead of Cys70 while maintaining the truncated protein construct in a soluble form. While Factor Xa proteolysis was attempted in 1% Triton X-100 to maintain the Factor Xa sequence tagged 50-GFP α A206K construct in solution, protein precipitated out of solution into a gelatinous mass upon addition of the enzyme. It

appears the proteolytic digestion of the central helix out of 50GFP α A206K construct destabilizes the protein barrel so much that 1% Triton X-100 is not enough keep the protein in solution. The use of urea above 250mM with Factor Xa is not recommended so other proteolytic enzymes were explored[36].

Enterokinase is able to tolerate a much broader set of buffer conditions and still maintain enzyme activity. As the 50-GFP α A206K construct remains soluble in a 1.5 molar urea buffer, enterokinase reactions were attempted with this level of denaturant. Additionally, the reaction was carried out at 4°C because the enzyme cleaved the enterokinase tag more efficiently at this temperature. To terminate the reaction and decrease the volume of reaction mixture (and numbers of HPLC injections), the reaction was flash frozen and lyophilized. Resuspended in 50% acetonitrile to maintain protein solubility, purification via a C18 HPLC isolated the needed 50T70GFP α A206K protein. The novel method demonstrated here permits the high yield isolation and enzymatic processing of a normally insoluble protein inclusion body into a purified, N-terminal cysteine peptide suitable for native ligation.

Methods for the generation of peptide thioesters needed for the native chemical ligation reaction were explored. Initial attempts were based on a modification of the backbone amide linker methodology. This modified method linked the side chain of a glutamate residue to resin, while retaining the carboxylic acid C-terminus as an allyl protected group. Synthesis of complete 21-amino acid peptide begins at the amine of this residue. The N-terminus of the peptide was capped using acetic anhydride and the C-terminal allyl group was deprotected. While

several methods to were attempted to generate a thioester at this position, only starting the carboxylic acid was isolated.

We developed a method to liberate the peptide from the resin in a fully protected form, thereby leaving only the carboxylic acid C-terminus in solution for esterification. Peptide synthesis was carried out on chlorotriyl resin and the protected peptide was liberated from resin under mildly acidic conditions (hexafluoroisopropanol:chloroform). This protected peptide was then esterified in good yield and purified by HPLC. Native chemical ligation of this thioester peptide with the endoprotease cut 50T70GFP α A206K fragment will then provide the protein sequence that can be *de novo* folded to study the pathways of GFP chromophore formation and investigate the photochemistry within GFP.

The native chemical ligation reaction was then used to join the previously expressed N-terminal cysteine protein and synthetically generated C-terminal thioester peptide. Initial reactions used established peptide ligation methods that had to be modified to accommodate the large 50T70GFP α A206K fragment. Reactions were therefore run at lower protein concentrations, slightly basic pH, and with several fold excess of thioester peptide using a water-soluble thiol catalyst. Ligation reactions were run for 24 hours and reactions were terminated by dilution into 8 molar urea denaturation buffer. Attempts to isolate the ligated product prior to dilution into urea buffer via HPLC did not produce elution peaks of pure ligated product, most likely due to the similarity in size, shape, and charge of the ligated and

unligated products. *De novo* folding of the urea buffer diluted ligation reaction yielded the full circular permutant with some protein forming chromophore. While ligation yields for each derivative were enough for an absorption spectrum, fluorescence measurements, and mass spectral analysis, not enough protein was generated for measurements requiring high concentrations of protein (i.e. Infrared spectroscopy, RAMAN, x-ray crystallography). Further optimization and accumulation of multiple rounds of ligation and *de novo* folding may yield enough protein for these studies. The absorption spectrum of the wild-type peptide ligation product while showing a high 280 nm absorption, does display the characteristic GFP absorption peaks at 397 nm and 475 nm, indicating that the folded ligation products behave photophysically as the natively expressed circular permutant 50-GFP α A206K and wtGFP do.

References

1. Hirata, R., et al., *Molecular structure of a gene, VMA1, encoding the catalytic subunit of H(+)-translocating adenosine triphosphatase from vacuolar membranes of Saccharomyces cerevisiae*. J. Biol. Chem., 1990. **265**(12): p. 6726-6733.
2. Paulus, H., *The chemical basis of protein splicing*. Chem. Soc. Rev., 1998. **27**: p. 375-386.
3. Muir, T.W., D. Sondhi, and P.A. Cole, *Expressed protein ligation: A general method for protein engineering*. Proc. Natl. Acad. of Sci. U. S. A., 1998. **95**(12): p. 6705-6710.
4. Dawson, P.E., et al., *Synthesis of Proteins by Native Chemical Ligation*. Science, 1994. **266**(5186): p. 776-779.
5. Tam, J.P., et al., *Peptide synthesis using unprotected peptides through orthogonal coupling methods*. Proc. Natl. Acad. Sci. U.S.A., 1995. **92**(26): p. 12485-9.
6. Camarero, J., A. Adeva, and T. Muir, *3-Thiopropionic acid as a highly versatile multidetachable thioester resin linker*. Lett. Pept. Sci., 2000. **7**(1): p. 17-21.
7. Muralidharan, V. and T.W. Muir, *Protein ligation: an enabling technology for the biophysical analysis of proteins*. Nat. Meth., 2006. **3**(6): p. 429-438.
8. Clippingdale, A.B., C.J. Barrow, and J.D. Wade, *Peptide thioester preparation by Fmoc solid phase peptide synthesis for use in native chemical ligation*. J. Pept. Sci., 2000. **6**(5): p. 225-34.
9. Blanco-Canosa, J.B. and P.E. Dawson, *An efficient Fmoc-SPPS approach for the generation of thioester peptide precursors for use in native chemical ligation*. Angew. Chem. Int. Ed. Engl., 2008. **47**(36): p. 6851-5.
10. Brik, A., E. Keinan, and P.E. Dawson, *Protein synthesis by solid-phase chemical ligation using a safety catch linker*. J. Org. Chem., 2000. **65**(12): p. 3829-35.

11. Jensen, K.J., et al., *Backbone Amide Linker (BAL) Strategy for Solid-Phase Synthesis of C-Terminal-Modified and Cyclic Peptides* 1,2,3. J. Am. Chem. Soc., 1998. **120**(22): p. 5441-5452.
12. Dolphin, G.T., *A designed well-folded monomeric four-helix bundle protein prepared by Fmoc solid-phase peptide synthesis and native chemical ligation*. Chemistry, 2006. **12**(5): p. 1436-47.
13. Ingenito, R., et al., *Solid Phase Synthesis of Peptide C-Terminal Thioesters by Fmoc/t-Bu Chemistry*. J. Am. Chem. Soc., 1999. **121**(49): p. 11369-11374.
14. Flavell, R.R., et al., *Efficient semisynthesis of a tetraphosphorylated analogue of the Type I TGFbeta receptor*. Org. Lett., 2002. **4**(2): p. 165-8.
15. Backes, B.J., D.R. Dragoli, and J.A. Ellman, *Chiral N-Acyl-tert-butanesulfinamides: The "Safety-Catch" Principle Applied to Diastereoselective Enolate Alkylations*. J. Org. Chem., 1999. **64**(15): p. 5472-5478.
16. Futaki, S., et al., *Preparation of Peptide Thioesters using Fmoc-Solid-Phase Peptide Synthesis and its Application to the Construction of a Template-Assembled Synthetic Protein (TASP)*. Tet. Lett., 1997. **38**(35): p. 6237-6240.
17. Kitagawa, K., et al., *Total chemical synthesis of large CCK isoforms using a thioester segment condensation approach*. Tetrahedron, 2004. **60**(4): p. 907-918.
18. Shogren-Knaak, M.A., C.J. Fry, and C.L. Peterson, *A Native Peptide Ligation Strategy for Deciphering Nucleosomal Histone Modifications*. J. Biol. Chem., 2003. **278**(18): p. 15744-15748.
19. von Eggelkraut-Gottanka, R., et al., *Peptide [alpha]thioester formation using standard Fmoc-chemistry*. Tet. Lett., 2003. **44**(17): p. 3551-3554.
20. Warren, J.D., et al., *Toward Fully Synthetic Glycoproteins by Ultimately Convergent Routes: A Solution to a Long-Standing Problem*. J. Am. Chem. Soc., 2004. **126**(21): p. 6576-6578.
21. Mezo, A.R., R.P. Cheng, and B. Imperiali, *Oligomerization of uniquely folded mini-protein motifs: development of a homotrimeric betabetaalpha peptide*. J. Am. Chem. Soc., 2001. **123**(17): p. 3885-91.
22. Camarero, J.A., et al., *Chemical ligation of unprotected peptides directly from a solid support*. J. Pept. Res., 1998. **51**(4): p. 303-16.

23. Johnson, E.C. and S.B. Kent, *Insights into the mechanism and catalysis of the native chemical ligation reaction*. J. Am. Chem. Soc., 2006. **128**(20): p. 6640-6.
24. Dawson, P.E., et al., *Modulation of Reactivity in Native Chemical Ligation through the Use of Thiol Additives*. Journal of the American Chemical Society, 1997. **119**(19): p. 4325-4329.
25. Lee, J.Y., B.J. Byun, and Y.K. Kang, *Conformational Preferences and pKa Value of Cysteine Residue*. J. Phys. Chem. B, 2008. **112**(36): p. 11189-11193.
26. Danford, J.J., A.M. Arif, and L.M. Berreau, *Thioester Hydrolysis Promoted by a Mononuclear Zinc Complex*. Inorg. Chem., 2009. **49**(3): p. 778-780.
27. Kaiser, E., et al., *Color test for detection of free terminal amino groups in the solid-phase synthesis of peptides*. Anal. Biochem, 1970. **34**(2): p. 595-8.
28. Hackenberger, C.P.R., M.M. Chen, and B. Imperiali, *Expression of N-terminal Cys-protein fragments using an intein refolding strategy*. Bioorg. & Med. Chem., 2006. **14**(14): p. 5043-5048.
29. Bollhagen, R., et al., *A new reagent for the cleavage of fully protected peptides synthesised on 2-chlorotrityl chloride resin*. J. Chem. Soc. Chem. Comm., 1994(22): p. 2559 - 2560.
30. Barlos, K., D. Gatos, and W. Schafer, *Synthese von Prothymosin alpha (ProTalpha), einem aus 109 Aminosäuren aufgebauten Protein*. Angew. Chem., 1991. **103**(5): p. 572-575.
31. Barlos, K., et al., *Darstellung geschützter peptid-Fragmente unter Einsatz substituierter triphenylmethyl-Harze*. Tet. Lett., 1989. **30**(30): p. 3943-3946.
32. Barlos, K., et al., *2-Chlorotrityl chloride resin. Studies on anchoring of Fmoc-amino acids and peptide cleavage*. Int. J. Pept. Prot. Res., 1991. **37**(6): p. 513-20.
33. Barlos, K., et al., *Application of 2-chlorotrityl resin in solid phase synthesis of (Leu15)-gastrin I and unsulfated cholecystinin octapeptide. Selective O-deprotection of tyrosine*. Int. J. Pept. Prot. Res., 1991. **38**(6): p. 555-61.
34. Barlos, K., et al., *Solid phase synthesis of partially protected and free peptides containing disulphide bonds by simultaneous cysteine oxidation-release from 2-chlorotrityl resin*. Int. J. Pept. Prot. Res., 1991. **38**(6): p. 562-8.

35. Hackeng, T.M., J.H. Griffin, and P.E. Dawson, *Protein synthesis by native chemical ligation: expanded scope by using straightforward methodology*. Proc. Natl. Acad. Sci. U.S.A., 1999. **96**(18): p. 10068-73.
36. Novagen User Protocol TB205 Factor Xa Kits
http://www.emdchemicals.com/chemdat/en_CA/Merck-US-Site/USD/ViewProductDocuments-File?ProductSKU=EMD_BIO-69037&DocumentType=USP&DocumentId=%2Femd%2Fbiosciences%2Fuserprotocols%2Fen-US%2FTB205.pdf&DocumentSource=GDS

Chapter 4

Use of unnatural amino acids to probe the mechanism of chromophore formation in green fluorescent protein

Introduction.....	107
Methods.....	110
Results and Discussion.....	119
Summary.....	129
References.....	131

Introduction

The previous chapter detailed the methods used to incorporate a synthetic peptide into the central helix of the GFP protein. Here, we use this method to generate synthetic peptides with unnatural amino acids that will be used to probe the formation of the GFP chromophore. To review, the proposed chromophore formation pathway within GFP follows three distinct steps; cyclization by amine attack of Gly67 to the carbonyl of Ser65, dehydration of the carboxyl at Ser65 to create a cyclic imine, and oxidation of the C α -C β side chain bond of Tyr66 to form the full chromophore [1-6]. Both the order of each modification step and the intermediate structures that are generated are still debated (Figure 4.1).

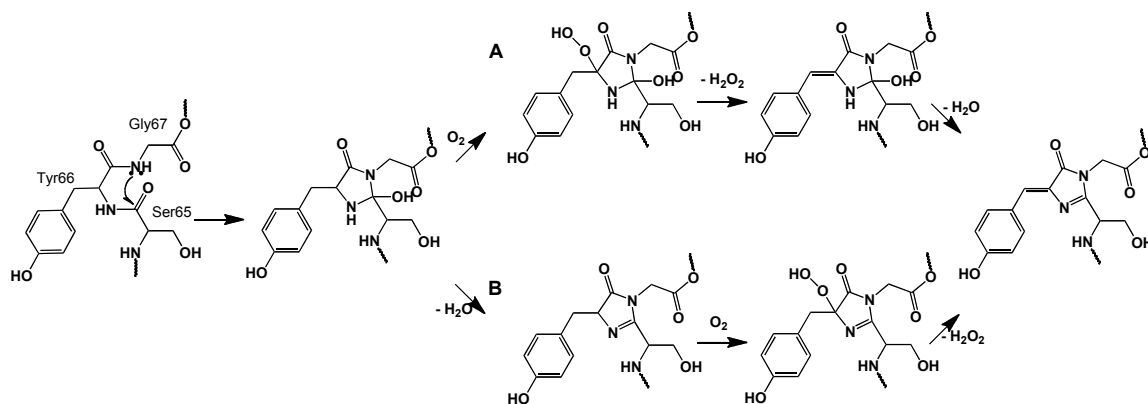


Figure 4.1 Current models of GFP chromophore formation. The strained central helix within the GFP protein barrel undergoes an initial cyclization reaction. Maturation of the chromophore follows either an oxidation-dehydration mechanism (A) or a dehydration-oxidation mechanism (B).

The use of unnatural amino acids at the Tyr66 position will permit the interruption of the chromophore maturation process. Incorporating an amino methyl tyrosine will prevent imine formation between the amine of the tyrosine and the carbonyl of the serine, thereby preventing the dehydration at the serine carbonyl. This modification preventing the dehydration step will test if the oxidation step is still possible. Similarly, the removal of the tyrosine carbonyl group prevents enolate formation at this position, proposed as essential for the oxidation at the C α of Tyr 66. This modification may even prevent the initial cyclization reaction, as some models propose the enolate at this position initiates the cyclization reaction [1] (Figure 4.2). This modification may even prevent the initial cyclization reaction, as some models propose the enolate at this position initiates the cyclization reaction [1] (Figure 4.2). While an N-methyl tyrosine derivative suitable for solid phase synthesis was available commercially, the tyrosine lacking the carbonyl oxygen had to be generated synthetically as a conjugate with the adjacent glycine.

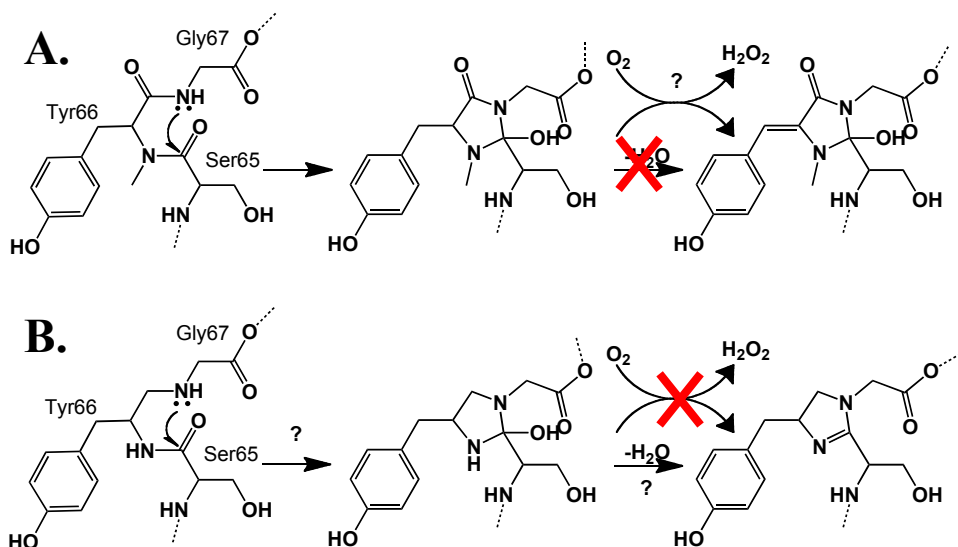


Figure 4.2 Proposed modifications to the Tyr66 backbone to interrupt chromophore formation. **A.** Introduction of an amino methyl tyrosine may prevent the dehydration step and evaluate if the oxidation step is still possible. **B.** Removal of the tyrosine carbonyl may prevent cyclization or dehydration.

Incorporation of these derivatives into peptide thioesters and native chemical ligation to truncated 50T70GFP α A206K will produce constructs that, upon *de novo* folding, contain structures in intermediate states of chromophore formation. These proteins may be studied by mass spectroscopy to elucidate the chromophore intermediates and shed light on the pathways of chromophore formation.

Methods

Insoluble Expression of Protein and Endoproteolytic Cleavage

Plasmid bearing the 50-GFP α A206K gene with the enterokinase recognition tag encoded N-terminal to Cys70 (p50-GFP α A206K-entero) was transformed into BL21(DE3)-pLysS cells and grown in 500 mL cultures of LB media supplemented with 200 μ g/mL of ampicillin and 30 μ g/mL of chloramphenicol at 37°C shaking 250 RPM for 2 ½ hours. The temperature was raised to 42°C and protein expression was induced with 0.8mM IPTG for 3 hours. Cells were harvested by centrifugation at 5,000 x g and stored at -20°C. Cells were resuspended in 50mM HEPES pH 7.9, 300mM NaCl, 5mM β -mercaptoethanol and passed through a French press 4 times. Whole cell lysate was centrifuged for 90 minutes at 125,000 x g and the soluble lysate fraction removed. The remaining inclusion body pellet was resuspended in 100mM Tris•HCl pH7.9, 500mM NaCl, 15mM EDTA, 5mM DTT, 2% Triton X-100 and centrifuged again for 30 minutes at 5,000 x g. The inclusion bodies were washed (i.e. resuspended and centrifuged at 5,000 x g) twice with 100mM Tris•HCl pH7.9, 500mM NaCl, 15mM EDTA, 5mM DTT, then washed twice with 50mM HEPES pH 7.9, 50mM NaCl, 1mM DTT, and finally flash frozen in 1.5 mL aliquots in liquid N₂ and stored at -80°C.

Thawed aliquots were solubilized in 13.5 mL 20mM Tris•HCl pH 7.9, 50mM NaCl, 2mM CaCl₂, 20mM methylamine, 1mM DTT (enterokinase buffer) supplemented with 8M Urea and incubated for one hour at room temperature. Protein was filtered through a 0.22 μM filter and dialyzed overnight to 1.5M urea in enterokinase buffer. The resulting protein was cleaved with the addition of 1:1000 U/ μg recombinant enterokinase (Novagen) or 0.000075% w/w recombinant enterokinase (New England Biolabs) at 4°C for 24 hours, flash frozen, and lyophilized. The dried protein was suspended in 50/49.9/0.1 acetonitrile/water/TFA and purified via HPLC on a Vydac protein and peptide C18 reverse phase column with a linear gradient from 95/4.9/0.1 water/acetonitrile/TFA to 95/4.9/0.1 acetonitrile/water/TFA over 47.5 minutes at 8 mL/min. Eluant that showed an intense 280 nm peak was collected, rotary evaporated, and lyophilized. This product was used subsequently for native chemical ligation.

Tyrosinyl-glycine Pseudo-dipeptide synthesis

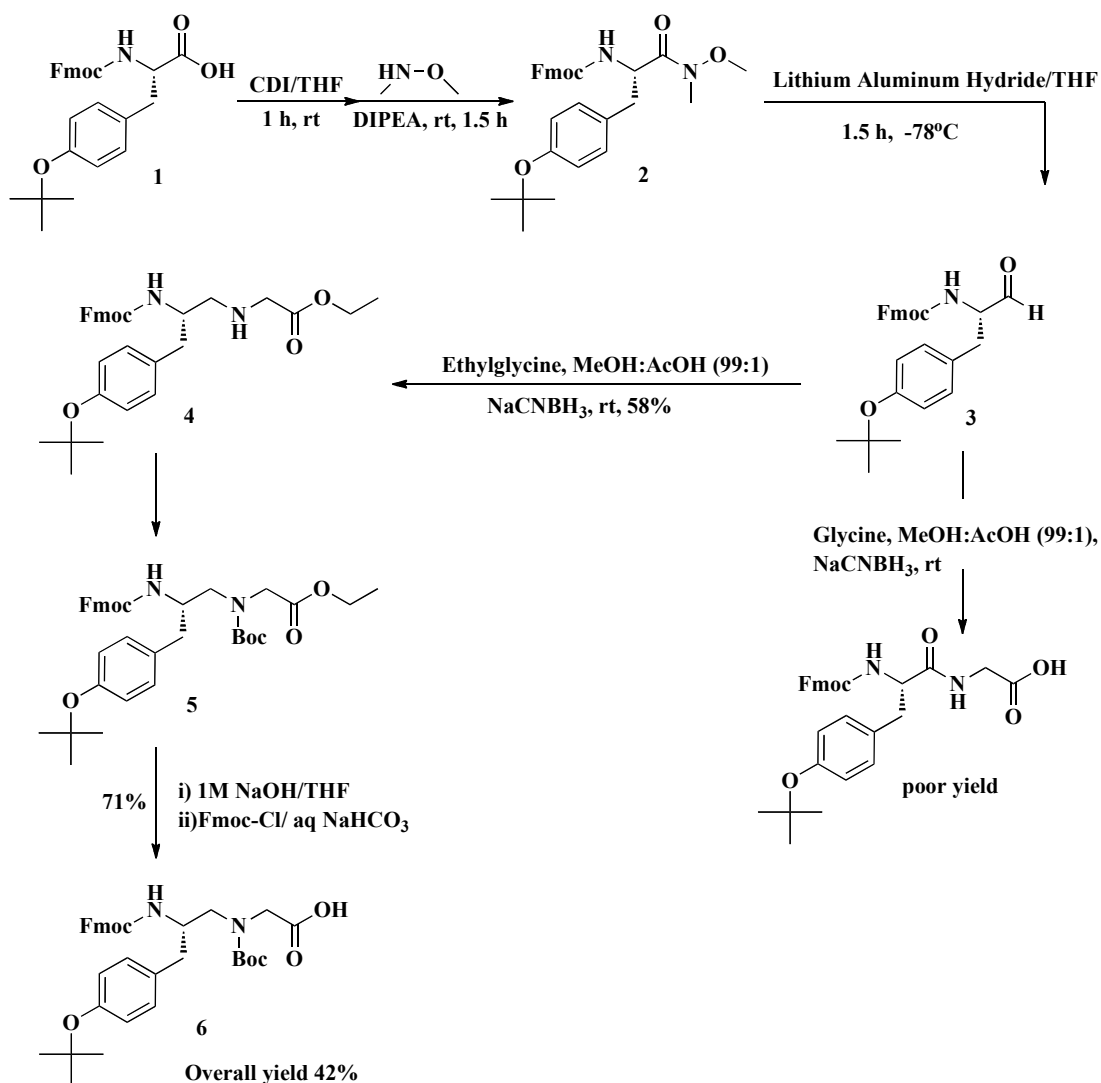


Figure 4.3 Synthesis of tyrosine-glycine pseudodipeptide product. Fmoc-tyrosine(tBu)-OH (**1**) is coupled with *N,O*-dimethylhydroxylamine generating a hydroxyamate (**2**) that is reduced with lithium aluminum hydride to form an aldehyde (**3**). Initial attempts to couple a glycine via reductive amidation to the aldehyde (**3**) gave unacceptable yields. Reductive amidation of ethylglycine with cyanoborohydride formed the pseudodipeptide (**4**). The secondary amine is Boc protected, the ethyl group of the glycine removed with base along with the Fmoc group producing (**5**). The Fmoc group is reintroduced to the primary amine generating the protected pseudodipeptide (**6**) in 42% yield overall.

Preparation of *N*-(Fluorenylmethyloxycarbonyl)-*L*-tyrosine (*tert*-butyl ester)-hydroxyamate (**2**)

Fmoc-Tyrosine(*tert*-butyl ester)-OH, 7.85 g (16.49 mmol), and 2.68 g (16.52 mmol) carbonyldiimidazole (CDI) was dissolved in approximately 75 mL of THF and stirred for 1 hour at room temperature. 2.88 mL (16.49 mmol) *N,N*-diisopropylethylamine (DIPEA) and 1.61 g (16.51 mmol) *N,O*-dimethylhydroxylamine•HCl was added and stirred for an additional 1.5 hours until shown complete by TLC. The reaction was rotary evaporated, suspended in dichloromethane, and successively extracted with saturated sodium bicarbonate, 10% hydrochloric acid, and brine. The organic layer was dried with magnesium sulfate and rotary evaporated, yielding a clear syrup. Resuspension in hexanes and rotary evaporation yielded white foam (**2**). Yield: 8.43 g (16.78 mmol), 100% yield. ESI mass expected $M+1$, 503.25; Found 503.3.

N-(Fluorenylmethyloxycarbonyl)-*L*-tyrosinal (*tert*-butyl ester) (**3**)

The resulting compound (**2**) 8.43 g (16.78 mmol) was added to 75 mL dry THF and cooled to -78 °C. 636.80 mg (16.78 mmol) of lithium aluminum hydride was added slowly portion-wise over an hour and allowed to react until completion by TLC. 3 volumes of diethyl ether were added to the reaction and the reaction was quenched with 3 volumes cold 20% citric acid. The organic layer was rotary

evaporated, resuspended in diethyl ether, and successively washed with saturated sodium bicarbonate and 10% citric acid. The ether layer was dried with magnesium sulfate, and evaporated to give compound (**3**). Yield: 7.76 g (17.51 mmol), 100 % yield. ESI mass expected M+1, 444.21; Found 444.1.

Preparation of *N*-(Fluorenylmethyloxycarbonyl)-L-tyrosyl (*tert*-butyl ester) amino ethyl-*N*-glycine ethyl ester (**4**)

The aldehyde product (**3**) 7.6 g (17.14 mmol) and 3.35 g (24.0 mmol) of glycine ethyl ester was stirred in 100 mL 99:1 methanol:acetic acid. 1.29 g (20.53 mmol) of sodium cyanoborohydride was added portion-wise over one hour and allowed to stir for an additional hour. The reaction was quenched with saturated sodium bicarbonate to pH 7 and methanol was rotary evaporated. The remaining solution was extracted with ethyl acetate and washed with brine. The organic layer was dried with sodium sulfate and rotary evaporated. Silica gel purification in petroleum ether:ethyl acetate yielded 5.32 g (10.03 mmol) of purified product (**4**), 58.5% yield. R_f 0.26 in 1:3 petroleum ether:ethyl acetate. ESI mass expected M+1 531.28; Found 531.3.

Preparation of *N*-(Fluorenylmethyloxycarbonyl)-L-tyrosyl (*tert*-butyl ester) amino ethyl-*N,N*-(*tert*-butyloxy carbamate) glycine ethyl ester (**5**)

The ethyl glycine –tyrosine product (**4**) 4.75g (8.97 mmol) was added to 100 mL dry dichloromethane and cooled to 0 °C on ice. Di-*tert*-butyl dicarbonate, 2.6 g

(11.92 mmol), and 1.75 mL (10.02 mmol) of DIPEA was added and allowed to react and warm to room temperature overnight. The reaction was extracted with water, saturated sodium bicarbonate, and 10% hydrochloric acid. The organic layer was dried with sodium sulfate and rotary evaporated to yield compound (5). Yield: 6.23 g (9.88 mmol). ESI mass expected M+1 631.3; Found 632.3.

Preparation of *N*-(Fluorenylmethyloxycarbonyl)-L-tyrosyl (*tert*-butyl ester) amino ethyl-*N,N*-(*tert*-butyloxy carbamate) glycine (6)

Compound (5), 6.23 g (9.88 mmol), was suspended in 100 mL 50:50 water:THF cooled to 0 °C and 1.82 g sodium hydroxide dissolved in water was added slowly over an hour and allowed to react overnight to room temperature. The reaction was acidified to pH 7 with 10% hydrochloric acid, rotary evaporated to remove THF, and lyophilized. The dried product was dissolved in 130 mL of 60:40 dioxanes:water, cooled to 0 °C, and 3.29 g (31 mmol) of sodium bicarbonate was added. 3.53 g (13.64mmol) of Fluorenylmethyloxycarbonyl chloride in 72 mL of dioxanes was added over 30 minutes and reacted overnight to room temperature. The reaction was quenched with ice cold 20% citric acid to pH 3 and extracted with ethyl ether. The organic layer was dried with magnesium sulfate and rotary evaporated. Purification on silica gel in methanol:dichloromethane yielded 4.23 g (7.02 mmol; 71% yield) of a white foam. R_f 0.15 in 5:95 methanol:dichloromethane. ESI mass expected M-1 601.30, 2M 1204.6; Found 601.3, 1204.6. Overall yield from (1) to (6) was 42.55% (Figure 4.3).

Synthesis of peptide thioesters via a protected peptide intermediate synthesized on chlorotryl resin

The 21-amino acid peptides were generated by manual peptide synthesis on 0.75 g of 2-chlorotryl resin pre-loaded with H-Gln(Trt) (Advanced Chemtech; 1 mmol/g substitution) in a 25 mL peptide synthesis reaction vessel. Fmoc-N-methyl-tyrosine(tBu)-OH was obtained from Chem-Impex. Fmoc-tyrosinyl(tBu, Boc)-glycine-OH pseudodipeptide was synthesized as described (*vide supra*). Resin was swollen in dimethylformamide (DMF; Fisher Scientific) for 45 minutes and drained. Fmoc-amino acids were coupled to the resin by pre-activating 3 equivalents of Fmoc-amino acid, 3 equivalents of 2-(1H-benzotriazole-1-yl)-1,1,3,3-tetramethylammonium hexafluorophosphate (HBTU), and 6 equivalents of N,N'-diisopropylethylamine (DIPEA; Fisher Scientific) in DMF for 10 minutes and then reacting with resin for 60 minutes, shaking 300 rpm. After extensive washing with DMF, Fmoc protecting groups were removed with 20% piperidine (Sigma-Aldrich) in DMF and washed again liberally with DMF. Fmoc-Trp(Boc)-OH was doubly coupled after ninhydrin testing revealed an incomplete coupling after a single reaction. The N-terminus was capped by reacting the deprotected amino-acid with 0.3M acetic anhydride (Sigma-Aldrich) in DMF. The resin was then successively washed with dichloromethane, DMF, isopropyl alcohol, methanol, and ethyl ether to remove any trace synthesis reagents and stored under nitrogen at 4°C.

The finished protected peptide was liberated from the resin by reacting with 1:4 hexafluoroisopropanol:chloroform for 1 hour. The suspension was filtered and the resin was washed thrice with chloroform. The flowthrough was collected and evaporated to dryness. The C-terminal carboxylic acid was esterified following a modified protocol reported by the Imperiali group with 4 equivalents of 3-mercaptopropionic acid ethyl ester (MPA; TCI America), 4 equivalents of HBTU, and 8 equivalents of DIPEA in DMF for 16 hours [7]. The resulting product was precipitated with six volumes of diethyl ether and pelleted by centrifugation. The pellet was washed with diethyl ether, centrifuged again, and dried under a stream of nitrogen gas. Amino-acid side chain protecting groups were removed by resuspending the dried thioester product in 95% TFA, 2.5% triisopropyl silane, 2.5% water for one hour. Precipitation with diethyl ether and centrifugation afforded a white pellet. This pellet was again washed with diethyl ether and centrifuged before suspending in water with 0.1%TFA and lyophilizing. The crude thioester peptide was further purified by C18 RP-HPLC and mass was confirmed by ESI-MS.

Native Chemical Ligation and *De Novo* Folding

Purified fragments of truncated 50T70GFP α A206K with an N-terminal cystein were combined with between 2 to 10 molar equivalent excess of purified 21-mer thioester peptides in 6M Guanidine•HCl, 0.1M sodium phosphate pH 8.0, 5mM 4-mercaptophenylacetic acid, 5mM TCEP to 0.25mM with respect to the truncated protein for 24 hours. Sampling 20 μ L of the reaction mixture at set intervals, isolating the protein fraction by TCA precipitation in 200 μ L 5% trichloroacetic acid, and evaluating resulting protein bands by SDS-PAGE, progress of the native chemical ligation was monitored. The ligation reaction mixture was diluted into 25mM MES pH 8.5, 8M Urea, 10mM EDTA, 0.1mM DTT so that the protein concentration was between 0.128 mg/mL and 0.0625 mg/mL and incubated at room temperature for one hour. The ligation reaction products were then folded by rapidly diluting dropwise into 100-fold 50mM Tris•HCl pH 8.5, 500mM NaCl, 1mM DTT. Folded protein was isolated from solution by passage over Ni²⁺-NTA resin, exchanging on resin into 20mM sodium phosphate pH 7.5, 300mM NaCl buffer, and then eluting with the same buffer supplemented with 250mM imidazole. The eluting protein was dialyzed against buffer without imidazole and concentrated to approximately 1.5 mL with centrifugal 3kDa MWCO filters.

Absorption measurements were taking on a BioCary 300 UV/visible spectrophotometer. Background absorption signals due to buffer were subtracted from spectra manually. Samples for MALDI mass spectrometry were prepared by C18 Zip-tip (Millipore).

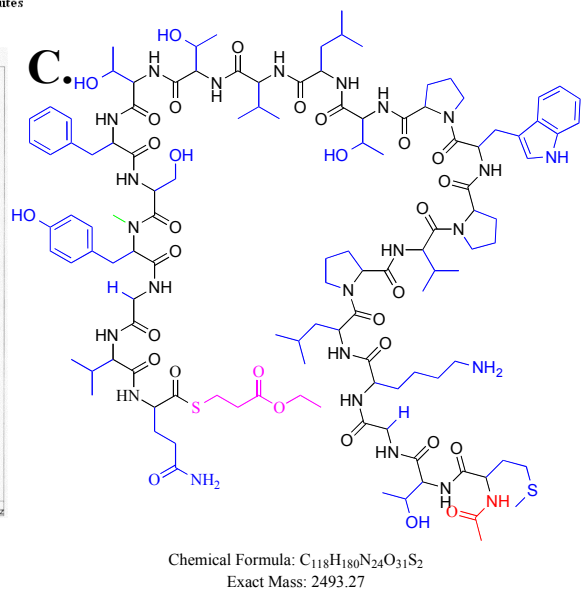
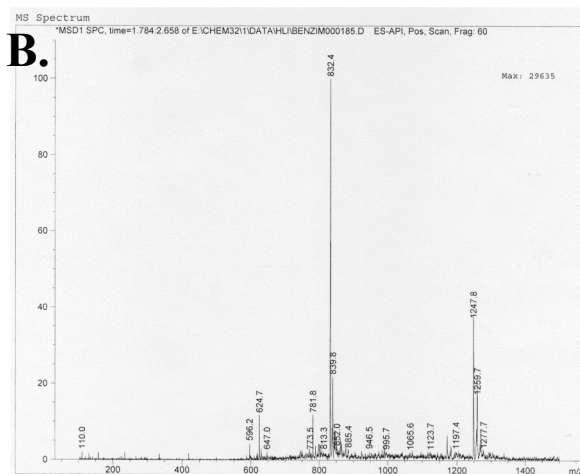
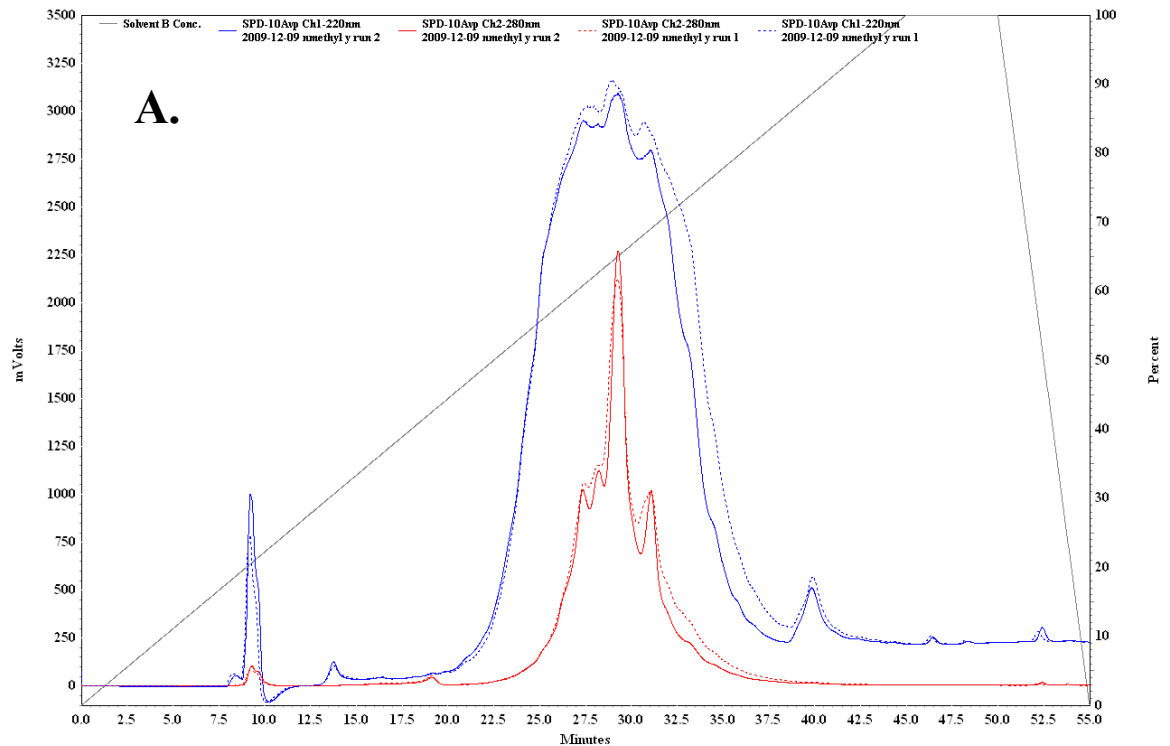
Results and Discussion

In an effort to study the chromophore formation process in GFP, we utilized a semi-synthetic native chemical ligation technique to incorporate unnatural amino acids into the central helix of the GFP protein. As previously described, the expression and endoproteolysis of a protein with an N-terminal cysteine was carried out under semi-denaturing conditions, yielding the 50T70GFP α A206K scaffold for native chemical ligation.

Prior to the generation of peptide thioesters, protected unnatural amino acids suitable for peptide synthesis were purchased (Fmoc-N-methyl-Tyrosine(tBu)-OH) or synthetically made. Synthesis of the tyrosinyl-glycine pseudo-dipeptide that lacks the carbonyl group on the tyrosine, was completed in sufficient yield through a 5 step synthesis. The original method followed a protocol from Salvi *et al* by coupling glycine to the Fmoc-Tyrosinal(tBu)-OH via sodium cyanoborohydride reductive amidation [8]. While this reaction was attempted under various organic and aqueous mixed solvent conditions, the major product was a double Fmoc-Tyrosinal(tBu)-OH coupling to the glycine molecule. Since the glycine zwitterion is not soluble in organic solutions and the Fmoc-Tyrosinal(tBu)-OH is preferably soluble in THF, any glycine that does solubilize into organic solution and couples to Fmoc-Tyrosinal(tBu)-OH will be preferably double coupled. The use of organic solvent

soluble ethylglycine as a coupling reagent alleviated this problem and produced compound **4** in good yield. As a SPPS accessible secondary amine, the bridging amine between the glycine and tyrosine was protected with a *tert*-butyloxycarbonyl group. Deprotection of the ethyl group with base generated the carboxylic acid of the glycine, but also deprotected the Fmoc group on the tyrosine. Reprotection of the tyrosine moiety with Fmoc-Cl yielded the pseudo-dipeptide adduct. ESI mass spectroscopy confirmed the generation of each intermediate and the final product (see Appendix 1).

The solid phase synthesis of peptide thioesters for native chemical ligation was carried out via a resin free protected peptide intermediate as described earlier, except these syntheses incorporated the N-methyl tyrosine and the tyrosine-glycine pseudodipeptide. ESI mass spectrometry confirmed the incorporation each unnatural amino acid into the final peptide thioester. Each peptide thioester was purified by HPLC, as was done with the wild type peptide thioester described earlier (Figure 4.4, 4.5).



Ac-MTGKLPVPWPTLVTTFS(NmeY)GVQ-Smpa

$$\begin{aligned} (M+1)/1 &= 2493.27 \\ (M+2)/2 &= 1247.64 \\ (M+3)/3 &= 832.09 \\ (M+4)/4 &= 624.32 \end{aligned}$$

Figure 4.4. HPLC purification of the N-methyl tyrosine containing thioester peptide. **A.** the major peak eluting at 29 minutes was isolated and lyophilized. **B.** the ESI mass of the 29 minute elute matches the calculated M+2/2 and M+3/3 for this peptide. **C.** Structure of the peptide thioester, with the thioester in pink and the N-methyl group in green.

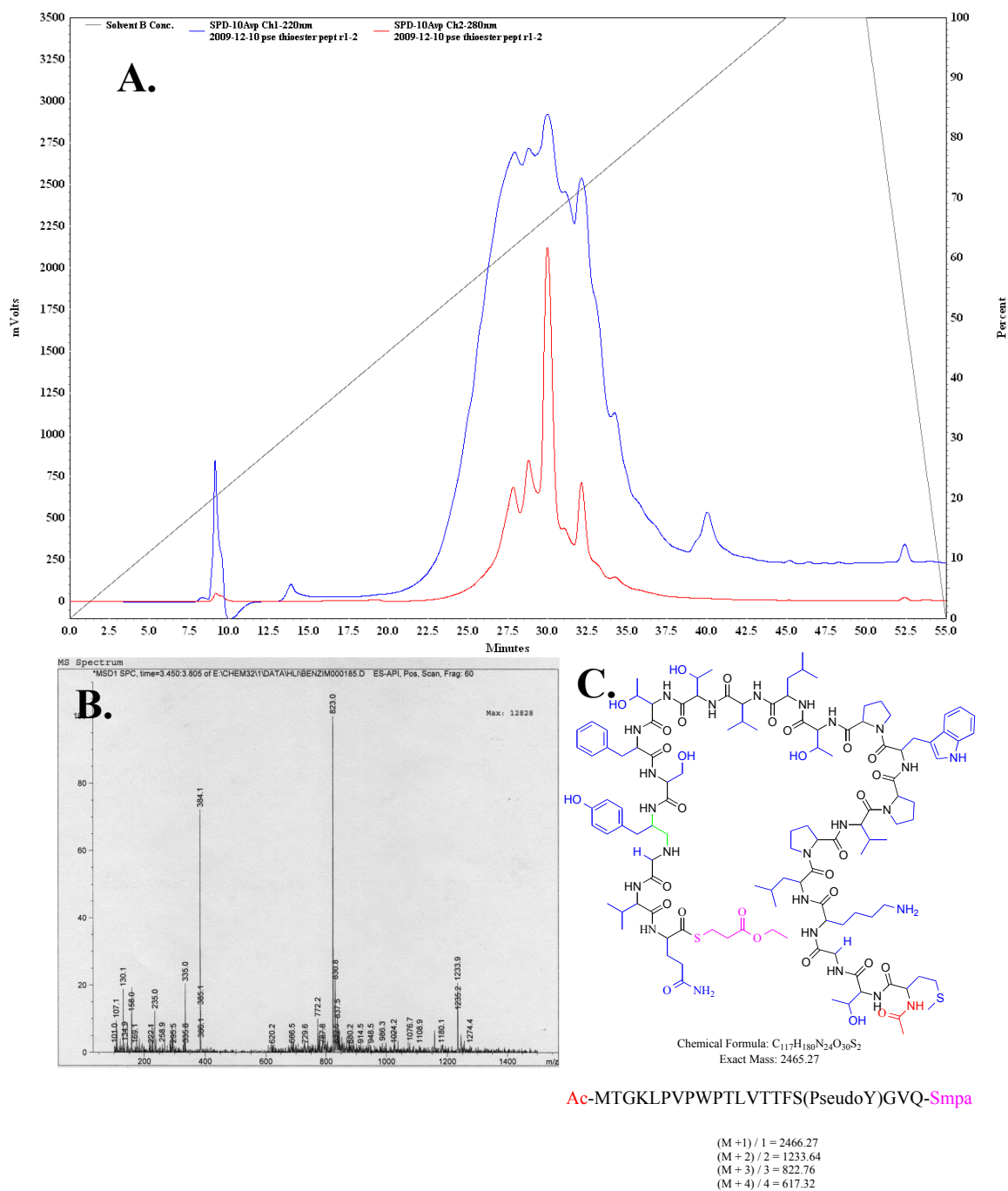


Figure 4.5. HPLC purification of the pseudo-dipeptide containing thioester peptide. **A.** the major peak eluting at 30 minutes was isolated and lyophilized. **B.** the ESI mass of the purified thioester peptide matches the calculated $M+2/2$ and $M+3/3$ for the peptide thioester. **C.** Structure of the pseudo-dipeptide thioester, thioester in pink, pseudo-dipeptide bond in green.

These purified peptides were then ligated in denaturing solution to the truncated 50T70GFP α A206K construct at 250 μ M for 24 hours, and the resulting protein was *de novo* folded (Table 4.1, Figure 4.6). The resulting protein was characterized by UV visible absorption and evaluated for interruptions in the chromophore forming process by mass spectrometry (Figure 4.7, 4.8).

Table 4.1. Protein and peptide amounts used in native chemical ligation reactions.

Amino-Acid derivative at Y66	N-terminal Cys 50T70GFP α A206K in mg (nmol)	C-terminal peptide thioester in mg (nmol)	Molar ratio
Tyrosine	11.9 (478)	11.68 (4711)	9.85
N-Methyl-Tyrosine	11.34 (455.2)	12.38 (4965)	10.9
Pseudo-Dipeptide Tyrosine	13.43 (539.1)	8.15 (2606)	4.8

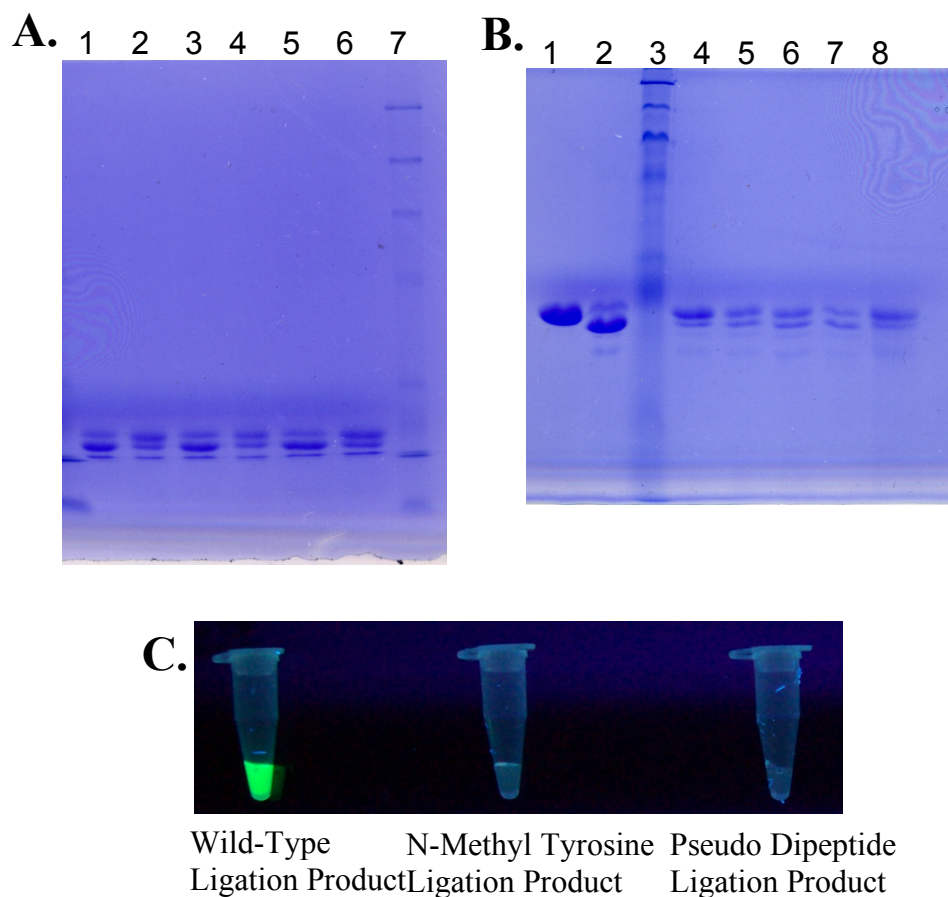


Figure 4.6. Native chemical ligation of the peptide thioester to 50T70GFP α A206K. **A.** Thioester peptides and the N-terminal cysteine protein 50T70GFP α A206K were ligated in denaturing solution using a water soluble thiol catalyst for 24 hours. Samples are taken when the reaction was initiated and terminated. Gel bands (1) Wild-type thioester ligation T=0, (2) T=24hrs., (3) N-methyl tyrosine ligation T=0, (4) T=24hrs., (5) Pseudodipeptide ligation T=0, (6) T=24hrs., (7) NEB Broad Range Ladder. **B.** ligation product was *de novo* folded and concentrated. Samples are as follows (1) 50-GFP α A206K-entero, (2) 50T70GFP α A206K, (3) NEB Broad Range Ladder, (4) wt peptide thioester to 50T70GFP α A206K, (5) N-methyl Tyrosine thioester to 50T70GFP α A206K, (6) Pseudo-dipeptide thioester to 50T70GFP α A206K **C.** Visible evaluation of chromophore formation in the *de novo* folded ligation sample by exposure to 365 nm transillumination.

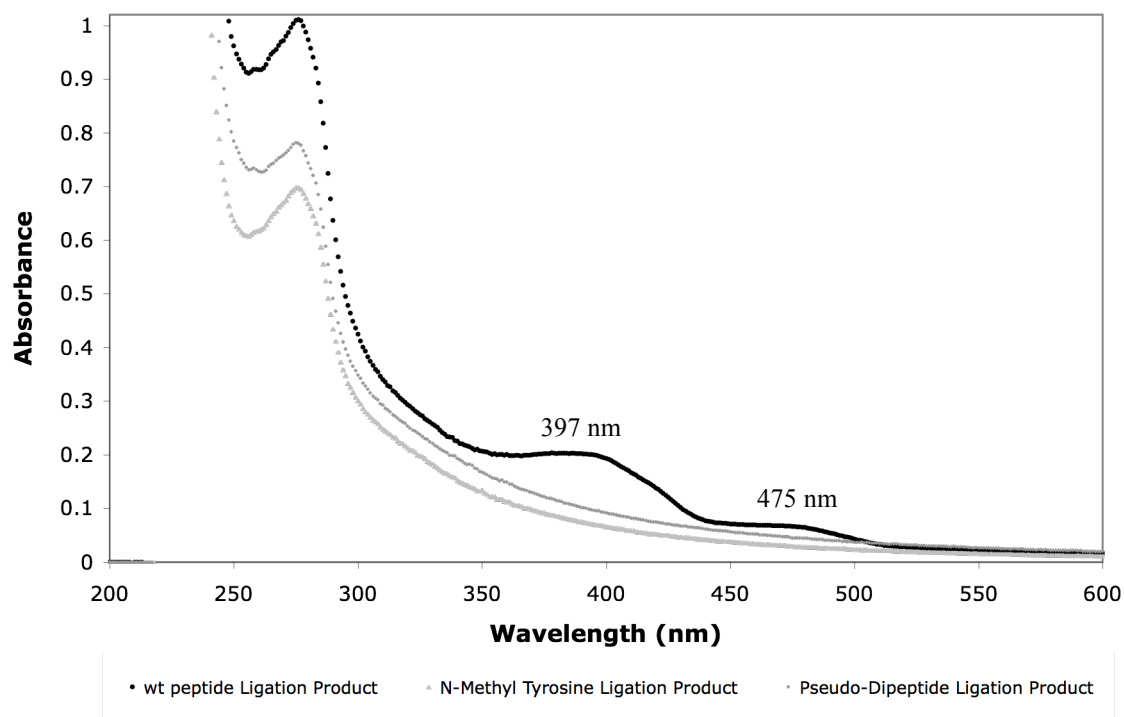


Figure 4.7 Absorption spectrum of the wild-type, N-methyl tyrosine, and pseudo-dipeptide ligation products after *de novo* folding. Absorption maxima for the wild-type sample are denoted, while the N-methyl tyrosine and the pseudo-dipeptide samples show no absorption peaks other than the 280 nm band corresponding to protein absorption.

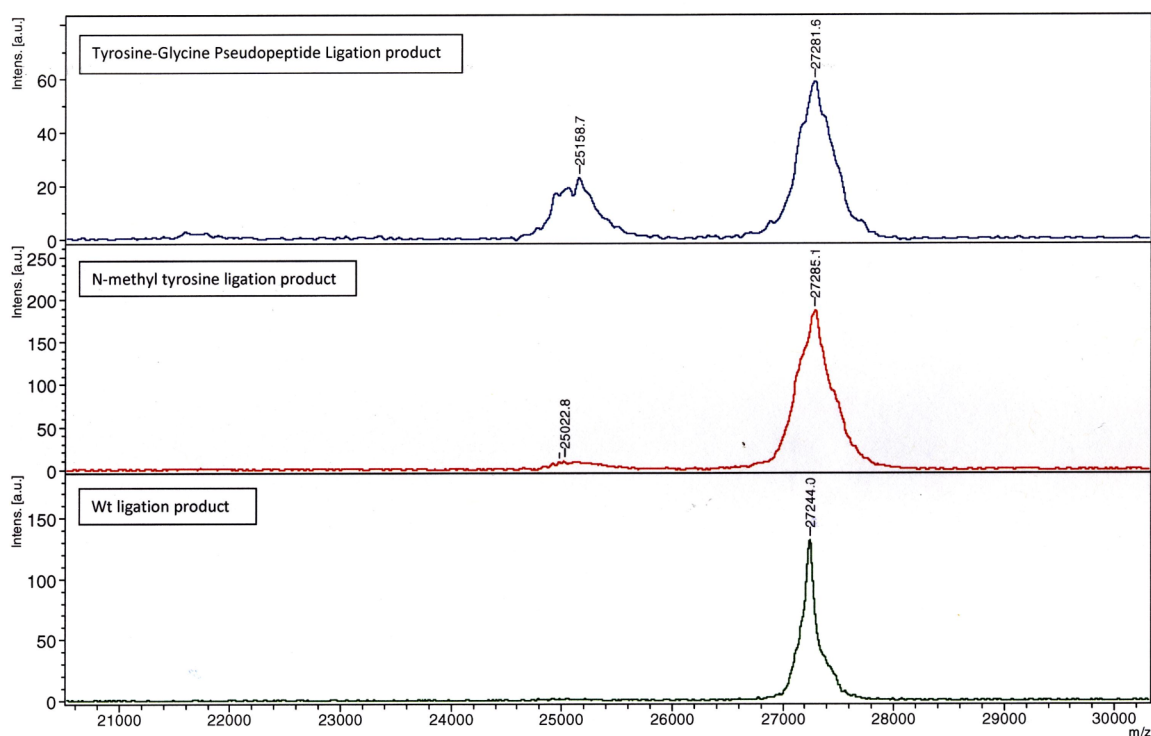


Figure 4.8 MALDI mass spectrum analysis of folded wild type peptide, N-methyl tyrosine, and pseudo-dipeptide ligation products. Ligated and *de novo* folded products with unnatural amino-acids that disrupt chromophore formation were purified from buffer using C18 Zip-Tips and submitted for MALDI mass analysis. The specific amino-acid derivatives introduced into the chromophore forming tyrosine are noted in the inset text boxes.

The N-methyl tyrosine derivative and the pseudo-dipeptide derivative show no absorption peak in the 300 to 600 nm range and no visible fluorescence emission indicating that chromophore formation has been interrupted. With no means of determining the intermediate chromophore structure based on absorption spectra within either the N-methyl tyrosine or the pseudo-dipeptide folded proteins, these samples were instead submitted for MALDI mass spectrometry. The mass for the wt product that is within 6 Da of the expected mass of 27238 (27216 (base amino acids) + 42 (N-acetyl) – 18 (dehydration) – 2 (oxidation)) found 27244. The mass of the N-

methyl tyrosine ligation product shows a mass of 27285, one Da difference from a structure that would predict a trapped peroxide with a dehydrated tyrosine intermediate in chromophore formation; $27286 (27230 \text{ (base amino acids)} + 42 \text{ (N-acetyl)} - 18 \text{ (dehydration)} + 32 \text{ (trapped peroxide)})$. These results indicate that the addition of a methyl group on the tyrosine that forms the chromophore does not prevent dehydration of the intermediate chromophore structure. This result is unexpected, because if the chromophore is in a dehydrated state, then the prechromophore structure must have formed a quaternary amine at either the glycine nitrogen or at the N-methyl modified tyrosine to accommodate the extra electron from the release of the hydroxide. This arrangement would create a positively charged imine that would need to rearrange to accommodate this added charge. Interestingly, the additional mass of 32 Da may correspond to a trapped peroxide intermediate at the C α of the tyrosine, although such a species would be highly unstable and the additional mass may simply be from oxidation of methionine residues within the protein. The addition of two oxygen atoms outside of the chromophore is not seen in the wild-type ligation product, which supports a trapped peroxide intermediate as the source of the additional mass. It may be possible that the methyl group on the tyrosine nitrogen is involved with the release of a peroxide moiety at the C α of the tyrosine after it is formed. With this nitrogen trapped in as a tertiary amine, the peroxide group may form, but not eject from the molecule.

The pseudo-dipeptide ligation product, lacking the carbonyl group of Tyr66, depicts a mass of 27281, corresponding to a mass of the protein with, again, a

potential peroxide intermediate but most likely no dehydration product; 27276 (27202 (base amino acids) + 42 (N-acetyl) + 32 (trapped peroxide)). In this case, though, we see no dehydration product suggesting that removal of the tyrosine backbone carbonyl either prevents the first cyclization step in chromophore formation (i.e. the carbonyl of the serine is still present) or prevents dehydration at this position. As with the N-methyl ligation product, the additional mass of two oxygens suggests either a peroxide intermediate at the C α of the tyrosine or two oxidized methionine residues (4 Met within the protein itself). Again, given that the wt ligation product (prepared identically to the unnatural amino-acid substitutions) shows no methionine oxidation, a peroxide adduct may be the cause of the added mass. If the protein does have a peroxide intermediate, with a hydrated intermediate chromophore, then removal of the carbonyl group from the tyrosine backbone prevents dehydration but does not prevent oxygen addition and peroxide formation. These results suggest that the carbonyl of the tyrosine is involved with the dehydration mechanism of the chromophore. Coupled with the unusual results seen from the N-methyl tyrosine adduct, the dehydration step may involve the formation of an imine along the C-N bond that initially cyclized the chromophore, generating a positive charge on the glycine nitrogen that may be dissipated through rearrangement of the imidazole ring, ending at the C α of the tyrosine that then traps the peroxide group. As these results are still preliminary, samples of these ligation products have been submitted for high-resolution mass spectroscopy to shed further light on the possible intermediate states seen in these molecules.

Summary

In this chapter, we use two unnatural amino acids to probe the formation of the GFP chromophore. These amino acids are incorporated into the critical Tyr66 position of the chromophore by a semi-synthetic native chemical ligation technique. We synthetically generated a tyrosine-glycine pseudodipeptide adduct that, along with an N-methyl tyrosine derivative, were incorporated into peptide thioesters. These thioesters were then ligated to a truncated version, 50T70GFP α A206K, of the previously established circular permutant of GFP, 50-GFP α A206K. *De novo* folding and analysis of the resulting protein ligation products by mass spectroscopy revealed unexpected chromophore intermediate structures. While the N-methyl tyrosine derivative was chosen to halt dehydration at the carbonyl of Ser65, the mass result from this ligated protein showed a dehydrated molecule. This implies that if the protein is dehydrated, it must have formed a positively charged cyclic imine at either the nitrogen of Gly67 or the N-methylated tyrosine. Rearrangement of this charge to the C α of the tyrosine may trap the peroxide group at this position, explaining the added mass of 32 Da for this sample. The added 32 Da mass is also seen in the pseudo-dipeptide product, but this sample may be in a hydrated form (as calculated by MALDI mass peaks). This data suggests that the carbonyl at position 65 is important for the dehydration reaction, since the absence of it prevents dehydration and, like the N-methyl tyrosine product. These data suggest interesting and alternate

pathways in GFP chromophore formation and will need further study to interpret our unexpected results.

References

1. Barondeau, D.P., et al., *Understanding GFP Chromophore Biosynthesis: Controlling Backbone Cyclization and Modifying Post-translational Chemistry*. *Biochemistry*, 2005. **44**: p. 1960-1970.
2. Reid, B.G. and G.C. Flynn, *Chromophore formation in green fluorescent protein*. *Biochemistry*, 1997. **36**(22): p. 6786-91.
3. Remington, S.J., *Fluorescent proteins: maturation, photochemistry and photophysics*. *Curr. Opin. Struct. Biol.*, 2006. **16**(6): p. 714-21.
4. Rosenow, M.A., et al., *The crystal structure of the Y66L variant of green fluorescent protein supports a cyclization-oxidation-dehydration mechanism for chromophore maturation*. *Biochemistry*, 2004. **43**(15): p. 4464-72.
5. 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. 12501-12504.
6. Pouwels, L.J., et al., *Kinetic isotope effect studies on the de novo rate of chromophore formation in fast- and slow-maturing GFP variants*. *Biochemistry*, 2008. **47**(38): p. 10111-22.
7. Mezo, A.R., R.P. Cheng, and B. Imperiali, *Oligomerization of uniquely folded mini-protein motifs: development of a homotrimeric betabetaalpha peptide*. *J. Am. Chem. Soc.*, 2001. **123**(17): p. 3885-91.
8. Salvi, J.-P., N. Walchshofer, and J. Paris, *Formation of bis (Fmoc-amino ethyl)-N-glycine derivatives by reductive amination of Fmoc-amino aldehydes with NaBH₃CN*. *Tet. Lett.*, 1994. **35**(8): p. 1181-1184.

Chapter 5

Use of unnatural amino acids to modulate the photophysical properties of green fluorescent protein

Introduction.....	133
Methods.....	135
Results and Discussion.....	144
Conclusions and Future Directions.....	155
References.....	160

Introduction

The semi-synthetic technique used in the previous chapters can also be extended to introduce novel functionality into the chromophore residues that do not disrupt the formation of the chromophore. These modifications then alter the spectral properties of the GFP chromophore and may elucidate interesting photophysical modifications within the protein matrix. Fluorination of the tyrosine ring lowers the pKa of the phenol group [1-3]. In the context of the GFP chromophore, a fluorinated tyrosine with reduced pKa would promote an ionized phenolate on the chromophore that would more easily release a proton through the ESPT network within GFP upon excitation. This modification may possibly shift the ground state ionization ratio of the GFP chromophore within the wild-type hydrogen-bonding network further to a B state from the favorably populated A state. While an ionized B form is seen in the S65T variant of GFP, this is due to the mutation that specifically disrupts the hydrogen bonding network in GFP by modifying the hydrogen bonding to Glu222 [4]. In a 3,5-difluorinated tyrosine, proposed here, the ionization state may change due to a lower pKa, but presumably, the hydrogen-bonding network would still be present in this wild-type context [5]. Spectroscopic studies on this variant would therefore prove interesting, especially if the hydrogen-bonding network is proven intact in this protein. An alternate modification, meta-hydroxyl-tyrosine, would relocate the phenol oxygen away from the usual para

position (Figure 5.1). The relocation of this group may have both implications on the formation of the chromophore, on the spectroscopy of this protein if it does form chromophore, and would be the first introduction of such a modification within the GFP protein.

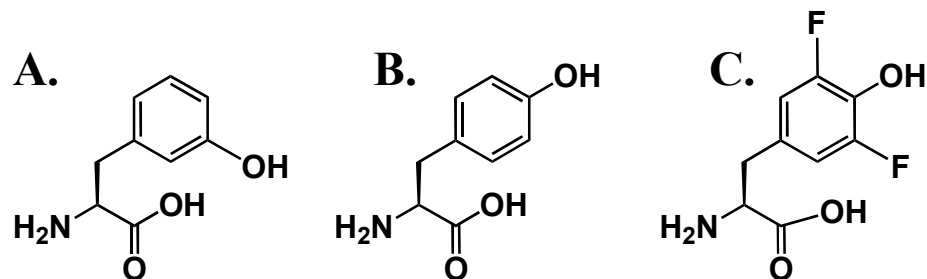


Figure 5.1. Proposed unnatural amino acids to incorporate into the GFP chromophore and modulate photophysics. A. meta-hydroxyl tyrosine relocates the hydroxyl group. B. native tyrosine, pK_a=10. C. 3,5-difluorotyrosine, pK_a=6.8[6].

As the fluorinated derivative of tyrosine was not commercially available, biosynthetic synthesis of this amino acid derivative was completed using tyrosine phenol lyase (TPL) from *Citrobacter Freundii* [2, 3, 7, 8]. After isolation from the enzymatic reaction, this fluorinated derivative was protected on the phenol group with a *tert* butyl group and protected on the backbone amine with an Fmoc group, rendering it usable in solid phase synthesis.

Methods

Insoluble Expression of Protein and Endoproteolytic Cleavage

Plasmid bearing the 50-GFP α A206K gene with the enterokinase recognition tag encoded N-terminal to Cys70 (p50-GFP α A206K-entero) was transformed into BL21(DE3)-pLysS cells and grown in 500 mL cultures of LB media supplemented with 200 μ g/ml of ampicilin and 30 μ g/ml of chloramphenicol at 37°C shaking 250 RPM for 2 ½ hours. The temperature was raised to 42°C and protein expression was induced with 0.8mM IPTG for 3 hours. Cells were harvested by centrifugation at 5,000 x g and stored at -20°C. Cells were resuspended in 50mM HEPES pH 7.9, 300mM NaCl, 5mM β -mercaptoethanol and passed through a French press 4 times. Whole cell lysate was centrifuged for 90 minutes at 125,000 x g and the soluble lysate fraction removed. The remaining inclusion body pellet was resuspended in 100mM Tris•HCl pH7.9, 500mM NaCl, 15mM EDTA, 5mM DTT, 2% Triton X-100 and centrifuged again for 30 minutes at 5,000 x g. The inclusion bodies were washed (i.e. resuspended and centrifuged at 5,000 x g) twice with 100mM Tris•HCl pH7.9, 500mM NaCl, 15mM EDTA, 5mM DTT, then washed twice with 50mM HEPES pH 7.9, 50mM NaCl, 1mM DTT, and finally flash frozen in 1.5 mL aliquots in liquid N₂ and stored at -80°C.

Thawed aliquots were solubilized in 13.5 mL 20mM Tris•HCl pH 7.9, 50mM NaCl, 2mM CaCl₂, 20mM methylamine, 1mM DTT (enterokinase buffer) supplemented with 8M Urea and incubated for one hour at room temperature. Protein was filtered through a 0.22 μM filter and dialyzed overnight to 1.5M urea in enterokinase buffer. The resulting protein was cleaved with the addition of 1:1000 U/ug recombinant enterokinase (Novagen) or 0.000075% w/w recombinant enterokinase (New England Biolabs) at 4°C for 24 hours, flash frozen, and lyophilized. The dried protein was suspended in 50/49.9/0.1 TFA and purified via HPLC on a Vydac protein and peptide C18 reverse phase column with a linear gradient from 95/4.9/0.1 water/acetonitrile/TFA to 95/4.9/0.1 acetonitrile/water/TFA over 47.5 minutes at 8 mL/min. Eluant that showed an intense 280 nm peak was collected, rotary evaporated, and lyophilized. This product was used subsequently for native chemical ligation.

Fluorinated Tyrosine Synthesis and Protection for SPPS

The *tpl* gene for tyrosine phenol lyase was amplified from *Citrobacter freundii* genomic DNA by polymerase chain reaction on whole *C. freundii* cells (ATCC: 29063) using the primers 5' CTAGCTAGCATGAATTATCCGGCAGAACCC 3' and 5' CCGCTCGAGGATATAGTCAAAGCGTGCAGT 3'. The PCR product and pET23b expression vector were digested with NheI and XhoI and cut fragments were isolated and extracted from a 2% agarose gel. The endonuclease cut *tpl* gene and

pET23b vector fragments were ligated together overnight and transformed into *E. coli* XL1-Blue cells. Colonies that grew on ampicillin selective media were grown in liquid culture and plasmid isolated via a Wizard mini-prep kit. Plasmids testing positive for insert were transformed into *E. coli* BL21(DE3)-pLysS, grown to mid-log phase in 0.5 L cultures, and induced with 0.8mM IPTG. Cell were isolated from media by centrifugation at 5,000 x g and resuspended in 100mM potassium phosphate pH 7.0, 150mM NaCl (TPL buffer), supplemented with 5mM imidazole, 5mM β -mercaptoethanol, and 0.1 mM pyridoxal 5' phosphate. Cells were lysed by sonication at 15W in 30-second intervals with 5 repetitions. Lysate was cleared by centrifugation at 125,000 x g, passed through a 0.22 μ m filter, and bound to a 5 mL bed volume of Ni²⁺-NTA resin. Protein was washed and eluted with TPL buffer supplemented with 20mM imidazole and 250mM imidazole, respectively. Eluted protein was dialyzed against 3 L of TPL buffer and concentrated to 3 mg/mL (6 units/mL)[9].

Synthesis of 3,5 difluortyrosine was carried out by adding 30 units of tyrosine phenol lyase in two portions over two days to a 4 L solution of 30mM ammonium acetate pH 8.0, 10mM 2,6-difluorophenol (5g), 60mM sodium pyruvate, 5mM β -mercaptoethanol, and 40 μ M pyridoxal-5-phosphate[3]. The reaction was allowed to proceed for four days at room temperature in the dark. The solution was acidified to pH 3 and filtered through a celite pad. Excess phenol was extracted twice with ethyl acetate and the aqueous phase loaded onto Dowex AG50W-X8 resin precharged with 1M HCl. The column was washed with 1L water, and tyrosine was eluted with 10% ammonium hydroxide. Fractions that showed positive ninhydrin tests were pooled,

rotary evaporated and lyophilized. 3,5-difluorotyrosine (**7**) yield; 546.05 mg (2.51 mmol), 6.53%. ESI mass expected M+1 218.06; found 218.0.

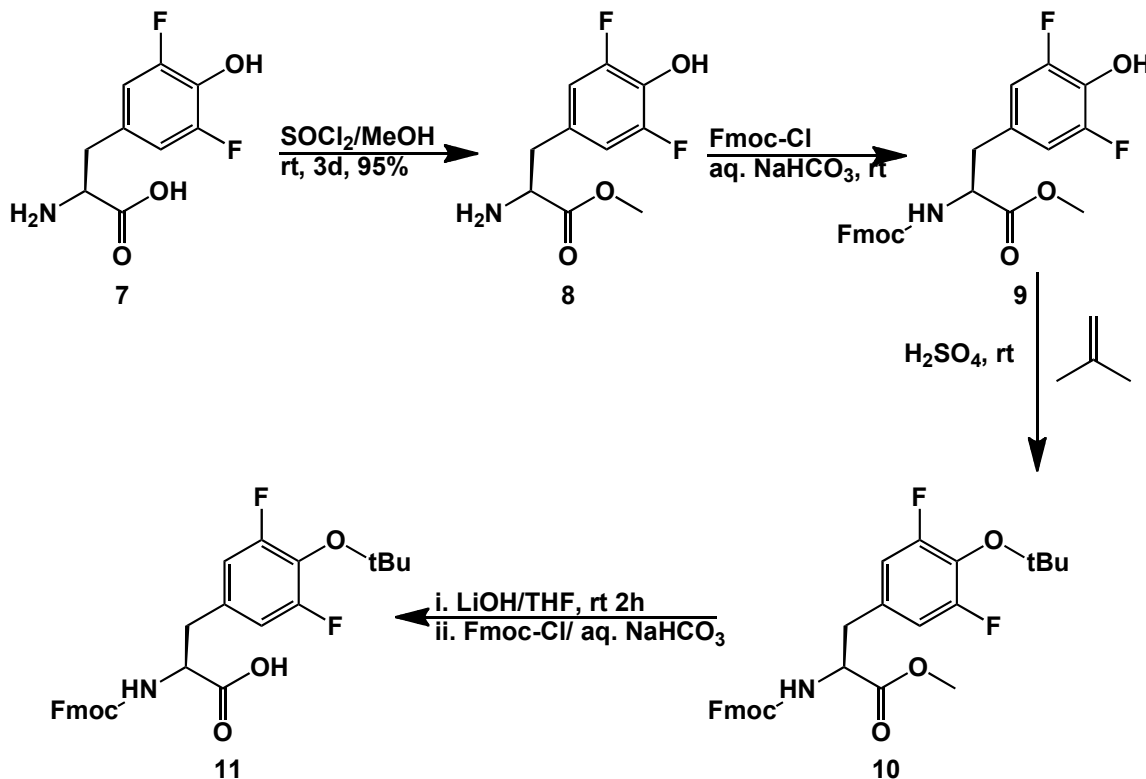


Figure 5.2. Synthesis of Fmoc-L-3,5-difluorotyrosine(tBu)-OH from L-3,5-difluorotyrosine. Biosynthetically generated L-3,5-difluorotyrosine (**7**) was methyl esterified (**8**) at the carboxylic acid and Fmoc protected (**9**) at the amine so that the remaining phenol could be protected with a tertiary butyl group (**10**). Deprotection of the methyl group in basic conditions and reprotection of the amine with Fmoc generated the SPPS compatible product (**11**).

Preparation of L-3,5-difluorotyrosine methyl ester. (**8**)

Methodology as listed in Seyedsayamdost et al was followed [3]. Briefly, L-3,5-difluorotyrosine was dissolved in methanol and thionyl chloride (10 equivalents) was added dropwise. The solution was stirred for 3 days at room temperature until

completion. Solvent was removed by rotary evaporation and diethyl ether added to precipitate the product.

Preparation of *N*-(9-Fluorenylmethoxycarbonyl)-L-3,5-difluorotyrosine methyl ester. **(9)**

L-3,5-difluorotyrosine methyl ester was dissolved in aqueous sodium bicarbonate and fluorenylmethoxycarbonyl chloride was added and stirred at room temperature until completion. The solution was acidified to pH 3 with 10% hydrochloric acid and extracted with ethyl acetate. The organic phase was rotary evaporated and provided the product.

Preparation of *N*-(9-Fluorenylmethoxycarbonyl)-*O*-*tert*-butyl-L-3,5-difluorotyrosine methyl ester. **(10)**

Methodology as listed in Gopishetty et al was followed [2]. Briefly, Fmoc-L-3,5-difluorotyrosine methyl ester was suspended in dichloromethane and acidified with sulfuric acid. Isobutylene gas was bubbled through the solution until the reaction was completed as judged by TLC. The reaction was quenched with addition of saturated sodium bicarbonate and extracted with ethyl acetate. The organic layer was rotary evaporated yielding the product.

Preparation of *N*-(9-Fluorenylmethoxycarbonyl)-*O*-*tert*-butyl-L-3,5-difluorotyrosine. (11)

Fmoc-L-3,5-difluorotyrosine(*t*Bu) methyl ester was suspended in 50:50 THF/H₂O at 0°C and lithium hydroxide added. The reaction was allowed to warm to room temperature and stirred until judged complete by TLC. The reaction was neutralized with 10% hydrochloric acid and extracted with ethyl acetate to remove byproducts. The aqueous fraction was lyophilized and resuspended in aqueous sodium bicarbonate and fluorenylmethoxycarbonyl chloride was added and stirred at room temperature until completion. The solution was acidified to pH 3 with 10% hydrochloric acid and extracted with ethyl acetate. The organic phase was rotary evaporated and provided the product.

Synthesis of peptide thioesters via a protected peptide intermediate synthesized on chlorotriyl resin

The 21-amino acid peptides were generated by automated peptide synthesis on 0.5 g of 2-chlorotriyl resin pre-loaded with H-Gln(*Trt*) (Advanced Chemtech; 1 mmol/g substitution) on a Liberty - 12-Channel Automated Peptide Synthesizer (CEM Corp.). Meta-hydroxyl (L)tyrosine(*t*Bu)-OH was obtained from RSP amino acids inc. Fmoc-(L)3,5Difluorotyrosine(*t*Bu)-OH was synthesized as described (*vide supra*). Fmoc-amino acids were coupled to the resin using 3 equivalents of Fmoc-amino acid, 3 equivalents of 2-(1H-benzotriazole-1-yl)-1,1,3,3-tetramethylammonium hexafluorophosphate (HBTU), and 6 equivalents of *N,N'*-diisopropylethylamine (DIPEA; Fisher Scientific) and deprotected with 20% piperidine (Sigma-Aldrich) as

per the instrument protocol. Fmoc-Trp(Boc)-OH was doubly coupled, as previous manual syntheses revealed inefficient coupling with one coupling step. The N-terminus was capped by reacting the deprotected amino-acid with 0.3M acetic anhydride (Sigma-Aldrich) in DMF. The resin was then successively washed with dichloromethane, DMF, isopropyl alcohol, methanol, and ethyl ether to remove any trace synthesis reagents and stored under nitrogen at 4°C.

The finished protected peptide was liberated from the resin by reacting with 1:4 hexafluoroisopropanol:chloroform for 1 hour. The suspension was filtered and the resin was washed thrice with chloroform. The flowthrough was collected and evaporated to dryness. The C-terminal carboxylic acid was esterified following a modified protocol reported by the Imperiali group with 4 equivalents of 3-mercaptopropionic acid ethyl ester (MPA; TCI America), 4 equivalents of HBTU, and 8 equivalents of DIPEA in DMF for 16 hours [10]. The resulting product was precipitated with six volumes of cold diethyl ether and pelleted by centrifugation. The pellet was washed with diethyl ether, centrifuged again, and dried under a stream of nitrogen gas. Amino-acid side chain protecting groups were removed by resuspending the dried thioester product in 95% TFA, 2.5% triisopropyl silane, 2.5% water for one hour. Precipitation with diethyl ether and centrifugation afforded a white pellet. This pellet was again washed with diethyl ether and centrifuged before suspending in water with 0.1%TFA and lyophilizing. The crude thioester peptide was further purified by C18 RP-HPLC and mass was confirmed by ESI-MS.

Native Chemical Ligation and *De Novo* Folding

Purified fragments of truncated 50T70GFP α A206K with an N-terminal cystein were combined with between 2 to 10 molar equivalent excess of purified 21-mer thioester peptides in 6M Guanidine•HCl, 0.1M sodium phosphate pH 8.0, 5mM 4-mercaptophenylacetic acid, 5mM tris-2-carboxyethyl phosphine (TCEP) to 0.25mM with respect to the truncated protein for 24 hours. Sampling 20 μ L of the reaction mixture at set intervals, isolating the protein fraction by TCA precipitation in 200 μ L 5% trichloroacetic acid, and evaluating resulting protein bands by SDS-PAGE, progress of the native chemical ligation was monitored. The ligation reaction mixture was diluted into 25mM MES pH 8.5, 8M Urea, 10mM EDTA, 0.1mM DTT so that the protein concentration was between 0.128 mg/mL and 0.0625 mg/mL and incubated at room temperature for one hour. The ligation reaction products were then folded by rapidly diluting dropwise into 100-fold 50mM Tris•HCl pH 8.5, 500mM NaCl, 1mM DTT. Folded protein was isolated from solution by passage over Ni-NTA resin, exchanging on resin into 20mM sodium phosphate pH 7.5, 300mM NaCl buffer, and then eluting with the same buffer supplemented with 250mM imidazole. The eluting protein was dialyzed against buffer without imidazole and concentrated to approximately 1.5 mL with centrifugal 3kDa MWCO filters.

Fluorescence measurements were taken on a Quanta Master fluorimeter (Photon Technology International). Absorption measurements were taking on a BioCary 300 UV/visible spectrophotometer. Background absorption and fluorescence signals

due to buffer were subtracted from spectra manually. Samples for MALDI mass spectrometry were prepared by C18 Zip-tip (Millipore).

Results and Discussion

Our aim in these studies was to incorporate unnatural amino acids into GFP that modulate the photophysics of the chromophore. By site-specifically introducing a fluorinated derivative of tyrosine into Tyr66, the pKa of the resulting chromophore is reduced markedly. Alternatively, the introduction of a *meta* hydroxyl-tyrosine displaces the phenol hydroxyl and may affect both the formation of the chromophore and the photophysics of the chromophore. The introduction of these two modified amino acids was accomplished through the semi-synthetic native chemical ligation method detailed earlier in this work. While strictly organic synthesis of fluorinated tyrosines is possible, such a synthesis is a multi-step reaction that produces racemic fluorinated tyrosine that is difficult to separate [11]. The stereoselective biosynthesis of fluorinated tyrosines was therefore produced from fluorinated phenol, pyruvic acid, and ammonia via a recombinantly expressed *C. freundii* tyrosine phenol lyase (TPL) enzyme [6-8, 12]. Reacting 2,6-difluorophenol with TPL yielded L-3,5-difluorotyrosine in 6.53% yield based on the amount of phenol added to the reaction (Figure 5.3). ESI mass spectrometry confirmed the identity and purity of the generated fluorotyrosine.

The fluorinated tyrosine was methyl protected at the carboxylic acid and Fmoc protected at the amine to allow *tert* butyl protection with isobutylene at the phenol group. Demethylation under basic conditions regenerated the carboxylic acid,

but also deprotected the amine. Subsequent Fmoc reprotection at the amine provided the protected amino-acid suitable for solid-phase synthesis.

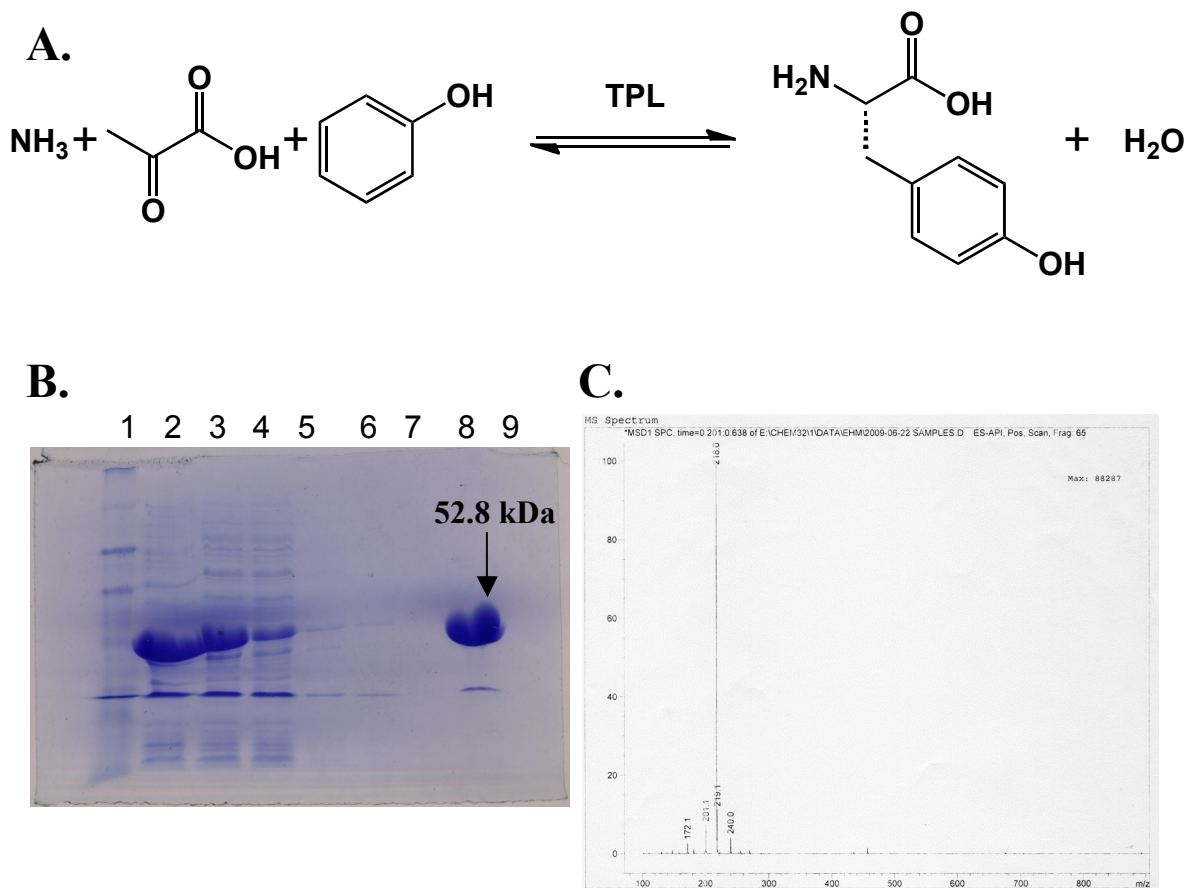
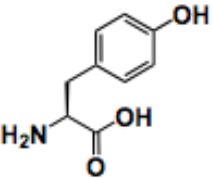
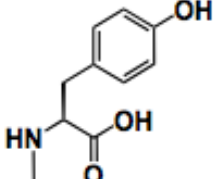
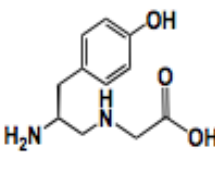
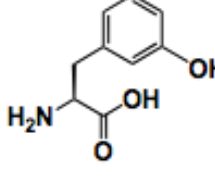
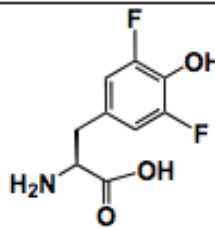


Figure 5.3. 3,5-difluorotyrosine biosynthesis using tyrosine phenol lyase. A. Reaction equation for tyrosine phenol lyase. Excess phenol and pyruvic acid substrates drive the reaction toward tyrosine synthesis. **B.** Recombinant expression profile of TPL, (1) New England Biolabs broad range ladder, (2) whole cell lysate, (3) soluble cell lysate, (4) Ni-NTA flowthrough, (5) 0mM Imidazole buffer wash, (6) 20mM Imidazole buffer wash, (7) empty lane, (8) concentrated sample of purified TPL (6.25mg/mL; 12.5U/mL). **C.** ESI mass spectrum of biosynthesized 3,5-difluorotyrosine; expected M+1 218.06, found 218.0.

Peptide thioesters with the designed unnatural amino-acids were generated by SPPS, with only the protected 3,5-difluorotyrosine derivative providing low final

Table 5.1. Overall yields of peptide thioesters generated for native chemical ligation. Each peptide incorporates a tyrosine at position 66 with either a side chain or a main-chain modification. Crude yields are determined after isolation of the peptide from the thioesterification reaction. Percent yield of each peptide is based on a 1mmol/g resin loading.

Amino-acid at Y66	Structure	mg resin	mg (μ mol) crude peptide thioester	mg (μ mol) HPLC purified peptide thioester	Percent Yield
Tyrosine		200	186.4 (59.3)	21.5 (8.67)	4.33%
N-methyl Tyrosine		200	178.2 (56.5)	20.7 (8.30)	4.15%
Glycine-Tyrosine Pseudodipeptide		200	170.1 (54.4)	18.8 (7.63)	3.81%
meta-hydroxyl Tyrosine		250	212.3 (67.5)	24.5 (9.88)	3.95%
3,5-difluorotyrosine		95	27.3 (8.59)	2.3 (0.914)	0.96%

yield (Table 5.1). Presumably, the phenol of 3,5-difluorotyrosine may be acidic enough to prematurely cleave the peptide from the acid sensitive chlorotryl resin (pKa 6.8 vs. pKa 10) over the course of solid phase synthesis [6]. Premature liberation of peptide from chlorotryl resin was seen in previous SPPS with 3,5-difluorotyrosine when only the amine of the 3,5-difluorotyrosine was protected. Incorporation of fluorinated tyrosine derivatives into a peptide thioester may require an alternate method of peptide synthesis using a more acid stable resin, such as unchlorinated trytil resin [13].

The native chemical ligation of these peptides to the enterokinase cut 50T70GFP α A206K scaffold was accomplished in the same manner detailed earlier, except that the 3,5-difluorotyrosine derivative was ligated at a lower molar ratio due to the limited amount of peptide thioester available (Table 5.2).

Table 5.2. Protein and peptide amounts used in native chemical ligation reactions.

Amino-Acid derivative at Y66	N-terminal Cys 50T70GFP α A206K in mg (nmol)	C-terminal peptide thioester in mg (nmol)	Molar ratio
Meta-Hydroxyl-Tyrosine	26.07 (1047)	27.14 (10950.0)	10.5
3,5-Difluorotyrosine	13.11 (526.2)	2.3 (720.0)	1.4

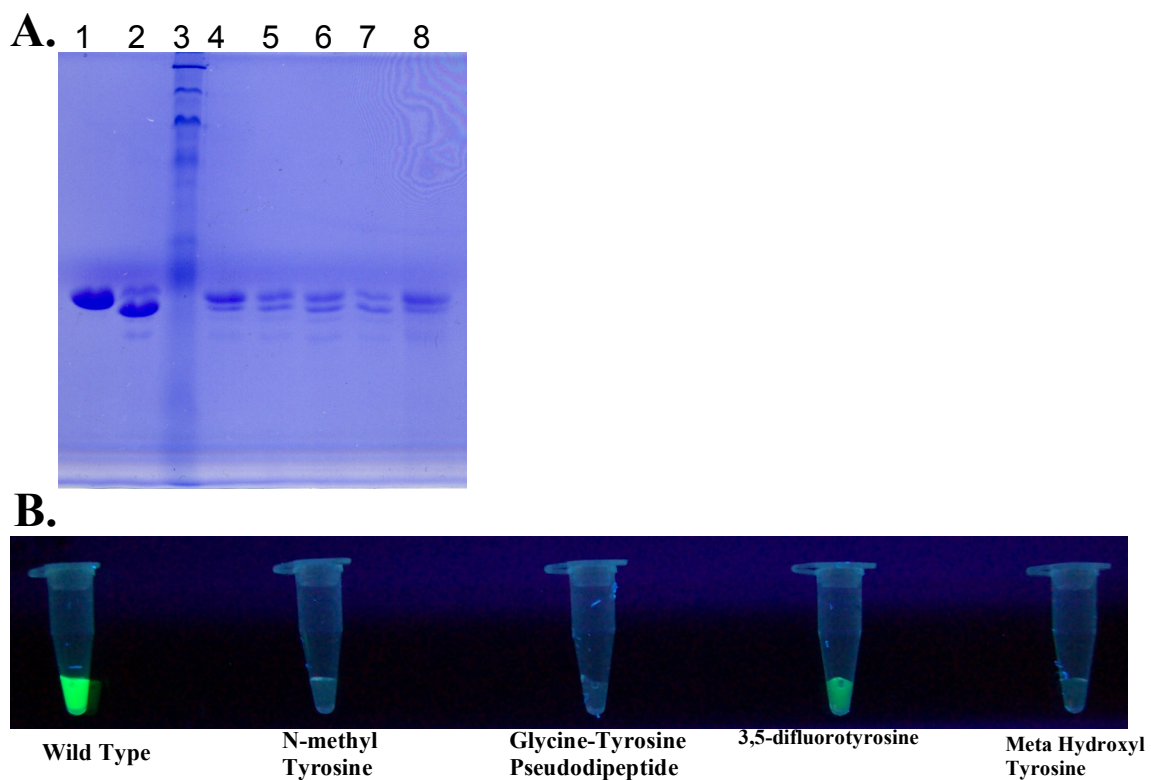


Figure 5.4. Native chemical ligation of peptide thioesters incorporating unnatural amino-acids to 50T70GFP α A206K. **A.** ligation products were *de novo* folded and concentrated. Samples are as follows from left to right; 50GFP α A206K entero, 50T70GFP α A206K, NEB Broad Range Ladder, ligation products- wt, N-methyl tyrosine, Pseudodipeptide, 3,5-difluorotyrosine, meta-hydroxyl tyrosine **C.** Visible evaluation of chromophore formation in each sample by exposure to 365 nm transillumination.

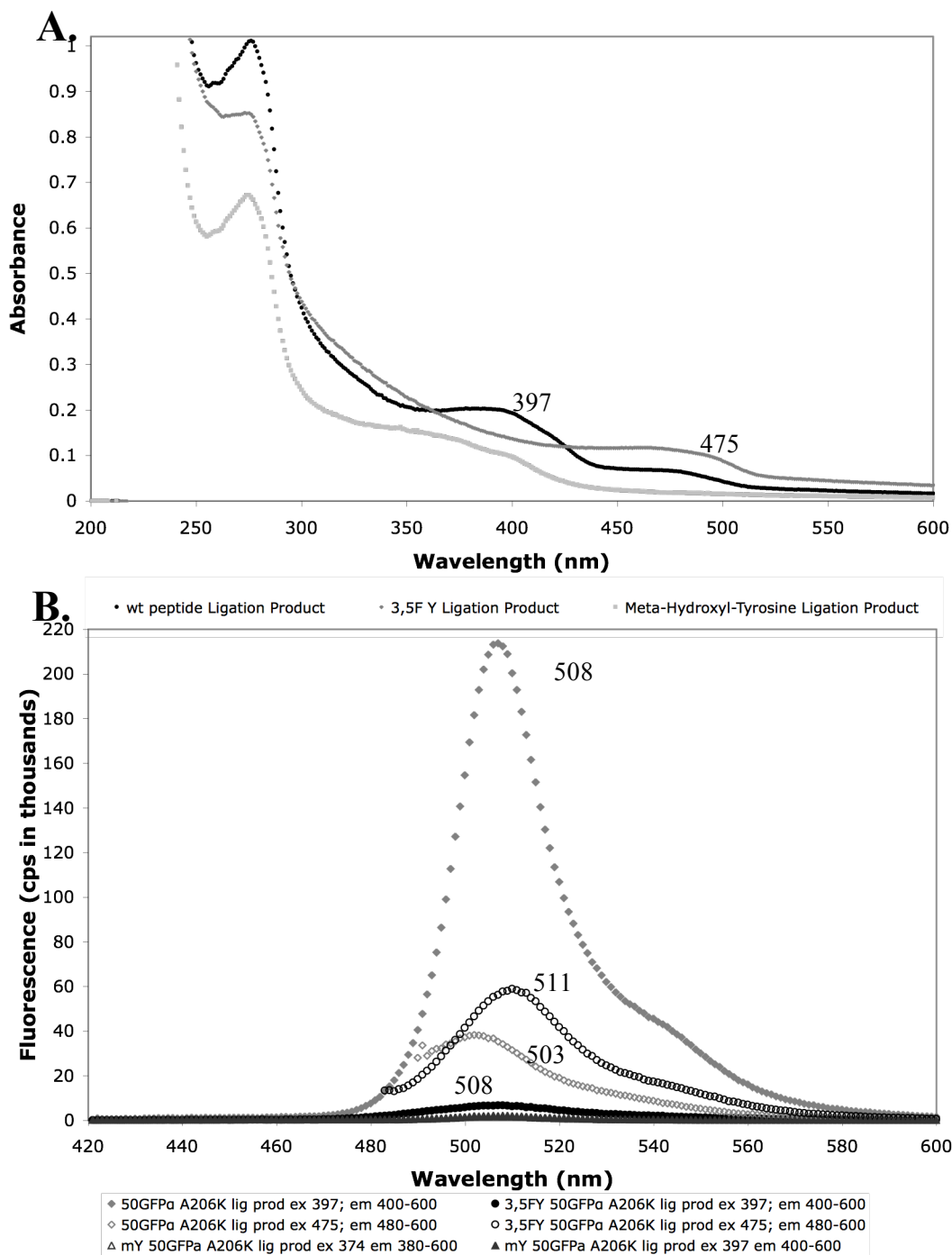


Figure 5.5. Absorption and fluorescence spectra of folded ligation products. *de novo* folded peptide ligation products concentrated to approximately 1.5 mL; absorption spectrum in (A) and fluorescence emission in (B) with listed excitation (ex) and emission (em) wavelengths. In the fluorescence spectrum, proteins are of equimolar concentration based on the major absorption band matching the wt peptide ligation product and fluorescence emission peaks are noted.

The ligation reactions were *de novo* folded and purified by application to Ni-NTA resin (Figure 5.4). The purified protein was then studied by UV visible absorption and fluorescence spectroscopy (Figure 5.5).

The substitution of two fluorine atoms adjacent to the *para*-hydroxyl of the tyrosine in the 3,5-difluorotyrosine peptide ligation product forces the formed chromophore into a completely anionic form, showing no absorption band at 397 nm and only weak fluorescence signal at 508 nm. Interestingly, the anionic fluorescence emission of this protein is shifted to 511 nm from the expected 503 nm anionic emission seen in the wild-type peptide ligation product.

As the 3,5-difluorotyrosine peptide construct demonstrates an absorption and fluorescence emission spectrum similar to S65T GFP by exhibiting a single absorption (475 nm) and fluorescence emission peak (511 nm vs. S65T GFP at 510 nm), investigation into whether this difluorinated protein exhibits a similar decrease in fluorescence upon a decrease in pH were undertaken [14-16] (Figure 5.6). While both the wild-type peptide ligation product and a sample of the natively expressed 50GFP α A206K protein show a decrease in protein fluorescence with a shift of the emission peak from approximately 503 nm to 499 nm, similar to natively expressed wtGFP, the 3,5-difluorotyrosine derivative shows a much larger shift from the 511 nm emission peak to approximately 501 nm. This result differs from the titration of S65T GFP in that the S65T GFP does not exhibit this shift in peak emission [14]. Based on a global incorporation of either 3-fluorotyrosine or 2-fluorotyrosine into EGFP (F64L, S65T mutations), emission maxima either blue shift (504 nm for 2-

fluorotyrosine) or red shift (514 nm for 3-fluorotyrosine) [5]. The electron withdrawing properties of two fluorine atoms adjacent to the phenol hydroxyl in the 3,5-difluorotyrosine derivative would suggest a larger red shift than the 3-fluorotyrosine derivative, therefore this discrepancy must be due to the S65T mutation in the 3-fluorotyrosine derivative that drives the emission further into the red than the 3,5-difluorotyrosine in the context of the wt sequence. This result would also suggest that the hydrogen-bonding network present in the 3,5-difluorotyrosine derivative that is disrupted in the S65T mutant has the effect of stabilizing the excited state and limiting the amount by which the emission is red shifted. Additionally, the shift in absorption and emission spectra in the 3,5-difluorotyrosine derivatives that appear spectrally like S65T may simply be due to the tendency of the phenol to be ionized (i.e. lower pKa), thereby exhibiting a preferentially ionized chromophore still within the wild-type hydrogen bonding network. This reasoning also explains the weak emission band seen at 508nm when irradiated into the neutral 397 nm band in the 3,5-difluorotyrosine ligation product, indicative of a minor neutral chromophore species that would not be present if this unnatural amino-acid truly behaved as S65T GFP does.

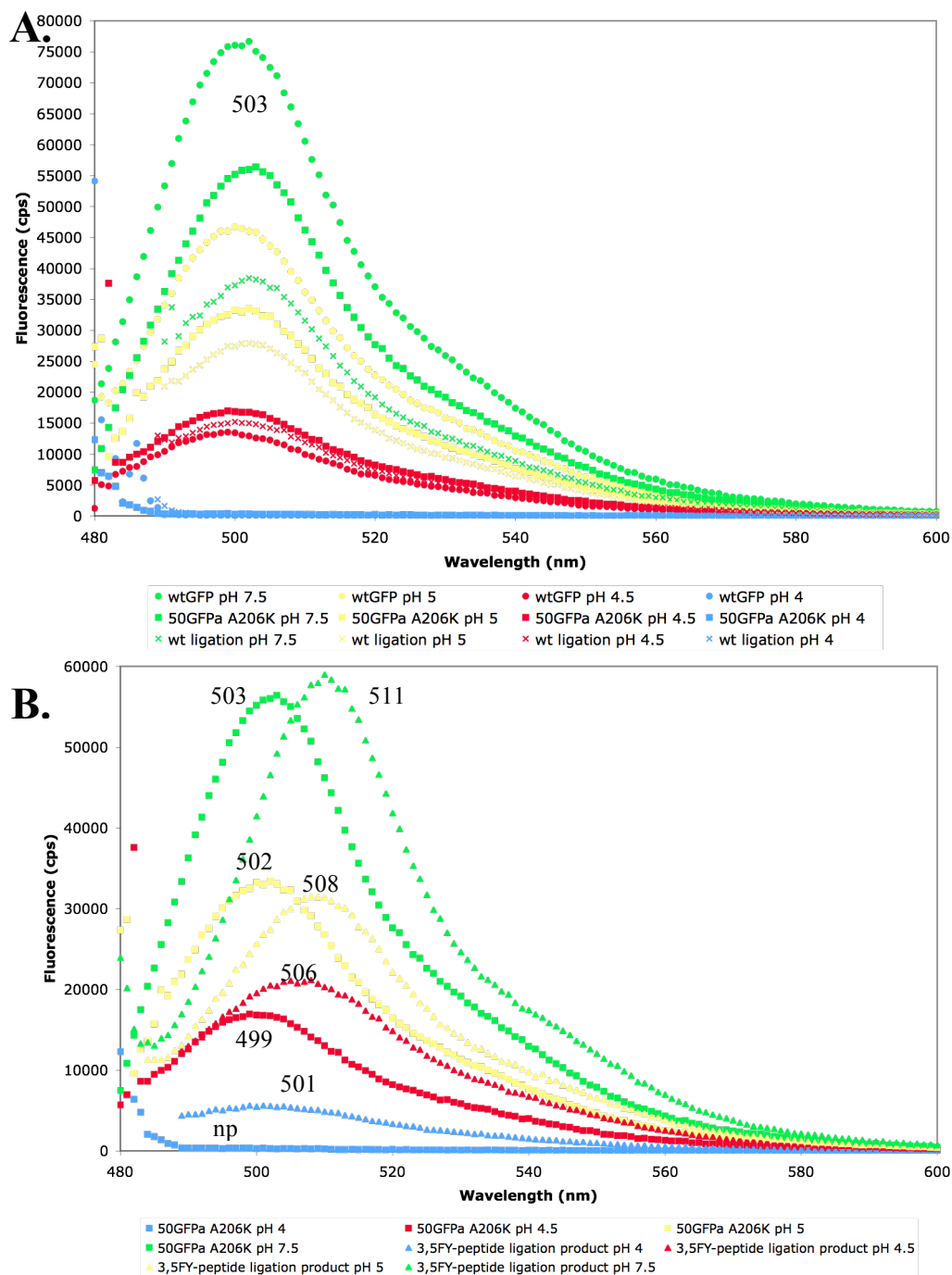


Figure 5.6. pH titration of 3,5-difluorotyrosine, wild-type ligation products compared to natively expressed wtGFP and 50-GFP α A206K. Equimolar protein based on 475 nm absorption was diluted into differing pH buffer (pH 7.5 green, pH 5 yellow, pH 4.5 red, pH 4 blue), excited at 475 nm, and fluorescence emission scanned from 480 to 600nm. **A.** Overlay of natively expressed wtGFP, 50-GFP α A206K, and the wild type peptide thioester ligation product **B.** overlay of 50-GFP α A206K with the 3,5-difluorotyrosine ligation product. Fluorescence maxima are as indicated, np = no peak.

The meta-hydroxyl tyrosine derivative shows only a band corresponding to the absorption of a neutral chromophore at 397 nm and, interestingly, an additional peak at 374 nm. The presence of absorption bands other than at 220 nm and 280 nm indicates that the conjugated chromophore is present in this sample, indicating that the movement of the hydroxyl group does not affect chromophore formation. Additionally, the presence of absorption bands at 397 nm indicates that the protein exists solely as a neutral A state species (lacking the anionic absorption at 475 nm). This result would suggest that the meta hydroxyl of the tyrosine is positioned so that the chromophore is prevented from ionizing, highlighting the importance of the hydrogen bonded residues around the para-hydroxyltyrosine that stabilize the excited anionic chromophore in wtGFP. This construct has an exceptionally weak fluorescence signal (Figure 5.5B). Further TRF work may elucidate the nature and timescales of fluorescence of this potential A state-only chromophore.

The 3,5-difluorinated peptide ligation product shows a major absorption peak corresponding to the anionic form of the chromophore. Irradiation into the neutral band of the chromophore shows a highly diminished fluorescence peak relative to the wild-type ligation product, confirming the shift in ionization state of the chromophore caused by fluorination. Interestingly, irradiation into the major absorption band at 475nm gives rise to a fluorescence peak at 511nm instead of the expected 503 nm. This is a larger shift than is seen in previous EGFP (with only one anionic absorption band) 2-fluorotyrosine (510 to 504 nm) and 3-fluorotyrosine (510 to 514 nm) constructs and is in agreement with the inductive effect of the fluorine isosteres seen

within these variants [5]. In an effort to induce the formation of a neutral chromophore within this protein, pH titration studies were conducted to establish the pKa of protonation for this chromophore (Figure 5.6). As with wtGFP, 50-GFP α A206K, and the ligated wt peptide product, the 3,5-difluorotyrosine loses fluorescence signal as the pH of the buffer is lowered. Loss of fluorescence in these variants is similar to the loss of fluorescence seen in the S65T variant, presumably due to increasing populations of low fluorescing neutral chromophore (A state) within the protein [5]. Interestingly, the 3,5-difluorotyrosine derivative still shows a minor fluorescence signal even at pH 4 where the non-fluorinated derivative shows no emission peak, indicating some ionized and fluorescing chromophore even at this pH (Figure 5.6). This variant seems to behave like a hydrogen bond network disrupted S65T GFP, but presumably within an intact hydrogen bonding network, therefore an investigation into the state of the hydrogen bonding network by ultrafast techniques may provide additional insights into this unique variant.

Conclusions and Future Directions

Previous attempts in the Tonge Lab by Dr. Deborah Stoner-Ma to create a GFP scaffold amenable to unnatural amino acid incorporation to elucidate the GFP chromophore formation mechanism focused on a three peptide approach with the synthetic peptide bearing the designed unnatural amino acids being the middle portion. While promising, these experiments were hampered by protein insolubility and low yields, but offered valuable guidance for the studies presented in this dissertation. Rather than continue with the three peptide approach, we decided to design a circular permutation that reduced the number of peptides to two and the ligation steps to one. At the onset of the project, our goal was to generate a circular permutant of GFP that behaved photophysically like wtGFP. After developing such a permutant, extensive testing revealed that it did indeed behave like wtGFP and could serve as a homologue to wtGFP. Importantly, this permutant retained the critical hydrogen-bonding network found within wtGFP that is often disrupted in other GFP mutants (i.e. S65T, EGFP). By developing a novel semi-denaturing endoproteolytic method to truncate this circular permutant, a GFP protein scaffold lacking the central chromophore bearing α helix with an N-terminal cysteine was produced in high yield. A method to generate peptide thioesters via Fmoc solid phase peptide synthesis that could be ligated to this truncated GFP scaffold was also developed.

Our aim in developing these methods was to incorporate unnatural amino acids into the synthetic peptide thioester that would, upon native chemical ligation to the truncated GFP scaffold and *de novo* folding of the ligation product, form the chromophore and study its formation, thereby contributing to our understanding of the biosynthetic route(s) of chromophore formation. In order to incorporate unnatural amino acid variants into peptide thioesters that would provide maximum insight into chromophore biosynthesis, we endeavored to synthesize several unnatural amino acids not yet commercially available.

Upon incorporation of unnatural amino acids into the circular permutant, we found that the N-methyltyrosine and tyrosine-glycine pseudodipeptide GFP variants did not exhibit any absorption peaks besides the protein absorption bands at 280 nm and 220 nm, indicating an interruption in chromophore formation. The route proposed by the Getzoff group places dehydration as a necessary prerequisite before oxidation during chromophore formation, while the Wachter model proposes an inverse of these steps [17-19]. The N-methyl tyrosine derivative was designed to prevent dehydration of the carbonyl of Ser65 and test if the C α - C β bond of the tyrosine is still capable of being oxidized. MALDI data indicates a protein mass with a dehydrated peroxide adduct intermediate that suggests dehydration is possible prior to oxidation, in support of the Getzoff model. Conversely, the pseudodipeptide ligation product shows a hydrated peroxide adduct, indicating that oxidation of the tyrosine side chain is possible prior to dehydration, in support of the Wachter model. These conflicting results may indicate that each step in chromophore formation occurs independent of

the other and mature chromophore is formed only when both steps have finished. In addition to preliminary MALDI mass spectroscopy, these samples are currently submitted to collaborators for high-resolution mass spectrometry to determine the intermediate states of chromophore formation. We believe this data will give us further insight into the progression of the chromophore biosynthesis.

Beyond elucidating the biosynthetic pathway for chromophore posttranslational modification, in this study we generated and incorporated other tyrosine modifications that alter the photophysical spectrum of GFP. Incorporation of a 3,5-difluorotyrosine derivative generates a chromophore, which based on absorption and emission spectra, is largely in the B state. While the excitation of A to A* and ESPT to I* can easily be studied, conversions of B states to I and A states are infrequent and therefore the photophysics are difficult to study [20, 21]. Our current experiments are focused on how the fluorination of the chromophore phenol affects both the pKa of the chromophore and the state of the hydrogen-bonding network surrounding this chromophore by pH titration studies. By incorporating other fluorotyrosine derivatives, we can effectively modulate the pKa of the phenol hydroxyl and look at trends of the A state neutral form converting to the anion B state and study the photophysical equilibrium between these states using TRF.

The meta-hydroxyl tyrosine derivative displayed only a neutral form of the chromophore (absorbance at 397 nm) with a second absorption peak at 374 nm. Interestingly, this meta-hydroxyl derivative has a very weak fluorescence emission indicating a serious disruption of the normal GFP photocycle simply due to the

relocating of the phenol hydroxyl group. This derivative may be maintained in a protonate state within the protein environment, thereby exhibiting A state absorption and emission behavior. Again, further TRF studies on this derivative may provide further insight into the modified photocycle of this variant.

The initial studies presented here open up extensive possibilities for future modification of the GFP chromophore using unnatural amino acids. Using the methods developed here, virtually any modification that will fit within the β barrel of GFP and that is compatible with solid phase peptide synthesis may be used with this system. Additionally, this method does not use cells to incorporate unnatural amino acids that are either toxic or incorporated inefficiently, as the modified amino-acid is synthetically incorporated into the peptide and then ligated to the cell expressed protein half. Other potential derivatives that modulate the chromophore include; the addition of extra conjugation that red shifts chromophore absorption and emission could be added to mimic the fluorescence of red fluorescent protein within the wtGFP environment, the phenol ring of the tyrosine could be linked to the imidiazolinone ring, thereby creating a three ring system within GFP, and incorporation of an *ortho*-tyrosine could lead to a tightened hydrogen bond network directly to Glu222. Spectroscopic analysis of each of these unnatural incorporations would provide valuable insight into photophysical behavior of the chromophore.

Additionally, our use of a dispeptide on within the active site of a protein is the first example in the field incorporating a tyrosine lacking a carbonyl into a protein. While this is a synthetically generated main chain modification, the loss of

the carbonyl makes the adjacent nitrogen a secondary amine as opposed to an amide peptide bond. This makes the amine more flexible without the steric influence from the carbonyl. This makes it a good candidate for incorporation into systems where decreased sterics or removal of a key carbonyl may aid in elucidating catalytic mechanisms.

While other groups have made circular permutants of GFP, we are first to exploit the position of our circular permutation to incorporate unnatural functionality and investigate chromophore formation and photophysics. Notably, our circular permutant creates a single polypeptide chain via the designed linker, which simplifies the native chemical ligation to one step.

Although analysis of the GFP variants characterized in this project is ongoing, we have already created a pivotal tool for the study of chromophore biosynthesis and GFP chromophore photophysics. While pH studies, additional mass spectrometry, and TFR spectroscopy will answer the open questions about the intrinsic properties of GFP in ways that previous studies could not, the strenuous tasks of creating the circular permutant, optimizing expression and solubility, incorporating several key unnatural functionalities, successfully ligating, and *de novo* folding the products is complete. Further investigations may now focus on generating deep insight into the structural and photophysical properties of GFP.

References

1. Cole, P.A., et al., *The Role of the Catalytic Base in the Protein Tyrosine Kinase Csk*. J. Biol. Chem., 1995. **270**(38): p. 22105-22108.
2. Gopishetty, B., et al., *Synthesis of 3,5-Difluorotyrosine-Containing Peptides: Application in Substrate Profiling of Protein Tyrosine Phosphatases*. Org. Lett. 2008. **10**(20): p. 4605-4608.
3. Seyedsayamdost, M.R., et al., *Mono-, di-, tri-, and tetra-substituted fluorotyrosines: new probes for enzymes that use tyrosyl radicals in catalysis*. J. Am. Chem. Soc., 2006. **128**(5): p. 1569-79.
4. Elsliger, M.A., et al., *Structural and spectral response of green fluorescent protein variants to changes in pH*. Biochemistry, 1999. **38**(17): p. 5296-301.
5. Pal, P.P., et al., *Structural and Spectral Response of Aequorea victoria Green Fluorescent Proteins to Chromophore Fluorination*. Biochemistry, 2005. **44**(10): p. 3663-3672.
6. Kim, K. and P.A. Cole, *Measurement of a Bronsted Nucleophile Coefficient and Insights into the Transition State for a Protein Tyrosine Kinase*. J. Am. Chem. Soc., 1997. **119**(45): p. 11096-11097.
7. Nagasawa, T., et al., *Syntheses of *l*-Tyrosine-Related Amino Acids by Tyrosine Phenol-lyase of *Citrobacter intermedius**. European Journal of Biochemistry, 1981. **117**(1): p. 33-40.
8. Phillips, R.S., R.L. Von Tersch, and F. Secundo, *Effects of Tyrosine Ring Fluorination on Rates and Equilibria of Formation of Intermediates in the Reactions of Carbon-Carbon Lyases*. Eur. J. Biochem., 1997. **244**(2): p. 658-663.
9. Haoyuan, C., G. Paul, and S.P. Robert, *Site-Directed Mutagenesis of His343 to Ala in *Citrobacter freundii* Tyrosine Phenol-Lyase*. Eur. J. Biochem., 1995. **229**(2): p. 540-549.

10. Mezo, A.R., R.P. Cheng, and B. Imperiali, *Oligomerization of uniquely folded mini-protein motifs: development of a homotrimeric betabetaalpha peptide*. J. Am. Chem. Soc., 2001. **123**(17): p. 3885-91.
11. Kirk, K.L., *Convenient synthesis of 3-fluoro-L-tyrosine and 3,5-difluoro-L-tyrosine*. J. Org. Chem., 1980. **45**(10): p. 2015-2016.
12. Yee, C.S., et al., *Generation of the R2 Subunit of Ribonucleotide Reductase by Intein Chemistry: Insertion of 3-Nitrotyrosine at Residue 356 as a Probe of the Radical Initiation Process*. Biochemistry, 2003. **42**(49): p. 14541-14552.
13. Wang, P. and L.P. Miranda, *Fmoc-Protein Synthesis: Preparation of Peptide Thioesters Using a Side-Chain Anchoring Strategy*. Int. J. Pept. Res. Ther., 2005. **11**(2): p. 117-123.
14. Kneen, M., et al., *Green fluorescent protein as a noninvasive intracellular pH indicator*. Biophys. J., 1998. **74**(3): p. 1591-9.
15. Tsien, R.Y., *The Green Fluorescent Protein*. Annual Review of Biochemistry, 1998. **67**(1): p. 509-544.
16. Heim, R., A.B. Cubitt, and R.Y. Tsien, *Improved green fluorescence*. Nature, 1995. **373**(6516): p. 663-664.
17. Rosenow, M.A., et al., *The crystal structure of the Y66L variant of green fluorescent protein supports a cyclization-oxidation-dehydration mechanism for chromophore maturation*. Biochemistry, 2004. **43**(15): p. 4464-72.
18. Barondeau, D.P., et al., *Understanding GFP Chromophore Biosynthesis: Controlling Backbone Cyclization and Modifying Post-translational Chemistry*. Biochemistry, 2005. **44**: p. 1960-1970.
19. Pouwels, L.J., et al., *Kinetic isotope effect studies on the de novo rate of chromophore formation in fast- and slow-maturing GFP variants*. Biochemistry, 2008. **47**(38): p. 10111-22.
20. Chatteraj, M., et al., *Ultra-fast excited state dynamics in green fluorescent protein: multiple states and proton transfer*. Proc. Natl. Acad. Sci. U.S.A., 1996. **93**(16): p. 8362-7.
21. van Thor, J.J., et al., *Phototransformation of green fluorescent protein with UV and visible light leads to decarboxylation of glutamate 222*. Nat. Struct. Biol., 2002. **9**(1): p. 37-41.

List of References

1. 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. 223-239.
2. Heim, R., A.B. Cubitt, and R.Y. Tsien, *Improved green fluorescence*. Nature, 1995. **373**(6516): p. 663-664.
3. Morise, H., O. Shimomura, F.H. Johnson, and J. Winant, *Intermolecular energy transfer in the bioluminescent system of Aequorea*. Biochemistry, 1974. **13**(12): p. 2656-2662.
4. Shimomura, O., *Structure of the chromophore of Aequorea green fluorescent protein*. Febs Lett., 1979. **104**(2): p. 220-222.
5. Prasher, D.C., V.K. Eckenrode, W.W. Ward, F.G. Prendergast, and M. J. Cormier, *Primary structure of the Aequorea victoria green-fluorescent protein*. Gene, 1992. **111**(2): p. 229-233.
6. Chalfie, M., Y. Tu, G. Euskirchen, W.W. Ward, and D.C. Prasher, *Green fluorescent protein as a marker for gene expression*. Science, 1994. **263**(5148): p. 802-805.
7. Inouye, S. and F.I. Tsuji, *Aequorea green fluorescent protein. Expression of the gene and fluorescence characteristics of the recombinant protein*. Febs Lett., 1994. **341**(2-3): p. 277-280.
8. Martin, C., *GREEN FLUORESCENT PROTEIN*. Photochemistry and Photobiology, 1995. **62**(4): p. 651-656.
9. Ormo, M., A.B. Cubitt, K. Kallio, L.A. Gross, and R.Y. Tsien, *Crystal Structure of the Aequorea victoria Green Fluorescent Protein*. Science, 1996. **273**(5280): p. 1392-1395.
10. Yang, F., L.G. Moss, and G.N. Phillips, Jr., *The molecular structure of green fluorescent protein*. Nat. Biotechnol., 1996. **14**(10): p. 1246-1251.

11. Cody, C.W., D.C. Prasher, W.M. Westler, F.G. Prendergast, and W.W. Ward, *Chemical structure of the hexapeptide chromophore of the Aequorea green-fluorescent protein*. *Biochemistry*, 1993. **32**(5): p. 1212-1218.
12. Brejc, K., T.K. Sixma, P.A. Kitts, S.R. Kain, R.Y. Tsien, M. Ormo, and S.J. Remington, *Structural Basis for Dual Excitation and Photoisomerization of the Aequorea victoria Green Fluorescent Protein*. *Proc. Natl. Acad. Sci. U.S.A.*, 1997. **94**(6): p. 2306-2311.
13. Jaye, A.A., D. Stoner-Ma, P. Matousek, M. Towrie, P.J. Tonge, and S.R. Meech, *Time Resolved Emission Spectra of Green Fluorescent Protein*. *Photochem. Photobiol.*, 2006 **82**(2): p. 373-379.
14. Lossau, H., A. Kummer, R. Heinecke, F. Pöllinger-Dammer, C. Kompa, G. Bieser, T. Jonsson, C.M. Silva, M.M. Yang, D.C. Youvan, and M.E. Michel-Beyerle, *Time-resolved spectroscopy of wild-type and mutant Green Fluorescent Proteins reveals excited state deprotonation consistent with fluorophore-protein interactions*. *Chemical Physics*, 1996. **213**(1-3): p. 1-16.
15. Chatteraj, M., B.A. King, G.U. Bublitz, and S.G. Boxer, *Ultra-fast excited state dynamics in green fluorescent protein: multiple states and proton transfer*. *Proc. Natl. Acad. Sci. U.S.A.*, 1996. **93**(16): p. 8362-8367.
16. Stoner-Ma, D., E.H. Melief, J. Nappa, K.L. Ronayne, P.J. Tonge, and S.R. Meech, *Proton relay reaction in green fluorescent protein (GFP): Polarization-resolved ultrafast vibrational spectroscopy of isotopically edited GFP*. *J. Phys. Chem. B*, 2006. **110**(43): p. 22009-22018.
17. Ward, W., W., H.J. Prentice, A.F. Roth, C.W. Cody, and S.C. Reeves, *Spectral Perturbations of the Aequorea Green-Fluorescent Protein*. *Photochem. Photobiol.*, 1982. **35**(6): p. 803-808.
18. Bizzarri, R., C. Arcangeli, D. Arosio, F. Ricci, P. Faraci, F. Cardarelli, and F. Beltram, *Development of a novel GFP-based ratiometric excitation and emission pH indicator for intracellular studies*. *Biophys. J.*, 2006. **90**(9): p. 3300-3314.
19. Ward, W.W. and S.H. Bokman, *Reversible Denaturation of Aequorea Green-Fluorescent Protein: Physical Separation and Characterization of the Renatured Protein*. *Biochemistry*, 1982. **21**: p. 4535-4540.
20. Pal, P.P., J.H. Bae, M.K. Azim, P. Hess, R. Friedrich, R. Huber, L. Moroder, and N. Budisa, *Structural and Spectral Response of Aequorea victoria Green*

- Fluorescent Proteins to Chromophore Fluorination*. Biochemistry, 2005. **44**(10): p. 3663-3672.
21. Cheng, L., J. Fu, A. Tsukamoto, and R.G. Hawley, *Use of green fluorescent protein variants to monitor gene transfer and expression in mammalian cells*. Nat. Biotechnol., 1996. **14**(5): p. 606-609.
 22. Cormack, B.P., R.H. Valdivia, and S. Falkow, *FACS-optimized mutants of the green fluorescent protein (GFP)*. Gene, 1996. **173**(1): p. 33-38.
 23. Delagrave, S., R.E. Hawtin, C.M. Silva, M.M. Yang, and D.C. Youvan, *Red-shifted excitation mutants of the green fluorescent protein*. Biotechnology (NY), 1995. **13**(2): p. 151-154.
 24. Reid, B.G. and G.C. Flynn, *Chromophore formation in green fluorescent protein*. Biochemistry, 1997. **36**(22): p. 6786-6791.
 25. Ehrig, T., D.J. O'Kane, and F.G. Prendergast, *Green-fluorescent protein mutants with altered fluorescence excitation spectra*. Febs Lett., 1995. **367**(2): p. 163-166.
 26. 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. 12501-12504.
 27. Miyawaki, A., O. Griesbeck, R. Heim, and R.Y. Tsien, *Dynamic and quantitative Ca²⁺ measurements using improved cameleons*. Proc. Natl. Acad. Sci. U.S.A., 1999. **96**: p. 2135-2140.
 28. Miyawaki, A., J. Llopis, R. Heim, J.M. McCaffery, J.A. Adams, M. Ikura, and R.Y. Tsien, *Fluorescent indicators for Ca²⁺ based on green fluorescent proteins and calmodulin*. Nature, 1997. **388**(6645): p. 882-887.
 29. Miyawaki, A., *Fluorescent proteins in a new light*. Nat. Biotechnol., 2004. **22**(11): p. 1374-1376.
 30. Cubitt, A.B., R. Heim, S.R. Adams, A.E. Boyd, L.A. Gross, and R.Y. Tsien, *Understanding, improving and using green fluorescent proteins*. Trends Biochem. Sci., 1995. **20**(11): p. 448-455.
 31. Pollok, B.A. and R. Heim, *Using GFP in FRET-based applications*. Trends Cell Biol., 1999. **9**(2): p. 57-60.

32. Barondeau, D.P., C.J. Kassmann, J.A. Tainer, and E.D. Getzoff, *Understanding GFP Chromophore Biosynthesis: Controlling Backbone Cyclization and Modifying Post-translational Chemistry*. *Biochemistry*, 2005. **44**: p. 1960-1970.
33. Barondeau, D.P., C.D. Putnam, C.J. Kassmann, J.A. Tainer, and E.D. Getzoff, *Mechanism and Energetics of Green Fluorescent Protein Chromophore Synthesis Revealed by Trapped Intermediate Structures*. *Proc. Natl. Acad. Sci. U.S.A.*, 2003. **100**(21): p. 12111-12116.
34. Rosenow, M.A., H.A. Huffman, M.E. Phail, and R.M. Wachter, *The crystal structure of the Y66L variant of green fluorescent protein supports a cyclization-oxidation-dehydration mechanism for chromophore maturation*. *Biochemistry*, 2004. **43**(15): p. 4464-4472.
35. Pouwels, L.J., L. Zhang, N.H. Chan, P.C. Dorrestein, and R.M. Wachter, *Kinetic isotope effect studies on the de novo rate of chromophore formation in fast- and slow-maturing GFP variants*. *Biochemistry*, 2008. **47**(38): p. 10111-10122.
36. Xie, J. and P.G. Schultz, *Adding amino acids to the genetic repertoire*. *Curr. Opin. Chem. Biol.*, 2005. **9**(6): p. 548-554.
37. Xie, J. and P.G. Schultz, *An expanding genetic code*. *Methods*, 2005. **36**(3): p. 227-238.
38. Liu, W., A. Brock, S. Chen, and P.G. Schultz, *Genetic incorporation of unnatural amino acids into proteins in mammalian cells*. *Nat. Meth.*, 2007. **4**(3): p. 239-244.
39. Wang, L., J. Xie, A.A. Deniz, and P.G. Schultz, *Unnatural amino acid mutagenesis of green fluorescent protein*. *J. Org. Chem.*, 2003. **68**(1): p. 174-176.
40. Muralidharan, V. and T.W. Muir, *Protein ligation: an enabling technology for the biophysical analysis of proteins*. *Nat. Meth.*, 2006. **3**(6): p. 429-438.
41. Khan, F., I. Kuprov, T.D. Craggs, P.J. Hore, and S.E. Jackson, *¹⁹F NMR Studies of the Native and Denatured States of Green Fluorescent Protein*. *J. Am. Chem. Soc.*, 2006. **128**(33): p. 10729-10737.
42. Kim, K. and P.A. Cole, *Measurement of a Bronsted Nucleophile Coefficient and Insights into the Transition State for a Protein Tyrosine Kinase*. *J. Am. Chem. Soc.*, 1997. **119**(45): p. 11096-11097.

43. Baird, G.S., D.A. Zacharias, and R.Y. Tsien, *Circular permutation and receptor insertion within green fluorescent proteins*. Proc. Natl. Acad. Sci. U.S.A., 1999. **96**(20): p. 11241-11246.
44. Topell, S., J. Hennecke, and R. Glockshuber, *Circularly permuted variants of the green fluorescent protein*. Febs Lett., 1999. **457**(2): p. 283-289.
45. Iwai, H., A. Lingel, and A. Plückthun, *Cyclic green fluorescent protein produced in vivo using an artificially split PI-Pful intein from Pyrococcus furiosus*. J. Biol. Chem., 2001. **276**(19): p. 16548-16554.
46. Bell, A.F., X. He, R.M. Wachter, and P.J. Tonge, *Probing the ground state structure of the green fluorescent protein chromophore using Raman spectroscopy*. Biochemistry, 2000. **39**(15): p. 4423-4431.
47. Demidov, V.V. and N.E. Broude, *Profluorescent protein fragments for fast bimolecular fluorescence complementation in vitro*. Nat. Protoc., 2006. **1**(2): p. 714-719.
48. Zacharias, D.A., J.D. Violin, A.C. Newton, and R.Y. Tsien, *Partitioning of lipid-modified monomeric GFPs into membrane microdomains of live cells*. Science, 2002. **296**(5569): p. 913-916.
49. Bell, A.F., D. Stoner-Ma, R.M. Wachter, and P.J. Tonge, *Light-driven decarboxylation of wild-type green fluorescent protein*. J. Am. Chem. Soc., 2003. **125**(23): p. 6919-6926.
50. van Thor, J.J., T. Gensch, K.J. Hellingwerf, and L.N. Johnson, *Phototransformation of green fluorescent protein with UV and visible light leads to decarboxylation of glutamate 222*. Nat. Struct. Biol., 2002. **9**(1): p. 37-41.
51. Hofmann, A., H. Iwai, S. Hess, A. Pluckthun, and A. Wlodawer, *Structure of cyclized green fluorescent protein*. Acta Cryst., 2002. **D58**(Pt 9): p. 1400-1406.
52. He, X., A.F. Bell, and P.J. Tonge, *Isotopic Labeling and Normal-Mode Analysis of a Model Green Fluorescent Protein Chromophore*. J. Phys. Chem. B, 2002. **106**(23): p. 6056-6066.
53. Hirata, R., Y. Ohsumi, A. Nakano, H. Kawasaki, K. Suzuki, and Y. Anraku, *Molecular structure of a gene, VMA1, encoding the catalytic subunit of H(+)-*

- translocating adenosine triphosphatase from vacuolar membranes of Saccharomyces cerevisiae*. J. Biol. Chem., 1990. **265**(12): p. 6726-6733.
54. Paulus, H., *The chemical basis of protein splicing*. Chem. Soc. Rev., 1998. **27**: p. 375-386.
 55. Muir, T.W., D. Sondhi, and P.A. Cole, *Expressed protein ligation: A general method for protein engineering*. Proc. Natl. Acad. Sci. U.S.A., 1998. **95**(12): p. 6705-6710.
 56. Dawson, P.E., T.W. Muir, I. Clark-Lewis, and S.B.H. Kent, *Synthesis of Proteins by Native Chemical Ligation*. Science, 1994. **266**(5186): p. 776-779.
 57. Tam, J.P., Y.A. Lu, C.F. Liu, and J. Shao, *Peptide synthesis using unprotected peptides through orthogonal coupling methods*. Proc. Natl. Acad. Sci. U.S.A., 1995. **92**(26): p. 12485-12489.
 58. Camarero, J., A. Adeva, and T. Muir, *3-Thiopropionic acid as a highly versatile multidetachable thioester resin linker*. Lett. Pept. Sci., 2000. **7**(1): p. 17-21.
 59. Clippingdale, A.B., C.J. Barrow, and J.D. Wade, *Peptide thioester preparation by Fmoc solid phase peptide synthesis for use in native chemical ligation*. J. Pept. Sci., 2000. **6**(5): p. 225-234.
 60. Blanco-Canosa, J.B. and P.E. Dawson, *An efficient Fmoc-SPPS approach for the generation of thioester peptide precursors for use in native chemical ligation*. Angew. Chem. Int. Ed. Engl., 2008. **47**(36): p. 6851-6855.
 61. Brik, A., E. Keinan, and P.E. Dawson, *Protein synthesis by solid-phase chemical ligation using a safety catch linker*. J. Org. Chem., 2000. **65**(12): p. 3829-3835.
 62. Jensen, K.J., J. Alsina, M.F. Songster, J. Vaigner, F. Albericio, and G. Barany, *Backbone Amide Linker (BAL) Strategy for Solid-Phase Synthesis of C-Terminal-Modified and Cyclic Peptides 1,2,3*. J. Am. Chem. Soc., 1998. **120**(22): p. 5441-5452.
 63. Dolphin, G.T., *A designed well-folded monomeric four-helix bundle protein prepared by Fmoc solid-phase peptide synthesis and native chemical ligation*. Chemistry, 2006. **12**(5): p. 1436-1447.

64. Ingenito, R., E. Bianchi, D. Fattori, and A. Pessi, *Solid Phase Synthesis of Peptide C-Terminal Thioesters by Fmoc/t-Bu Chemistry*. J. Am. Chem. Soc., 1999. **121**(49): p. 11369-11374.
65. Flavell, R.R., M. Huse, M. Goger, M. Trester-Zedlitz, J. Kuriyan, and T.W. Muir, *Efficient semisynthesis of a tetraphosphorylated analogue of the Type I TGFbeta receptor*. Org. Lett., 2002. **4**(2): p. 165-168.
66. Backes, B.J., D.R. Dragoli, and J.A. Ellman, *Chiral N-Acyl-tert-butanesulfinamides: The "Safety-Catch" Principle Applied to Diastereoselective Enolate Alkylations*. J. Org. Chem., 1999. **64**(15): p. 5472-5478.
67. Futaki, S., K. Sogawa, J. Maruyama, T. Asahara, M. Niwa, and H. Hojo, *Preparation of Peptide Thioesters using Fmoc-Solid-Phase Peptide Synthesis and its Application to the Construction of a Template-Assembled Synthetic Protein (TASP)*. Tet. Lett., 1997. **38**(35): p. 6237-6240.
68. Kitagawa, K., H. Adachi, Y. Sekigawa, T. Yagami, S. Futaki, Y.J. Gu, and K. Inoue, *Total chemical synthesis of large CCK isoforms using a thioester segment condensation approach*. Tetrahedron, 2004. **60**(4): p. 907-918.
69. Shogren-Knaak, M.A., C.J. Fry, and C.L. Peterson, *A Native Peptide Ligation Strategy for Deciphering Nucleosomal Histone Modifications*. J. Biol. Chem., 2003. **278**(18): p. 15744-15748.
70. von Eggelkraut-Gottanka, R., A. Klose, A.G. Beck-Sickinger, and M. Beyermann, *Peptide [alpha]thioester formation using standard Fmoc-chemistry*. Tet. Lett., 2003. **44**(17): p. 3551-3554.
71. Warren, J.D., J.S. Miller, S.J. Keding, and S.J. Danishefsky, *Toward Fully Synthetic Glycoproteins by Ultimately Convergent Routes: A Solution to a Long-Standing Problem*. J. Am. Chem. Soc., 2004. **126**(21): p. 6576-6578.
72. Mezo, A.R., R.P. Cheng, and B. Imperiali, *Oligomerization of uniquely folded mini-protein motifs: development of a homotrimeric betabetaalpha peptide*. J. Am. Chem. Soc., 2001. **123**(17): p. 3885-3891.
73. Camarero, J.A., G.J. Cotton, A. Adeva, and T.W. Muir, *Chemical ligation of unprotected peptides directly from a solid support*. J. Pept. Res., 1998. **51**(4): p. 303-316.

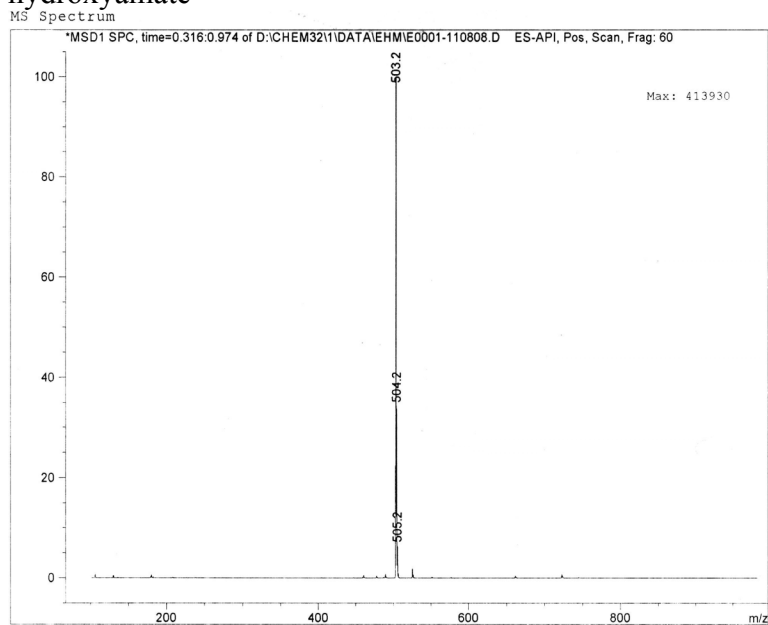
74. Johnson, E.C. and S.B. Kent, *Insights into the mechanism and catalysis of the native chemical ligation reaction*. J. Am. Chem. Soc., 2006. **128**(20): p. 6640-6646.
75. Dawson, P.E., M.J. Churchill, M.R. Ghadiri, and S.B.H. Kent, *Modulation of Reactivity in Native Chemical Ligation through the Use of Thiol Additives*. J. Am. Chem. Soc., 1997. **119**(19): p. 4325-4329.
76. Lee, J.Y., B.J. Byun, and Y.K. Kang, *Conformational Preferences and pKa Value of Cysteine Residue*. J. Phys. Chem. B, 2008. **112**(36): p. 11189-11193.
77. Danford, J.J., A.M. Arif, and L.M. Berreau, *Thioester Hydrolysis Promoted by a Mononuclear Zinc Complex*. Inorg. Chem., 2009. **49**(3): p. 778-780.
78. Kaiser, E., R.L. Colescott, C.D. Bossinger, and P.I. Cook, *Color test for detection of free terminal amino groups in the solid-phase synthesis of peptides*. Anal. Biochem., 1970. **34**(2): p. 595-598.
79. Hackenberger, C.P.R., M.M. Chen, and B. Imperiali, *Expression of N-terminal Cys-protein fragments using an intein refolding strategy*. Bioorg. Med. Chem., 2006. **14**(14): p. 5043-5048.
80. Bollhagen, R., M. Schmiedberger, K. Barlos, and E. Grell, *(A new reagent for the cleavage of fully protected peptides synthesised on 2-chlorotrityl chloride resin*. J. Chem. Soc. Chem. Comm., 1994(22): p. 2559 - 2560.
81. Barlos, K., D. Gatos, and W. Schafer, *Synthese von Prothymosin alpha (ProTalpha), einem aus 109 Aminosäuren aufgebauten Protein*. Angew. Chem., 1991. **103**(5): p. 572-575.
82. Barlos, K., D. Gatos, J. Kallitsis, G. Papaphotiu, P. Sotiriu, Y. Wenqing, and W. Schafer, *Darstellung gesch, tzter peptid-fragmente unter einsatz substituierter triphenylmethyl-harze*. Tet. Lett., 1989. **30**(30): p. 3943-3946.
83. Barlos, K., O. Chatzi, D. Gatos, and G. Stavropoulos, *2-Chlorotrityl chloride resin. Studies on anchoring of Fmoc-amino acids and peptide cleavage*. Int. J. Pept. Prot. Res., 1991. **37**(6): p. 513-520.
84. Barlos, K., D. Gatos, S. Kapolos, C. Poulos, W. Schafer, and W.Q. Yao, *Application of 2-chlorotrityl resin in solid phase synthesis of (Leu15)-gastrin I and unsulfated cholecystinin octapeptide. Selective O-deprotection of tyrosine*. Int. J. Pept. Prot. Res., 1991. **38**(6): p. 555-561.

85. Barlos, K., D. Gatos, S. Kutsogianni, G. Papaphotiou, C. Poulos, and T. Tsegenidis, *Solid phase synthesis of partially protected and free peptides containing disulphide bonds by simultaneous cysteine oxidation-release from 2-chlorotrityl resin*. *Int. J. Pept. Prot. Res*, 1991. **38**(6): p. 562-568.
86. Hackeng, T.M., J.H. Griffin, and P.E. Dawson, *Protein synthesis by native chemical ligation: expanded scope by using straightforward methodology*. *Proc. Natl. Acad. Sci. U.S.A.*, 1999. **96**(18): p. 10068-10073.
87. Novagen User Protocol TB205 Factor Xa Kits
http://www.emdchemicals.com/chemdat/en_CA/Merck-US-Site/USD/ViewProductDocuments-File?ProductSKU=EMD_BIO-69037&DocumentType=USP&DocumentId=%2Femd%2Fbiosciences%2Fuserprotocols%2Fen-US%2FTB205.pdf&DocumentSource=GDS
88. Remington, S.J., *Fluorescent proteins: maturation, photochemistry and photophysics*. *Curr. Opin. Struct. Biol.*, 2006. **16**(6): p. 714-721.
89. Salvi, J.P., N. Walchshofer, and J. Paris, *Formation of bis (Fmoc-amino ethyl)-N-glycine derivatives by reductive amination of Fmoc-amino aldehydes with NaBH₃CN*. *Tet. Lett.*, 1994. **35**(8): p. 1181-1184.
90. Cole, P.A., M.R. Grace, R.S. Phillips, P. Burn, and C.T. Walsh, *The Role of the Catalytic Base in the Protein Tyrosine Kinase Csk*. *J. Biol. Chem.*, 1995. **270**(38): p. 22105-22108.
91. Gopishetty, B., L. Ren, T.M., Waller, A.S. Wavreille, M. Lopez, A. Thakkar, J. Zhu, and D. Pei, *Synthesis of 3,5-Difluorotyrosine-Containing Peptides: Application in Substrate Profiling of Protein Tyrosine Phosphatases*. *Org. Lett.*, 2008. **10**(20): p. 4605-4608.
92. Seyedsayamdost, M.R., S.Y. Reece, D.G. Nocera, and J. Stubbe, *Mono-, di-, tri-, and tetra-substituted fluorotyrosines: new probes for enzymes that use tyrosyl radicals in catalysis*. *J. Am. Chem. Soc.*, 2006. **128**(5): p. 1569-1579.
93. Elsliger, M.A., R.M. Wachter, G.T. Hanson, K. Kallio, and S.J. Remington, *Structural and spectral response of green fluorescent protein variants to changes in pH*. *Biochemistry*, 1999. **38**(17): p. 5296-5301.
94. Nagasawa, T., T. Utagawa, J. Goto, C.J. Kim, Y. Tani, H. Kumagai, and H. Yamada, *Syntheses of L-Tyrosine-Related Amino Acids by Tyrosine Phenol-Lyase of *Citrobacter intermedius**. *Eur. J. Biochem.*, 1981. **117**(1): p. 33-40.

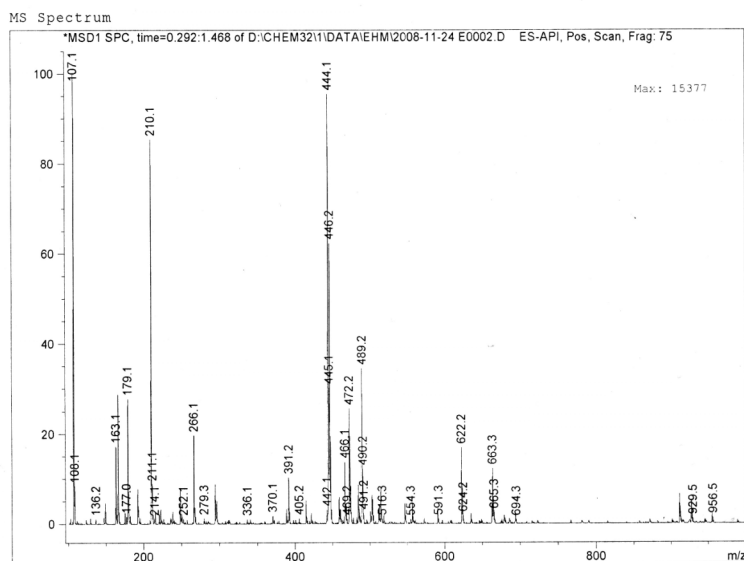
95. Phillips, R.S., R.L. Von Tersch, and F. Secundo, *Effects of Tyrosine Ring Fluorination on Rates and Equilibria of Formation of Intermediates in the Reactions of Carbon-Carbon Lyases*. Eur. J. Biochem., 1997. **244**(2): p. 658-663.
96. Haoyuan, C., G. Paul, and S.P. Robert, *Site-Directed Mutagenesis of His343Ala in Citrobacter freundii Tyrosine Phenol-Lyase*. Eur. J. Biochem. 1995. **229**(2): p. 540-549.
97. Kirk, K.L., *Convenient synthesis of 3-fluoro-L-tyrosine and 3,5-difluoro-L-tyrosine*. J. Org. Chem., 1980. **45**(10): p. 2015-2016.
98. Yee, C.S., M.R. Seyedsayamdost, M.C.Y Chang, D.. Nocera, and J. Stubbe, *Generation of the R2 Subunit of Ribonucleotide Reductase by Intein Chemistry: Insertion of 3-Nitrotyrosine at Residue 356 as a Probe of the Radical Initiation Process*. Biochemistry, 2003. **42**(49): p. 14541-14552.
99. Wang, P. and L.P. Miranda, *Fmoc-Protein Synthesis: Preparation of Peptide Thioesters Using a Side-Chain Anchoring Strategy*. Int. J. Pept. Res. Ther., 2005. **11**(2): p. 117-123.
100. Kneen, M., J. Farinas, Y. Li, and A.S. Verkman, *Green fluorescent protein as a noninvasive intracellular pH indicator*. Biophys. J., 1998. **74**(3): p. 1591-1599.
101. Tsien, R.Y., *The Green Fluorescent Protein*. Annual Review of Biochemistry, 1998. **67**(1): p. 509-544.

Appendix 1 – Pseudo dipeptide synthesis intermediate mass spectra.

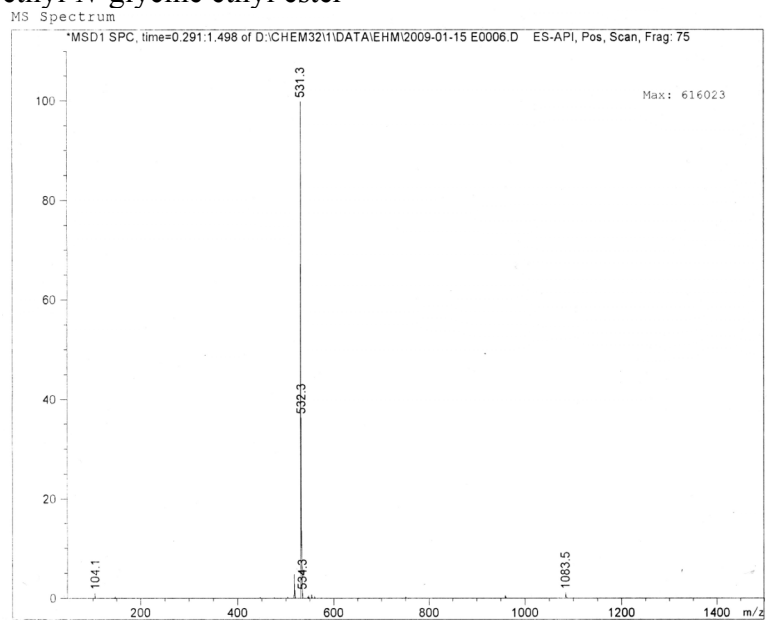
Compound 2: *N*-(Fluorenylmethyloxycarbonyl)-L-tyrosine (*tert*-butyl ester)-hydroxyamate



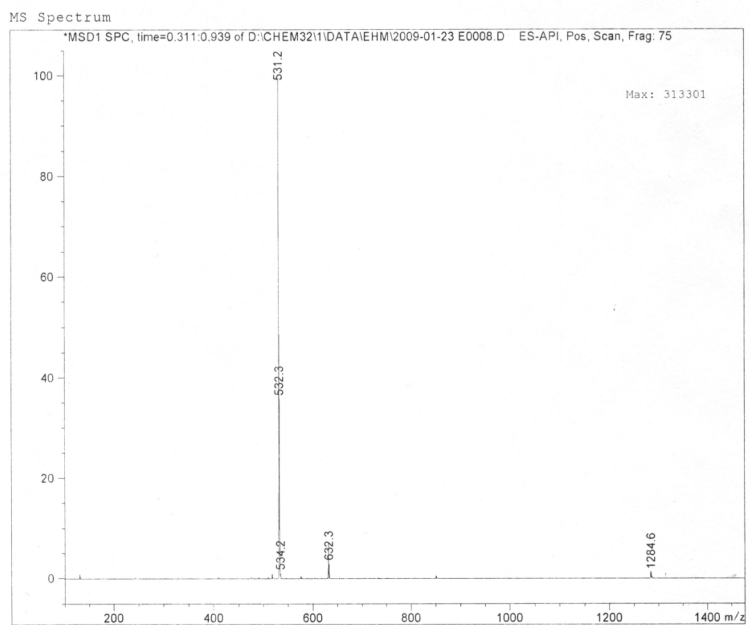
Compound 3: *N*-(Fluorenylmethyloxycarbonyl)-L-tyrosinal (*tert*-butyl ester)



Compound 4: *N*-(Fluorenylmethyloxycarbonyl)-L-tyrosyl (*tert*-butyl ester) amino ethyl-*N*-glycine ethyl ester



Compound 5: *N*-(Fluorenylmethyloxycarbonyl)-L-tyrosyl (*tert*-butyl ester) amino ethyl-*N,N*-(*tert*-butyloxy carbamate) glycine ethyl ester



Compound 6: *N*-(Fluorenylmethoxycarbonyl)-*L*-tyrosyl (*tert*-butyl ester) amino ethyl-*N,N*-(*tert*-butyloxy carbamate) glycine

